(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118591625 A
(43)申请公布日 2024.09.03

(21)申请号 202380018612.0
(22)申请日 2023.01.27
(30)优先权数据
63/304,565 2022.01.28 US
(85)PCT国际申请进入国家阶段日
2024.07.24
(86)PCT国际申请的申请数据
PCT/US2023/061500 2023.01.27
(87)PCT国际申请的公布数据
WO2023/147507 EN 2023.08.03
(71)申请人 桑格摩生物治疗股份有限公司
地址 美国
(72)发明人 S·阿朗根迪 F·A·福瑟 J·C·米勒 N·斯卡洛特
(74)专利代理机构 北京市铸成律师事务所 11313
专利代理师 陈艳娟 王艳波
(51)Int.Cl.
C12N 9/22(2006.01)
C12N 15/10(2006.01)
C12N 15/90(2006.01)

权利要求书2页 说明书41页
序列表（电子公布） 附图16页

(54)发明名称

用于位点特异性基因编辑的丝氨酸重组酶系统

(57)摘要

本公开提供了工程化大丝氨酸重组酶和含有所述重组酶的用于基因编辑的系统。

CN 118591625 A

1.一种Bxb1重组酶的非天然存在的变体,所述变体相对于野生型Bxb1重组酶具有改变的DNA靶特异性,包含锌带域(ZD)或重组酶域(RD)内的一个或多个氨基酸突变。

2.如权利要求1所述的Bxb1重组酶变体,其中所述一个或多个氨基酸突变发生在位置147、148、149、154、155、156、158、197、198、230、231、232、233、237、257、309、312、314、315、316、318、323、324、325、326和335(根据SEQ ID NO:1编号)中的一者或多者处。

3.如权利要求1所述的Bxb1重组酶变体,其中所述所述一个或多个氨基酸突变选自F314A、F314C、F314D、F314E、F314G、F314H、F314I、F314L、F314N、F314Q、F314S、F314T、F314V、F314W、F314Y、A315F、A315G、A315H、A315I、A315M、A315N、A315S、A315T、A315W、A315Y、G316A、G316C、G316D、G316E、G316F、G316H、G316I、G316K、G316L、G316M、G316P、G316Q、G316R、G316S、G316T、G316V、G316W、G316Y、G318I、G318K、G318R、G318W、R323G、R323K、R325D、R325E、R325K、R325L、R325M、R325N、R325Q、R325S、R325W、F147A、F147K、F147R、N148Q、N148T、L158D、L158N、L158S、L158T、L158W、P197H、P197R、P197T、S231F、S231G、S231H、S231K、S231R、S231V、S231Y、T233F、T233H、T233K、T233R、T233W、T233Y、R237A、R237C、R237D、R237E、R237F、R237G、R237H、R237I、R237K、R237L、R237M、R237N、R237P、R237Q、R237S、R237T、R237V、R237W、R237Y和D257K(根据SEQ ID NO:1编号)。

4.一种φC31整合酶、Pa557重组酶或Pa570重组酶的非天然存在的变体,所述变体包含一个或多个相对于野生型对应物产生改变的DNA靶特异性的氨基酸突变。

5.如权利要求4所述的φC31整合酶变体,其中所述一个或多个氨基酸突变发生在位置273、275、279、375、376、377、379和386(根据SEQ ID NO:77编号)中的一者或多者处。

6.如权利要求5所述的φC31整合酶变体,其中所述一个或多个氨基酸突变选自D273、A275、R279、K375、R376、G377、E379和R386(根据SEQ ID NO:77编号)。

7.如权利要求4所述的Pa557重组酶变体,其中所述一个或多个氨基酸突变发生在位置236、238、242、327、328、329、331和338(根据SEQ ID NO:78编号)中的一者或多者处。

8.如权利要求7所述的Pa557重组酶变体,其中所述一个或多个氨基酸突变选自G236、A238、A242、R327、T328、G329、G331和R338(根据SEQ ID NO:78编号)。

9.如权利要求4所述的Pa570重组酶变体,其中所述一个或多个氨基酸突变发生在位置243、245、249、333、334、335、337和344(根据SEQ ID NO:79编号)中的一者或多者处。

10.如权利要求9所述的Pa570重组酶变体,其中所述一个或多个氨基酸突变选自E243、S245、K249、I333、N334、P335、I337和Q344(根据SEQ ID NO:79编号)。

11.如权利要求1-10中任一项所述的重组酶变体,其中所述DNA靶标对基因的特定等位基因具有特异性。

12.如权利要求1-11中任一项所述的重组酶变体,其中所述DNA靶标位于细胞中。

13.如权利要求12所述的重组酶变体,其中所述细胞是真核细胞,任选地是哺乳动物细胞,进一步任选地是人类细胞。

14.一种编码如权利要求1-13中任一项所述的重组酶变体的核酸分子。

15.一种包含如权利要求14所述的核酸分子的载体。

16.如权利要求15所述的载体,其中所述载体是质粒或病毒载体,任选地选自腺相关病毒载体、腺病毒载体或慢病毒载体。

17.一种用于编辑细胞中的DNA的系统,所述系统包括如权利要求1-13中任一项所述的

重组酶变体、如权利要求14所述的核酸分子、或如权利要求15或16所述的载体。

18.如权利要求17所述的系统,所述系统还包括供体DNA,任选地其中所述供体DNA是环状的或线性的。

19.如权利要求18所述的系统,其中所述供体DNA包含选自以下的序列:SED ID NO:80-81和134-136以及表3、4和5中发现的序列。

20.如权利要求18或19所述的系统,其中所述供体DNA通过质粒或病毒载体递送。

21.如权利要求17-20中任一项所述的系统,其中所述编辑包括将DNA整合至所述细胞的基因组中。

22.如权利要求17-20中任一项所述的系统,其中所述编辑包括所述细胞的基因组中的DNA的切除或倒位。

23.如权利要求17-20中任一项所述的系统,其中所述编辑包括所述细胞的基因组中的染色体易位。

24.如权利要求17-23中任一项所述的系统,其中所述编辑发生在包含选自以下的序列的基因组区域处:SED ID NO:80-81和134-136以及表3、4和5中发现的序列。

25.一种编辑细胞的基因组的方法,所述方法包括向所述细胞提供如权利要求17-24中任一项所述的系统。

26.如权利要求26所述的方法,其中编辑多于一个基因组区域。

27.如权利要求25或26所述的方法,其中所述编辑导致从所述基因组中切除DNA、使所述基因组中的DNA倒位、染色体交换、重组酶介导的盒式交换(RMCE)和/或将供体DNA整合至所述基因组中。

28.一种包含如权利要求17-24中任一项所述的系统的细胞或其后代。

29.一种通过如权利要求25-27中任一项所述的方法编辑的细胞或其后代。

30.如权利要求28或29所述的细胞或后代,其中所述细胞是真核细胞,任选地是哺乳动物细胞,进一步任选地是人类细胞。

31.一种治疗有需要的受试者的疾病的方法,所述方法包括向所述受试者施用如权利要求30所述的细胞或后代。

32.如权利要求30所述的细胞或后代,其用于治疗有需要的受试者的疾病。

33.一种如权利要求30所述的细胞或后代的用途,其用于制造供治疗有需要的受试者的疾病用的药物。

34.如权利要求3所述的Bxb1重组酶变体,其中相对于野生型Bxb1重组酶,所述变体介导真核基因组中内源性靶位点处增加的转基因插入。

用于位点特异性基因编辑的丝氨酸重组酶系统

[0001] 相关申请的交叉引用

[0002] 本申请要求2022年1月28日提交的美国临时申请第63/304,565号的优先权,所述申请的内容以全文引用的方式并入本文中。

[0003] 序列表

[0004] 本申请含有序列表,所述序列表已以XML格式以电子方式提交。所述序列表特此以全文引用的方式并入。所述XML副本创建于2023年1月26日,名为025297.WO043.xml且大小为361,227字节。

[0005] 发明背景

[0006] 大丝氨酸重组酶(LSR)是由温和噬菌体编码的酶家族。这些酶能够以位点特异性的方式精确地切割和重组DNA,将DNA元件移入和移出噬菌体(attP)与宿主细菌(attB)中的短DNA附着位点之间的细菌染色体(Duyne和Rutherford,Crit Rev Biochem Mol Biol.(2013)48(5):476-91;Smith,Microbiol Spectr.(2015)3(4):doi10.1128/microbiolspec)。LSR介导的DNA重组的高度定向和受控过程使其成为基因工程化和基因疗法中的潜在工具,其中需要在基因组中进行位点特异性DNA整合、切除、倒位或盒式交换。

[0007] Bxb1重组酶,也称为Bxb1整合酶,是一种由噬菌体Bxb1编码的LSR,其促进噬菌体DNA整合至结核分枝杆菌(Mycobacterium smegmatis)基因组中(Russell等人,Biotechniques(2018)40(4):doi.org/10.2144/000112150)。重组酶含有一个与较小的分解酶/转化酶相似的N端催化域和一个较大的C端域,负责协调LSR特有的活性。此C端域进一步分为重组酶域(RD)和含有卷曲螺旋(CC)基序的锌带域(ZD)(Rutherford等人,Nucleic Acids Res.(2013)41(17):8341-56)。在重组事件期间,Bxb1重组酶二聚体的ZD和RD域与双链DNA片段上的附着位点(attB或attP)结合,而另一Bxb1重组酶二聚体的ZD和RD域与另一双链DNA片段上的另一附着位点(attP或attB)结合。随后两个二聚体的复合导致Bxb1重组酶四聚体的形成,并将两个DNA片段上的attP和attB位点结合在一起。四聚体活性位点处的丝氨酸残基在中心二核苷酸碱基对处形成共价键并裂解双链DNA。这种裂解使中心二核苷酸变成3'突出端。如果attB和attP序列的中心二核苷酸相同,则旋转所得蛋白质/DNA复合物,并连接先前分离的DNA分子或片段。此过程会产生附着R(attR)和附着(attL)位点,在不存在额外辅因子的情况下,它们不再是重组酶的底物。

[0008] 尽管大丝氨酸重组酶提供了精确的基因编辑,但由于其在人类基因组(包括治疗相关基因)中的结合位点的稀有性,其作为基因工程化和治疗工具的实用性受到限制。因此,仍然需要新型重组酶,其可识别多种DNA附着位点并实现此类酶的治疗潜力。

发明内容

[0009] 本公开提供了大丝氨酸重组酶变体,其与野生型相比具有改变的DNA靶标特异性。在一个方面,本公开提供了Bxb1重组酶的非天然存在的变体,其相对于野生型Bxb1重组酶(例如SEQ ID NO:1)具有改变的DNA靶标特异性,包含ZD或RD内的一个或多个氨基酸突变(例如取代、缺失或插入)。

[0010] 在一些实施方案中,一个或多个氨基酸突变发生在位置147、148、149、154、155、156、158、197、198、230、231、232、233、237、257、309、312、314、315、316、318、323、324、325、326和335(根据SEQ ID NO:1编号)中的一者或多者处。在进一步的实施方案中,一个或多个氨基酸突变选自F314A、F314C、F314D、F314E、F314G、F314H、F314I、F314L、F314N、F314Q、F314S、F314T、F314V、F314W、F314Y、A315F、A315G、A315H、A315I、A315M、A315N、A315S、A315T、A315W、A315Y、G316A、G316C、G316D、G316E、G316F、G316H、G316I、G316K、G316L、G316M、G316P、G316Q、G316R、G316S、G316T、G316V、G316W、G316Y、G318I、G318K、G318R、G318W、R323G、R323K、R325D、R325E、R325K、R325L、R325M、R325N、R325Q、R325S、R325W、F147A、F147K、F147R、N148Q、N148T、L158D、L158N、L158S、L158T、L158W、P197H、P197R、P197T、S231F、S231G、S231H、S231K、S231R、S231V、S231Y、T233F、T233H、T233K、T233R、T233W、T233Y、R237A、R237C、R237D、R237E、R237F、R237G、R237H、R237I、R237K、R237L、R237M、R237N、R237P、R237Q、R237S、R237T、R237V、R237W、R237Y和D257K。

[0011] 在一些实施方案中,相对于野生型Bxb1重组酶,Bxb1重组酶变体介导真核基因组中内源性靶位点处增加的转基因插入。

[0012] 在其它方面,本公开提供了φC31整合酶、Pa557重组酶或Pa570重组酶的非天然存在的变体,其包含一个或多个相对于野生型对应物产生改变的DNA靶标特异性的氨基酸突变(例如取代、缺失或插入)。

[0013] 在一些实施方案中,φC31整合酶变体在位置273、275、279、375、376、377、379和386(根据SEQ ID NO:77编号)中的一者或多者处包含一个或多个氨基酸突变。在进一步的实施方案中,一个或多个氨基酸突变选自D273、A275、R279、K375、R376、G377、E379和R386。

[0014] 在一些实施方案中,Pa557重组酶变体在位置236、238、242、327、328、329、331和338(根据SEQ ID NO:78编号)中的一者或多者处包含一个或多个氨基酸突变。在进一步的实施方案中,一个或多个氨基酸突变选自G236、A238、A242、R327、T328、G329、G331和R338。

[0015] 在一些实施方案中,Pa570重组酶变体在位置243、245、249、333、334、335、337和344(根据SEQ ID NO:79编号)中的一者或多者处包含一个或多个氨基酸突变。在进一步的实施方案中,一个或多个氨基酸突变选自E243、S245、K249、I333、N334、P335、I337和Q344。

[0016] 在一些实施方案中,DNA靶标对于基因的特定等位基因具有特异性。

[0017] 在一些实施方案中,DNA靶标位于细胞,诸如真核细胞,例如哺乳动物细胞(例如人类细胞)中。

[0018] 本文还提供了编码本文中的重组酶变体的核酸分子和包含编码序列的表达载体。载体可以是例如质粒或病毒载体(例如腺相关病毒载体、腺病毒载体或慢病毒载体)。

[0019] 本公开还提供了一种用于编辑细胞中的DNA的系统,其包含本文中的重组酶变体、核酸分子或载体。在一些实施方案中,所述系统还包含供体DNA,诸如环化DNA或线性DNA。在进一步的实施方案中,供体DNA包含选自以下的序列:SED ID NO:80-81和134-136以及表3、4和5中发现的序列。供体DNA可通过例如质粒或病毒载体来递送。

[0020] 本文中的系统所进行的编辑可包括将DNA整合至细胞基因组中、细胞基因组中DNA的切除或倒位、或细胞基因组中的染色体易位。在一些实施方案中,编辑发生在包含选自以下的序列的基因组区域处:SED ID NO:80-81和134-136以及在表3、4和5中发现的序列。

[0021] 在另一方面,本公开提供了一种编辑细胞基因组的方法,所述方法包括向细胞提

供本文中的基因编辑系统。在一些实施方案中,编辑多于一个基因组区域。编辑方法可导致从基因组中切除DNA、使基因组中的DNA倒位、染色体交换、重组酶介导的盒式交换(RMCE)和/或将供体DNA整合至基因组中。

[0022] 本文还提供了包含本系统的细胞、通过本方法编辑的细胞或其后代细胞。细胞可以是真核细胞(例如哺乳动物细胞,诸如人类细胞)。

[0023] 在另一方面,本公开提供了治疗有需要的受试者的疾病的方法,所述方法包括向受试者施用编辑的细胞。还包括用于治疗有需要的受试者的疾病的细胞,以及所述细胞在制造用于治疗有需要的受试者的疾病的药物中的用途。

[0024] 本发明的其它特征、目标和优点将在下面的详细描述中显而易见。然而,应理解,虽然详细描述指出了本发明的实施方案和方面,但其仅以说明的方式给出,而不是限制。另外,根据详细描述,本发明范围内的各种改变和修改对于本领域技术人员而言是显而易见的。

附图说明

[0025] 图1是示意图,示出(A)Bxb1重组酶的结构和(B)其作为结合的attB和attP位点上的四聚体的作用机制(B.NTD:催化N端域(NTD);αE:保守的α螺旋;CTD:C端域;RD:重组酶域;CC:卷曲螺旋基序;和ZD:锌带域(ZD))。(A)中蛋白质结构下方的数字表示以下各者距蛋白质N端的大致距离:NTD的C端边界(约150个氨基酸)、CTD的N端边界(300个氨基酸)和CTD的C端(600个氨基酸)。

[0026] 图2是示出attB和attP序列的核苷酸编号系统的示意图,包括由Bxb1重组酶的ZD和RD域结合的核苷酸。所示的attP序列与野生型序列(SEQ ID NO:3)相同,而attB序列被编辑为与野生型attB位点(SEQ ID NO:80)相比更加对称。

[0027] 图3A-E是示出重组酶介导的(A)整合、(B)切除、(C)倒位、(D)染色体交换和(E)盒式交换的一组示意图。对于重组酶介导的盒式交换(RMCE),供体DNA可以是环状的(如图所示)或线性的(未示出)。

[0028] 图4是示出基于质粒的重组测定与NGS结合使用以确定突变型Bxb1重组酶变体的DNA序列特异性的示意图。一个质粒以浅灰色显示,并且另一个以黑色显示。白色和深灰色框分别表示attB和attP位点。P1和P2是指引物结合位置。底部显示的是成功整合事件产生的产品。

[0029] 图5是示出人类K562细胞中的染色体重组测定的示意图。Bxb1促进供体质粒靶向整合(TI)至染色体Bxb1 attB伪位点(与天然Bxb1attB靶位点具有一定同源性的内源性人类序列)。白色和深灰色框分别表示attB和attP位点。F-Primer和R-Primer是指用于基于PCR的NGS测定以量化TI事件的引物结合位置。底部显示的是成功TI事件产生的产品。

[0030] 图6A-C示出Bxb1变体S231F的靶位点偏好的改变。(A)从图4中所示的基于质粒的实验系统获得的Bxb1变体S231F的靶标偏好转变。与位置10处的野生型(WT)Bxb1相比,S231F显示对C和T的靶向性均有所改善。(B)内源性人类Bxb1靶位点s5-8(SEQ ID NO:141和264)、s5-1(SEQ ID NO:137和265)、s5-11(SEQ ID NO:143和266)和s1-41(SEQ ID NO:161和267)显示两条DNA链,中心二核苷酸呈灰色阴影,RD基序带下划线,并且左半位点和右半位点的位置10处的碱基均以粗体显示。WT Bxb1所偏好的A碱基的变化由箭头表示。(C)对于

S231F Bxb1变体和WT Bxb1的50％混合物或100％ WT Bxb1,人类细胞中这些内源性位点处的靶向整合(TI)值(含有靶向整合的内源性等位基因的百分比)。第二列中指示了相对于WT Bxb1的偏好靶标,在靶位点的位置10处的任何变化。对WT Bxb1进行了多次重复。WT Bxb1的数据是所有重复的平均值+/-标准差。最后一列显示了具有S231F变体的TI与具有WT Bxb1的TI的比率。S231F变体在内源位点s5-1、s5-11和s1-41处的TI活性增加与基于质粒的实验系统中观察到的靶标偏好变化以及表1中所示的靶向规则一致。

[0031] 图7是Bxb1变体的列表,所述变体与野生型(WT)Bxb1相比,在内源性人类靶位点处改善了靶向整合(TI)活性,与表1中所示的DNA靶序列的改变一致。在并列实验中,同一靶标处的TI值比WT Bxb1的平均值高出至少3个标准差的变体被认为相对于WT Bxb1有所改善。TI和变体/WT数据如图6C中所示。这些位点上与表1中的改变相匹配的序列改变显示于第二、第三和第四列中,其中第二列显示改变在靶位点中的相关位置,第三列显示第一列中列出的靶位点在左半位点给定位置处的改变,并且第四列显示靶位点在右半位点相关位置处的改变。如果指定半位点的指定位置处的碱基与WT Bxb1的靶标偏好相匹配(例如,位置10处的A、位置9处的C或位置7处的A),则将所述碱基表示为“WT Bxb1”。

[0032] 图8A-C示出Bxb1变体F314G和G316Y的靶位点偏好的改变。(A)从图4中所示的基于质粒的实验系统获得的Bxb1变体F314G在位置19处的靶标偏好特异性转变(左图)和Bxb1变体G316Y在位置21处的靶标偏好特异性转变(右图)。碱基按照图2中所示的方案进行编号。请注意,这些内源性人类靶位点均类似于attB位点。由于attP位点相对于attB序列插入了五个碱基,并且插入的碱基位于位置13-17(如果存在),因此标记为位置18的碱基与标记为位置12的碱基相邻。(B)图4中所示的基于质粒的实验系统中使用的对称attB位点(SEQ ID NO:268和269)与内源性人类Bxb1靶位点s5-16(SEQ ID NO:146和270)和s3-28(SEQ ID NO:163和271)的比较。中心二核苷酸呈灰色阴影,并且每个半位点的ZD基序带下划线。请注意,s5-16位点左半位点的ZD基序与基于质粒的系统中使用的ZD基序差异很大,以至于其可能代表不同的序列基序,并且不能被视为基于质粒的系统中所表征的相同ZD序列基序;因此,其可能不遵循表1中所示的相同靶标偏好规则。位点s5-16具有与右半位点中基于质粒的系统中的ZD基序相似的ZD基序;此位点右半位点中位置19的T以粗体显示并用箭头表示。s3-28的两个半位点均含有ZD基序,所述基序与基于质粒的系统中使用的ZD基序类似,且因此两个半位点中的ZD基序均带下划线。右半位点位置21处的G以粗体显示并用箭头表示。(C)使用Bxb1变体F314G和野生型(WT)Bxb1的50％混合物与100％ WT Bxb1的人类细胞中内源性人类靶位点s5-16处的靶向整合(TI)数据的比较(上图),以及使用Bxb1变体G316Y和WT Bxb1的50％混合物与100％ WT Bxb1的人类细胞中内源性人类靶位点s3-28处的TI数据的比较(下图)。靶标改变和TI数据如图6C中所示。

[0033] 图9是在指定内源性人类靶位点处与野生型(WT)Bxb1相比改善靶向整合(TI)活性的Bxb1变体列表。与WT Bxb1的靶标偏好相匹配的碱基(例如位置19处的G)用“WT Bxb1”表示。具有ZD域的半位点与质粒系统中表征的ZD基序差异很大,以至于其可能代表不同的序列基序,在相关半位点的条目中用空白表示。数据如图6C、7和8C中所示。

[0034] 图10是在指定内源性人类靶位点处与野生型(WT)Bxb1相比改善靶向整合(TI)活性的Bxb1变体列表,所述变体在一个或两个半位点的ZD基序中的位置21、22、23或24处发生改变。靶位点改变和数据如图9中所示。WT Bxb1的靶序列偏好是位置21和22处的T,以及位

置23和24处的G。

[0035] 图11显示的数据表明,D257K Bxb1变体在多种不同的内源性人类靶位点处的靶向整合(TI)活性有所改善。与野生型(WT)Bxb1相比,D257K提高了TI活性,无论是作为与WT Bxb1的50%混合物还是作为不含任何WT Bxb1的100% D257K变体。在这些实例中,100% D257K变体比50% D257K或100% WT Bxb1具有更高的活性。因此,D257K似乎代表一种Bxb1变体,与WT Bxb1的偏好靶位点相比,其可在大多数或所有内源性靶位点处增加活性,而不管确切的靶位点改变。

具体实施方式

[0036] 本公开提供了大丝氨酸重组酶变体,诸如Bxb1重组酶变体和直系同源重组酶变体,以及包含所述变体的用于基因编辑的系统。与野生型相比,这些重组酶变体已经改变了DNA靶序列,并且可用来靶向感兴趣的生物(包括人类)基因组内的内源性DNA序列。这些酶可将供体DNA整合至基因组中,或者切除或反转靶基因组序列。因此,其可用于将治疗基因整合至基因组中,或者修复或去除基因组中的致病基因,以达到治疗效果。

[0037] 包含重组酶变体的本发明基因编辑系统在几个重要方面优于其它基因编辑系统。首先,本发明系统避免了对基因组的不期望改变。最广泛使用的基因编辑系统是CRISPR/Cas(成簇的规律间隔的短回文重复序列(CRISPR)和CRISPR相关蛋白(Cas))系统、锌指核酸酶(ZFN)系统和转录激活因子样效应核酸酶(TALEN)系统。这些系统依赖于核酸酶的活性和DNA修复机制,诸如同源重组和非同源末端连接,所述机制容易出错并且可能引入插入或缺失(插入/缺失)和/或易位。通过同时具有切割和连接功能,本发明的重组酶以高度精确、位点特异性的方式切割DNA并连接其断裂处,避免将有害的插入/缺失引入基因组中。其次,使用本发明的编辑系统,可整合至宿主基因组中的DNA没有固有的大小限制。再次,如果没有称为重组方向因子(RDF)的辅助蛋白,本发明系统产生的重组就不容易逆转;因此,基因编辑是稳定且可遗传的。总之,本发明的重组酶变体可用于在由本发明的重组酶变体特异性识别的位点处稳定且精确地从宿主基因组中去除DNA、将大的合成和/或外源供体DNA整合至宿主基因组中、或倒转宿主基因组的片段。

[0038] I.变体丝氨酸重组酶系统

[0039] 本公开提供了基因编辑系统,其包含大丝氨酸重组酶变体,诸如Bxb1重组酶变体或直系同源重组酶变体,以及任选地在靶位点整合至宿主基因组中的供体DNA。还提供了用于递送任何上述组分的表达构建体,以及使用一种或多种上述组分进行基因编辑的方法。下面将进一步详细描述本基因编辑系统的每种组分。

[0040] A.具有改变的DNA靶序列的大丝氨酸重组酶变体

[0041] 本文中使用的术语“重组酶”是指催化重组的蛋白质。术语“重组”是指靶DNA序列(诸如宿主基因组)中的DNA的切除、倒位、整合、染色体交换、RMCE或重排。术语“大丝氨酸重组酶”和“丝氨酸重组酶”是指诱导双链DNA断裂并利用其活性位点中的丝氨酸残基在重组过程中结合单独的DNA片段的重组酶蛋白家族。

[0042] 本文所用的术语“Bxb1重组酶”是指丝氨酸重组酶,其示例性氨基酸序列如下所示:

[0043] 野生型Bxb1

[0044]

```
          10         20         30         40         50
MRALVVIRLS RVTDATTSPE RQLESCQQLC AQRGWDVVGV AEDLDVSGAV
          60         70         80         90        100
DPFDRKRRPN LARWLAFEEQ PFDVIVAYRV DRLTRSIRHL QQLVHWAEDH
         110        120        130        140        150
KKLVVSATEA HFDTTTPFAA VVIALMGTVA QMELEAIKER NRSAAHFNIR
         160        170        180        190        200
AGKYRGSLPP WGYLPTRVDG EWRLVPDPVQ RERILEVYHR VVDNHEPLHL
         210        220        230        240        250
VAHDLNRRGV LSPKDYFAQL QGREPQGREW SATALKRSMI SEAMLGYATL
         260        270        280        290        300
NGKTVRDDDG APLVRAEPIL TREQLEALRA ELVKTSRAKP AVSTPSLLLR
         310        320        330        340        350
VLFCAVCGEP AYKFAGGGRK HPRYRCRSMG FPKHCGNGTV AMAEWDAFCE
         360        370        380        390        400
EQVLDLLGDA ERLEKVWVAG SDSAVELAEV NAELVDLTSL IGSPAYRAGS
         410        420        430        440        450
PQREALDARI AALAARQEEL EGLEARPSGW EWRETGQRFG DWWREQDTAA
         460        470        480        490        500
KNTWLRSMNV RLTFDVRGGL TRTIDFGDLQ EYEQHLRLGS VVERLHTGMS
```

(SEQ ID NO: 1)

[0045] Bxb1重组酶的域结构和其作用机制在图1中示出。以下序列是由野生型Bxb1重组酶识别的attB和attP位点,其中中心二核苷酸以斜体和粗体表示,单下划线区域由ZD域识别,并且双下划线区域由RD域识别:

[0046] 野生型attB

[0047] GGCTTGTCGACGACGGCG***GT***CTCCGTCGTCAGGATCAT (SEQ ID NO: 2)

[0048] 野生型attP

[0049] GGTTTGTCTGGTCAACCACCGCG***GT***CTCAGTGGTGTACGGTACAAACC (SEQ ID NO: 3)

[0050] 为了简化对具有改变的靶DNA特异性的Bxb1重组酶变体的搜索,可使用更对称型式的attB靶位点,其示例性核苷酸序列如下所示(SEQ ID NO:80;另见图2)。中心二核苷酸为斜体和粗体,单下划线区域由ZD域识别,并且双下划线区域由RD结合域识别:

[0051] 对称attB

[0052] GGTTTGTCGACGACGGCG***GT***CTCCGTCGTCAACAAACC (SEQ ID NO: 80)

[0053] 然而,通过此attB序列鉴定的变体将不局限于识别对称attB序列。

[0054] 本公开提供了具有改变的DNA识别序列的上述SEQ ID NO:1的变体。本文中使用的术语“Bxb1重组酶变体”、“变体Bxb1重组酶”、“Bxb1变体”或“变体Bxb1”是指相对于野生型(WT)Bxb1重组酶具有一个或多个氨基酸突变(例如取代、插入和/或缺失)的Bxb1重组酶,所述突变改变了变体的DNA靶向特异性。本文所用的术语“直系同源重组酶变体”或“直系同源丝氨酸重组酶变体”是指相对于其WT对应物具有一个或多个氨基酸突变的直系同源丝氨酸重组酶,所述突变改变了变体的DNA靶向特异性。

[0055] 丝氨酸重组酶(包括本发明的重组酶变体)包含参与序列特异性DNA识别的催化

NTD和CTD。CTD进一步分为RD和含有CC基序的ZD(图1)。NTD域可大致对应于野生型Bxb1重组酶中的氨基酸1-145(根据SEQ ID NO:1编号)或Bxb1重组酶变体中的功能类似序列,或直系同源丝氨酸重组酶和其变体中的功能类似序列。术语“功能类似序列”是指具有相同或基本相同生物学功能的氨基酸残基或蛋白质域的序列。RD域可大致对应于野生型Bxb1重组酶中的氨基酸140-287(根据SEQ ID NO:1编号)或Bxb1重组酶变体中的功能类似序列,或直系同源丝氨酸重组酶和其变体中的功能类似序列。ZD域可大致对应于野生型Bxb1重组酶中的氨基酸302-500(根据SEQ ID NO:1编号)或Bxb1重组酶变体中的功能类似序列,或直系同源丝氨酸重组酶和其变体中的功能类似序列。卷曲螺旋(CC)基序可大致对应于在野生型Bxb1重组酶的ZD域中发现的基序或Bxb1重组酶变体中的功能类似序列,或直系同源丝氨酸重组酶和其变体中的功能类似序列。

[0056] 在一些实施方案中,所提供的重组酶变体可在C末端包含氨基酸序列GSGSGSHHHHHHGSGPKKKRKV(SEQ ID NO:249),或类似的接头、His标签和/或核定位序列。在一些实施方案中,提供的重组酶变体在N末端或C末端包含核定位信号。

[0057] 在一些实施方案中,本发明的Bxb1重组酶变体可包括ZD域或RD域氨基酸序列内的突变。例如,ZD域中的取代可在氨基酸311-335中的一者或多者处进行。RD域的取代可在氨基酸137-160、195-201、229-239、249-252和284-287中的一者或多者处进行。ZD与RD之间的区域被认为充当这些域之间的接头,因此可在氨基酸288-301中的一者或多者处进行对此接头区域的取代。在一些实施方案中,本发明的重组酶变体可包括以下氨基酸位置处的一个或多个突变:147、148、149、154、155、156、158、197、198、230、231、232、233、237、257、309、312、314、315、316、318、323、324、325、326和335。除非另有说明,否则本公开中的Bxb1重组酶中的氨基酸位置按照SEQ ID NO:1进行编号。

[0058] 在进一步的实施方案中,Bxb1变体ZD域可包含以下一个或多个氨基酸取代:

[0059] F314:取代为A、C、D、E、G、H、I、L、N、Q、S、T、V、W或Y;

[0060] A315:取代为F、G、H、I、M、N、S、T、W或Y;

[0061] G316:取代为A、C、D、E、F、H、I、K、L、M、P、Q、R、S、T、V、W或Y;

[0062] G318:取代为I、K、R或W;

[0063] R323:取代为G或K;和

[0064] R325:取代为D、E、K、L、M、N、Q、S或W。

[0065] 在进一步的实施方案中,Bxb1变体RD域可包含以下任何一个或多个氨基酸突变:

[0066] F147:突变为A、K或R;

[0067] N148:突变为Q或T;

[0068] L158:突变为D、N、S、T或W;

[0069] P197:突变为H、R或T;

[0070] S231:突变为F、G、H、K、R、V或Y;

[0071] T233:突变为F、H、K、R、W或Y;以及

[0072] R237:突变为A、C、D、E、F、G、H、I、K、L、M、N、P、Q、S、T、V、W或Y。

[0073] 在一些实施方案中,本发明的重组酶变体可包括NTD域或CC卷曲基序氨基酸序列内的突变。

[0074] Bxb1重组酶(SEQ ID NO:1)的示例性单个氨基酸取代对重组酶识别位点的影响如

下表1和表2中所示。在这些表中,“WT Nt”是指野生型attB或attP位点中的核苷酸(例如SEQ ID NO:2和3)。表1显示了氨基酸147、148、149、154、155、156、158、197、198、230、231、232、233、237、257、309、312、314、315、316、318、323、324、325、326和335的饱和诱变研究的结果(参见下文实施例3)。在表中,“新Nt”是指Bxb1重组酶突变使野生型Bxb1重组酶的核苷酸偏好发生至少3倍变化或将野生型Bxb1重组酶的偏好核苷酸转移至不同核苷酸的核苷酸。例如,在WT attB或attP位点的第6个核苷酸处,当Bxb1重组酶变体具有F147A、N148T、L158W、T233F或T233Y取代时,所述Bxb1重组酶变体对核苷酸A的结合亲和力与野生型Bxb1重组酶相比至少高3倍。

[0075] 表1.具有改变的DNA靶序列的Bxb1重组酶突变体

[0076]

位置*	WT Nt	新 Nt	突变体
6	C	A	F147A、N148T、L158W、T233F、T233Y
		G	F147A、F147R、N148Q、N148T、P197H
7	A	C	F147K、F147R、N148Q、N148T、L158D、L158N、L158S、L158T、T233F、T233Y
9	C	A	N148T、R237A、R237C、R237F、R237G、R237H、R237I、R237K、R237L、R237M、R237N、R237Q、R237S、R237V、R237W、R237Y
		G	F147A、F147R、N148T、P197H、P197R、P197T、R237A、R237C、R237D、R237E、R237F、R237G、R237H、R237I、R237M、R237N、R237P、R237Q、R237S、R237T、R237V、R237W、R237Y
		T	R237A、R237C、R237H、R237M、R237N、R237S、R237T、R237V
10	A	C	F147A、F147R、N148Q、N148T、S231F、S231G、S231H、S231K、S231R、S231V、S231Y、T233H、T233K、T233W
		G	T233R
		T	N148Q、N148T、P197H、S231F、S231G、S231R、S231V、T233F、T233W、T233Y
18	T	A	R325W
		C	R325K
19	G	A	F314D、F314Q、A315T
		C	F314L、A315F、A315I、A315M、A315T、A315W、A315Y、G318R、R325K

[0077]

位置*	WT Nt	新 Nt	突变体
		T	F314A、F314C、F314G、F314N、F314Q、F314S、F314T、A315F、A315I、A315S、A315W、A315Y、G318R
20	T	A	G318R、R325K、R325W
		C	A315F、A315G、A315W、A315Y、R325D、R325E、R325K、R325L、R325M、R325N、R325Q、R325S
		G	F314D、F314N、F314S、F314T、A315T、G318K、G318R
21	T	A	F314D、F314I、F314L、A315F、A315W、A315Y
		C	F314D、F314E、R323G、R325W
		G	F314A、F314C、F314H、F314N、F314S、F314T、F314V、F314W、G316F、G316H、G316K、G316P、G316Y
22	T	A	F314T、A315F、A315W、A315Y、G316C、G316S、G318I
		G	F314D、F314T、F314Y、A315F、A315H、A315N、A315W、A315Y、G316A、G316H、G316Q、G316R、G316S、R323K
23	G	A	A315F、A315H、A315W、A315Y、G316T、G316V、G316W、G318I、G318K、G318R
		C	A315F、A315H、A315N、A315Y、G316A、G316C、G316D、G316E、G316F、G316H、G316I、G316L、G316M、G316Q、G316S、G316T、G316V、G316W、G316Y
		T	G316C、G316I、G316V、G316H
24	G	C	A315F、A315Y、G316W、G318W

[0078] *重组酶识别位点中与图2对应的核苷酸位置。

[0079] 在一些实施方案中,Bxb1重组酶变体可包含两个或更多个取代,例如上表1中所示的取代。这种变体可能对相对于WT attB或attP识别序列含有两个或更多个核苷酸差异的靶序列具有更高的结合亲和力。例如,包含突变R237K和T233R的Bxb1重组酶变体将对包含位置9处的A和位置10处的G的靶序列具有更高的结合亲和力。一些单一取代将使多个核苷酸位置处的结合亲和力增加。举例来说,包含突变F147R的Bxb1重组酶变体将对包含位置6处的G、位置7处的C、位置9处的G和位置10处的C的靶序列具有更高的结合亲和力。一些取代将使单个位置处的多个核苷酸的结合亲和力增加。例如,包含突变R237V的Bxb1重组酶变体将对包含位置9处的A、G或T的靶序列具有更高的结合亲和力。

[0080] 在某些实施方案中,Bxb1重组酶变体可包含一个或多个取代,例如下表2中所示的取代。表2显示了两个Bxb1重组酶突变,其在靶向DNA结合序列中产生对指定核苷酸位置(位置编号根据图2)处的A、C、G或T的结合亲和力的最大增加。例如,虽然突变F147A、N148T、L158W、T233F和T233Y产生的Bxb1重组酶变体对包含位置6处的A的靶序列的结合亲和力增加,但T233F和N148T产生的亲和力增加最多。带下划线的突变产生的Bxb1重组酶变体识别在指定位置处包含新的优选核苷酸的靶序列。例如,具有T233R突变的Bxb1重组酶变体优先识别具有位置10处的G的序列,而野生型Bxb1重组酶优先识别具有位置10处的A的序列。

[0081] 表2.结合亲和力特别增加的Bxb1重组酶突变体

[0082]

域	位置*	WT Nt	A	C	G	T
RD	6	C	T233F、N148T	**WT Bxb1**	N148T、F147R	
	7	A	**WT Bxb1**	N148T、T233Y		
	9	C	R237G/K	**WT Bxb1**	R237D/E	R237A/T
	10	A	**WT Bxb1**	S231R/Y	T233R	T233Y、S231R
ZD	18	T	R325W	R325K		**WT Bxb1**
	19	G	F314D/Q	R325K、F314L	**WT Bxb1**	F314G/S
	20	T	R325K、G318R	R325Q/M	F314T/D	**WT Bxb1**
	21	T	F314L/I	F314E/D	F314N/V	**WT Bxb1**
	22	T	G316C/S		F314Y、G316R	**WT Bxb1**
	23	G	G316W、A315Y	G316Y/D	**WT Bxb1**	G316C/I
	24	G		G318W、G316W	**WT Bxb1**	G316H

[0083] *重组酶识别位点中与图2对应的核苷酸位置。

[0084] 在本公开的一些方面,还可使用与Bxb1重组酶直系同源的重组酶。在一些实施方案中,直系同源丝氨酸重组酶是φC31整合酶,其包含以下示例性氨基酸序列:

[0085] 野生型φC31

[0086]

```
          10         20         30         40         50
MDTYAGAYDR QSRERENSSA ASPATQRSAN EDKAADLQRE VERDGGRFRF
          60         70         80         90        100
VGHFSEAPGT SAFGTAERPE FERILNECRA GRLNMIIVYD VSRFSRLKVM
         110        120        130        140        150
DAIPIVSELL ALGVTIVSTQ EGVFRQGNVM DLIHLIMRLD ASHKESSLKS
         160        170        180        190        200
AKILDTKNLQ RELGGYVGGK APYGFELVSE TKEITRNGRM VNVVINKLAH
         210        220        230        240        250
STTPLTGPFE FEPDVIRWWW REIKTHKHLP FKPGSQAAIH PGSITGLCKR
         260        270        280        290        300
MDADAVPTRG ETIGKKTASS AWDPATVMRI LRDPRIAGFA AEVIYKKKPD
         310        320        330        340        350
GTPTTKIEGY RIQRDPITLR PVELDCGPII EPAEWYELQA WLDGRGRGKG
         360        370        380        390        400
LSRGQAILSA MDKLYCECGA VMTSKRGEES IKDSYRCRRR KVVDPSAPGQ
         410        420        430        440        450
HEGTCNVSMA ALDKFVAERI FNKIRHAEGD EETLALLWEA ARRFGKLTEA
         460        470        480        490        500
PEKSGERANL VAERADALNA LEELYEDRAA GAYDGPVGRK HFRKQQAALT
         510        520        530        540        550
LRQQGAEERL AELEAAEAPK LPLDQWFPED ADADPTGPKS WWGRASVDDK
         560        570        580        590        600
RVFVGLFVDK IVVTKSTTGR GQGTPIEKRA SITWAKPPTD DDEDDAQDGT

EDVAA (SEQ ID NO: 77)
```

[0087] 在一些实施方案中,直系同源丝氨酸重组酶是Pa557重组酶,其包含以下示例性氨基酸序列:

[0088] 野生型Pa557

[0089]

```
          10         20         30         40         50
MNMHSPTVTT RAALYLRVST ARQAEHDISI PDQKRQGEAY CEQRGFQLVE
          60         70         80         90        100
TYVEPGATAT NDKRPEFQRM IEAGTSKPAP FDIVVVHSFS RFFRDHFEME
         110        120        130        140        150
FYVRKLAKNG VKLVSITQEM GDDPMHQMMR QIMALFDEYQ SKENAKHVLR
         160        170        180        190        200
AMNENARQGF WNGARPPIGY RIVAAEQRGS KTKKKLEIDP LHADTVRLIY
         210        220        230        240        250
RLFLEGDGTR GAMGVKAIAT YLNERRFFTR DGGRWGLAQI HAILTRTTYI
         260        270        280        290        300
GEHRFNTRSH KDREKKPESE IAIMAVPPLI EREIYDAVQA RLKSRNPMVT
         310        320        330        340        350
PARVSSGPTL LTGICFCAKC GGAMTLRTGQ GSTGATYRYY TCSTKARQGK
         360        370        380        390        400
TGCKGRTIPM DKLDHLVADH IGDRLLQPKR LETVLASVID RRQERAERRR
         410        420        430        440        450
EHLAELNRRI TEADQRLGRL FDAIEAGMVD KDDAMAKERM VSLKALRDQA
         460        470        480        490        500
AADAERTQLA LDSSGNQGVS PDMLKGFARK ARERIRLDDG GYRRDHLRAL
         510        520        530        540        550
AQRVEVADDE VRIMGSKSEL LRTLVAASSV ETAAFGVQSS VLKWRTQEDS

NLRPLGS (SEQ ID NO: 78)
```

[0090] 在一些实施方案中,直系同源丝氨酸重组酶是Pa570重组酶,其包含以下示例性氨基酸序列:

[0091] 野生型Pa570

[0092]

```
          10          20          30          40          50
MARLISYLRM STSEQLRGFS LERQRKLIAD FAAKNGLSVE EENTLEDIGR
          60          70          80          90         100
SSFSDDAQQK ELTRFFENLN AGKYEPGDVF ALENIDRLTR RGPVDAILKV
         110         120         130         140         150
NQIISKGLKL AIISGNEQRI IEDVNDVFTI INLSIDASRA NKESKNKSDK
         160         170         180         190         200
GLSNWQEKRN LASTYKIAMT AQAPAWLDTE IFYIFDEEKK KNTKRRKYVL
         210         220         230         240         250
NEEKAEAVRL IFDLYSNGNG ALKIKNILNE RNIPTFKGAP YWEPSIITKI
         260         270         280         290         300
LKNPATFGLY QPKKQGTGKR DLIAAGEPIN DYFPPVITRD LFEQCEHIRE
         310         320         330         340         350
GNSTRKGRKG KLFTNLFTGL LTCSKCGGPV HLINPGIDKR NKVQKSIYYL
         360         370         380         390         400
VCKRAKFTKE CTTKRVRYDD FEIALLKAIQ EINLADILNE NNPLEILVKK
         410         420         430         440         450
QRSKETEINK KRKLIENFQR QFLENDGDLP SFMISQAKDA EISIKELEED
         460         470         480         490         500
QREIASEIAQ LNIYNSNVDN AIEELKENAD YGTRSKINLL FHEIIKNISL
         510         520         530         540         550
DTENQFYTVR FKNGVMRVIT AAGFIATTEE QTQADINAIL QSIEGPRIPR
         560         570
EIATDAEKLI EYLKAREVIE (SEQ ID NO: 79)
```

[0093] 在一些实施方案中,直系同源丝氨酸重组酶是LI整合酶,其包含以下示例性氨基酸序列:

[0094] 野生型LI

[0095]

```
          10          20          30          40          50
MKAAIYIRVS TQEQVENYSI QAQTEKLTAL CRSKDWDVYD TFIDGGYSGS
          60          70          80          90         100
NMNRPALNEM LSKLHEIDAV VVYRLDRLSR SQKDTITLIE EYFLKNNVEF
         110         120         130         140         150
VSLSETLDTS SPFGRAMIGI LSVFAQLERE TIRDRMVMGK IKRIEAGLPL
         160         170         180         190         200
TTAKGRTFGY DVIDTKLYIN EEEAKQLRLI YDIFEEEQSI TFLQKRLKKL
         210         220         230         240         250
GFKVRTYNRY NNWLTNDLYC GYVSYKDKVH VKGIHEPIIS EEQFYRVQEI
         260         270         280         290         300
FSRMGKNPNM NKESASLLNN LVVCSKCGLG FVHRRKDTVS RGKKYHYRYY
         310         320         330         340         350
SCKTYKHTHE LEKCGNKIWR ADKLEELIID RVNNYSFASR NIDKEDELDS
         360         370         380         390         400
LNEKLKIEHA KKKRLFDLYI NGSYEVSELD SMMNDIDAQI NYYEAQIEAN
         410         420         430         440         450
EELKKNKKIQ ENLADLATVD FNSLEFREKQ LYLKSLINKI YIDGEQVTIE

WL (SEQ ID NO: 131)
```

[0096] 在一些实施方案中,直系同源丝氨酸重组酶是A118整合酶,其包含以下示例性氨

基酸序列：

[0097] 野生型A118

```
          10          20          30          40          50
MKAAIYIRVS TQEQVENYSI QAQTEKLTAL CRSKDWDVYD IFIDGGYSGS
          60          70          80          90         100
NMNRPALNEM LSKLHEIDAV VVYRLDRLSR SQRDTITLIE EYFLKNNVEF
         110         120         130         140         150
VSLSETLDTS SPFGRAMIGI LSVFAQLERE TIRDRMVMGK IKRIEAGLPL
         160         170         180         190         200
TTAKGRTFGY DVIDTKLYIN EEEAKQLQLI YDIFEEEQSI TFLQKRLKKL
         210         220         230         240         250
GFKVRTYNRY NNWLTNDLYC GYVSYKDKVH VKGIHEPIIS EEQFYRVQEI
         260         270         280         290         300
FTRMGKNPNM NRDSASLLNN LVVCSKCGLG FVHRRKDTMS RGKKYHYRYY
         310         320         330         340         350
SCKTYKHTHE LEKCGNKIWR ADKLEELIIN RVNNYSFASR NVDKEDELDS
         360         370         380         390         400
LNEKLKIEHA KKKRLFDLYI NGSYEVSELD SMMNDIDAQI NYYESQIEAN
         410         420         430         440         450
EELKKNKKIQ ENLADLATVD FDSLEFREKQ LYLKSLINKI YIDGEQVTIE

WL (SEQ ID NO: 132)
```

[0098]

[0099] 在一些实施方案中,直系同源丝氨酸重组酶是TP901重组酶,其包含以下示例性氨基酸序列：

[0100] 野生型TP901

```
          10          20          30          40          50
MTKKVAIYTR VSTTNQAEEG FSIDEQIDRL TKYAEAMGWQ VSDTYTDAGF
          60          70          80          90         100
SGAKLERPAM QRLINDIENK AFDTVLVYKL DRLSRSVRDT LYLVKDVFTK
         110         120         130         140         150
NKIDFISLNE SIDTSSAMGS LFLTILSAIN EFERENIKER MTMGKLGRAK
         160         170         180         190         200
SGKSMMWTKT AFGYYHNRKT GILEIVPLQA TIVEQIFTDY LSGISLTKLR
         210         220         230         240         250
DKLNESGHIG KDIPWSYRTL RQTLDNPVYC GYIKFKDSLF EGMHKPIIPY
         260         270         280         290         300
ETYLKVQKEL EERQQQTYER NNNPRPFQAK YMLSGMARCG YCGAPLKIVL
         310         320         330         340         350
GHKRKDGSRT MKYHCANRFP RKTKGITVYN DNKKCDSGTY DLSNLENTVI
         360         370         380         390         400
DNLIGFQENN DSLLKIINGN NQPILDTSSF KKQISQIDKK IQKNSDLYLN
         410         420         430         440         450
DFITMDELKD RTDSLQAEKK LLKAKISENK FNDSTDVFEL VKTQLGSIPI
         460         470         480
NELSYDNKKK IVNNLVSKVD VTADNVDIIF KFQLA (SEQ ID NO: 133)
```

[0101]

[0102] 一些上述直系同源重组酶的比对可在Rutherford等人,前述的图3中找到。

[0103] 在一些实施方案中,直系同源重组酶变体可包括ZD或RD内的突变。在一些实施方

案中,直系同源重组酶变体可包括NTD域或CC卷曲基序氨基酸序列内的突变。在一些实施方案中,直系同源重组酶可包括对应于Bxb1重组酶的残基S231、T233、R237、F314、A315、G316、G318和R325的位置处的一个或多个突变。例如,φC31整合酶变体可在氨基酸位置D273、A275、R279、K375、R376、G377、E379和R386(根据SEQ ID NO:77编号)包含一个或多个突变;Pa557重组酶变体可在氨基酸位置G236、A238、A242、R327、T328、G329、G331和R338(根据SEQ ID NO:78编号)包含一个或多个突变;且Pa570重组酶变体可在氨基酸位置E243、S245、K249、I333、N334、P335、I337和Q344(根据SEQ ID NO:79编号)包含一个或多个突变。

[0104] 在一些实施方案中,系统包含不同的Bxb1重组酶变体和/或其它直系同源丝氨酸重组酶变体,和/或编码其的核酸,以及任何相关供体DNA分子的混合物。

[0105] 在一些实施方案中,相对于野生型Bxb1重组酶,Bxb1重组酶变体介导真核基因组中内源性靶位点处增加的转基因插入。在一些实施方案中,相对于野生型Bxb1重组酶,Bxb1重组酶变体使转基因插入增加至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20倍。

[0106] B.重组酶变体识别位点

[0107] 本发明的重组酶变体总体上比其WT对应物更通用,因为这些变体可结合更广泛的DNA识别位点。下表(表3)显示与野生型Bxb1重组酶相比,本文变体中的一者或其混合物可更有效地结合attB样序列。下表中的序列按出现顺序为SEQ ID NO:83至130。在表中,带下划线的核苷酸为中心二核苷酸;“hg38坐标”是指根据基因组参考联盟人类构建(Genome Reference Consortium Human Build)38标准的位置;并且带粗体和斜体的核苷酸表示与以下共有attB序列不匹配(对于位点1-14,这是SEQ ID NO:82(参见实施例);对于位点15-24,这是SEQ ID NO:134;对于位点25-36,这是SEQ ID NO:135;并且对于位点37-48,这是SEQ ID NO:136):

[0108] 5′-GGTTTGTNNACNACNGNNNNNNCNGTNGTNAGGATCNN(SEQ ID NO:134)

[0109] 5′-NNGATCCTGACGACNGNNNNNNCNGTNGTNNACAAACC(SEQ ID NO:135)

[0110] 5′-NNGATCCTGACGACNGNNNNNNCNGTNGTNAGGATCNN(SEQ ID NO:136)

[0111] 表3.野生型Bxb1重组酶的内源性人类结合位点

[0112]

SEQ ID NO	#	hg38 坐标		内源性人类序列
83	1	chr18	23228819	***T***G***C***TTGTTG***C***CTACAGCCTCTGCAG***A***AG TTCACAAAC***T***
84	2	chr21	39796718	GG***G***TTGTTGACTACAGATTGCACT***A***TT***T*** TGCA***G***AAACC
85	3	chr5	130870404	***TAG***TTGTAGACCACAGAAATTTCTG***C***T GTTCA***G***AAAC***T***
86	4	chr10	114431581	G***T***TTTGTGGACAA***T***GGTTTCCTCTG***C***TG TAAA***G***AAAC***G***
87	5	chr4	78331600	GG***G***TTG***G***TG***G***CAA***T***GGAGATCTCAGTG GTGAA***A***AA***T***C***A***
88	6	chr3	21564386	***A***G***C***TTGTTTACA***TG***TGAATTTTCTGTAG TAAACAA***C***C***A***
89	7	chr17	37733272	***T***GTTTGTAC***T***CTACCGAAGATTCC***A***TAG TTAACAA***C***C***A***
90	8	chr7	37683631	***TA***TTT***C***TTGACAA***T***TGGCCTCTCTGTAG TTT***G***CAA***C***C***T***
91	9	chr9	94861248	GGTTTG***A***GG***G***CA***G***CAGCAGCAGCTG***C***T GTCA***C***CAAAC***A***
92	10	chrX	50461727	G***T***TTTGTTCACT***GT***AGAATCCTCAGTGG TTA***G***CA***C***AA***T***
93	11	chr1	203596869	***CTC***TTG***A***TGACT***G***CAGAGTATTCC***A***TTG TTGACAAA***T***C
94	12	chr5	35969016	***T***GTTT***C***TCAACTACAGCATGAGCTGTG GTTGACA***GTAA***
95	13	chr1	183975066	G***A***TTTG***G***ATACA***C***CTGAGATGACAGTT GTCAA***G***AAA***T***C
96	14	chr6	120401383	***AT***TT***GT***TTAACTACTGAGTTTTCAGTAG TTAA***A***AA***GC T***
97	15	chr2	131113644	GG***C***TT***T***TTCA***T***TACAGAACACCCAGTTG TCAG***T***ATCAG
98	16	chr18	68095661	***TTA***TTGTTGACAACTGATATTTCC***AC***GG TGAG***A***ATCCC
99	17	chr10	95438042	GG***G***TT***T***TTAA***A***CACTGACTTGCCTG***G***TG TGAGGATCCA
100	18	chr11	93987223	***AA***TTTG***C***TGACAACAGGTGCTTCTGTGG TTAGGA***GT***GG
101	19	chr18	38351472	GG***C***TT***T***TAGACT***GA***TGAGTGATCAGTC***T*** TCAGGATCCC
102	20	chr6	84802299	***ATAG***TGTTTACC***T***CAGAAAGATCTGTG GTAAGGA***C***CAA
103	21	chrX	54472830	GGT***G***TG***A***GGACAA***G***AGAGTGAGCAGT AGTCA***A***GA***G***CTC

[0113]

SEQ ID NO	#	hg38 坐标		内源性人类序列
104	22	chr3	138022040	***T***GT***A***TGTGGACCA***T***AGGAAAAGCTG***G***A GTCAGGA***G***CTC
105	23	chr7	125717751	***T***GTTTGTGGA***G***GA***G***TGATGGCTCAGTG GTA***T***GGATCTA
106	24	chr6	29957229	G***C***TTT***T***TTGACC***T***CGGCCTCCGCT***CA***GG TCAGGA***C***CAG
107	25	chr2	118918350	CTGAT***A***CTT***G***CAACTGAACACACTGTA GTCCACA***C***ACC
108	26	chr1	218729444	GAGATC***T***TG***G***CAA***T***AGGAACCTCAG***C***T GTCA***G***CAA***G***C***A***
109	27	chr12	82686460	GAG***T***T***TG***TGACAACCGAAGCTGCAG***C***A GTCG***C***CAAACC
110	28	chr1	12564817	TTG***C***T***T***CTCACC***GT***AGATGACCCT***C***TTG TCCACAAACC
111	29	chr14	33253493	CAG***C***TCCTCACAACTGATTTCCCA***A***TGG TAA***T***CAA***G***C***T***
112	30	chr5	168835161	GAGATCCTA***G***CCACAGGCGGGCCC***AG***G GTTG***G***CAAACC
113	31	chr1	205211683	TGG***C***TCC***C***AACC***T***CCGTCACTGCCGTTG ***C***AAACAAACC
114	32	chr17	51831373	AG***AT***TCCTGACTACAGTGGATGCTGTT GTTT***G***CAA***TT***C
115	33	chr7	73438826	CAG***C***TCCTGA***A***CACAGCAGTCCCAG***G***T GTTGA***A***AA***G***CC
116	34	chr8	138569746	TTG***G***TC***T***TTACAACTGGGGCTACTGTGG TGGA***T***AA***T***C***G***
117	35	chr7	26280167	CAG***G***TCCTGACC***T***CAGAGCCAGCTG***G***A ***A***TGAA***A***AAACC
118	36	chr18	64064874	CAGAT***G***CTC***T***CCACAGGGTCTGCTG***C***A GTTTACAA***T***CC
119	37	chr15	26397212	ATGATCCTCACTACAGCTATGGCA***A***TT GTA***TT***GATCCT
120	38	chr11	121591710	GTG***C***TCCTAACA***G***CAGGCGGAACAG***A***G GTGAGGATCCT
121	39	chr1	84080800	AAGATCCTGACC***T***CCGAAGAAGCTG***G***A GTTG***A***GATCAT
122	40	chr6	124844046	ATG***G***TCCCA***G***CTACTGGGGAGGCTG***A***G GTGAGGATCGC
123	41	chr3	194380608	TTG***C***TCCTCACA***G***CAGAGGACTCTG***A***A G***A***CAGGA***G***CAT
124	42	chr21	38902753	TGGATC***T***TCACTA***T***GGCACTCCCA***T***TGG TA***T***GGATCTC

[0114]

SEQ ID NO	#	hg38 坐标		内源性人类序列
125	43	chr18	33954365	CAGATC***TA***TACG***G***CTGACTACACAGTG GTGAG***A***A***C***CAT
126	44	chr14	42626161	ATGATCCTCACA***G***CAGATATATCTGTT***C*** TCA***A***GAT***T***TA
127	45	chr10	27365656	ATGATC***T***TGACAACAGGGAGTGCTG***G***G ***T***TC***T***GGA***A***CAT
128	46	chr4	9830819	CTGAT***G***CTGACCACAGCAAATGCTG***C***C ***AA***CAGGATCCC
129	47	chr12	69787548	ATGATCCTAACT***CT***TGCTTTATCT***AC***TG TGA***AA***ATCAT
130	48	chr11	97082317	TTG***G***TCCTG***GA***T***C***CTGAGGTTGCTGTGG TGA***A***GATCCC

[0115] 野生型Bxb1重组酶或其直系同源物可如上文所述通过例如采用表1和表2中所示的氨基酸取代进行工程化,使得工程化蛋白质现在可以高亲和力与人类细胞中的这些内源性基因组靶位点结合。

[0116] C.具有工程化attB或attP序列的供体DNA。

[0117] 如果供体和宿主DNA包含野生型或变体attB和/或attP序列,则所提供的重组酶变体可将供体DNA(例如环化供体DNA)整合至宿主基因组中。本公开提供了包含由所提供的重组酶变体识别的变体attB和attP序列的核酸分子。

[0118] 在一些实施方案中,本文所述的attB序列包括SEQ ID NO:83-130的序列中的任一者,如表3中所示。在一些实施方案中,本文所述的attB序列包括SEQ ID NO:4-40的序列中的任一者,如表4中所示。在一些实施方案中,本文所述的attP序列包括SEQ ID NO:41-76的序列中的任一者,如表5中所示。在一些实施方案中,attB或attP序列可包括选自SEQ ID NO:80-82和134-136的序列。

[0119] D.重组事件

[0120] Bxb1重组酶可产生一系列二聚体,其识别并特异性结合DNA识别attB和attP位点。当attB和attP位点存在于同一细胞中时,一对结合attB和attP位点的Bxb1重组酶二聚体形成四聚体,将含有attB和attP位点的DNA片段聚集在一起并启动DNA重组。当DNA片段位于同一DNA分子(例如染色体)上时,这种重组事件可导致DNA倒位或缺失。当DNA片段位于不同的DNA分子上时,这种重组事件可能导致DNA整合或盒式交换。本发明的重组酶变体含有突变,使其能够催化变体attB与attP序列之间的重组,大大扩展了可识别内源位点的库并提高了酶的多功能性。

[0121] 本发明的重组酶变体可产生感兴趣的重组事件,诸如整合、倒位、切除、染色体交换和RMCE。图3A示出了由本公开的重组酶变体介导的整合。整合需要存在具有与靶DNA附着位点相容的附着位点的供体DNA。图3B示出了由本公开的重组酶变体介导的切除。切除需要同一DNA分子上两个方向相似的互补附着位点。图3C示出了由本公开的重组酶变体介导的倒位。倒位需要同一DNA分子上两个方向相反的互补附着位点。图3D示出了由本公开的重组酶变体介导的染色体交换。染色体交换需要两个互补的附着位点,其在两个独立的线性DNA

分子上具有相似的方向。图3E示出了由本公开的重组酶变体介导的RMCE。RMCE要求供体和靶DNA分子各自含有两个不交叉相容的互补附着位点。例如,供体DNA分子含有基因X,其两侧有两个不同的attB位点,而靶DNA分子含有基因Y,其两侧有两个不同的attP位点,其中基因X和Y上游的相容附着位点是互补的,并且基因X和Y下游的相容附着位点是互补的,但上游和下游位点不交叉相容。例如,可通过在上游和下游附着位点使用不同的中心二核苷酸来避免交叉相容性。此系统允许基因X和Y在本公开的重组酶变体存在下进行交换。只有在适当的RDF存在的情况下,丝氨酸重组酶才能与attR和attL位点结合并介导反向重组事件。

[0122] II.变体重组酶系统的递送

[0123] 本公开的Bxb1重组酶变体或直系同源重组酶变体可通过多种方法(例如电穿孔、脂质纳米颗粒、阳离子或阴离子脂质体、或核定位信号(例如与脂质体组合))作为蛋白质引入靶细胞中。在一些实施方案中,提供的重组酶变体通过编码其的核酸分子(例如DNA质粒或mRNA)引入靶细胞中。核酸分子可位于核酸表达载体中,所述载体可包含允许用于提供的重组酶变体的编码序列表达的表达控制序列,诸如启动子、增强子、转录信号序列和转录终止序列。本文所述的"系统递送"可指包含本文所述的Bxb1重组酶变体或直系同源重组酶变体的系统的递送,或编码提供的重组酶变体的所述系统的核酸分子或包含所述核酸分子的载体或表达构建体的递送。

[0124] 在一些实施方案中,载体上用于指导提供的重组酶变体的表达的启动子是组成型活性启动子或诱导型启动子。合适的启动子包括但不限于Rous肉瘤病毒(RSV)长末端重复(LTR)启动子(任选地带有RSV增强子)、巨细胞病毒(CMV)启动子(任选地带有CMV增强子)、CMV立即早期启动子、猿猴病毒40(SV40)启动子、二氢叶酸还原酶(DHFR)启动子、β-肌动蛋白启动子、磷酸甘油酸激酶(PGK)启动子、EF1α启动子、莫洛尼鼠白血病病毒(MoMLV)LTR、肌酸激酶基(CK6)启动子、转甲状腺素蛋白启动子(TTR)、胸苷激酶(TK)启动子、四环素反应性启动子(TRE)、乙型肝炎病毒(HBV)启动子、人α1-抗胰蛋白酶(hAAT)启动子、嵌合肝特异性启动子(LSP)、E2因子(E2F)启动子、人端粒酶逆转录酶(hTERT)启动子、CMV增强子/鸡β-肌动蛋白/兔β-珠蛋白启动子(CAG启动子;Niwa等人,Gene(1991)108(2):193-9)和RU-486反应性启动子。另外,启动子可包含一个或多个自我调节元件,由此提供的重组酶变体可结合并抑制其自身的表达水平至预设的阈值。

[0125] 可采用将核苷酸序列引入细胞中的任何方法,包括但不限于电穿孔、磷酸钙沉淀、显微注射、阳离子或阴离子脂质体、与核定位信号组合的脂质体、天然存在的脂质体(例如外泌体)或病毒转导。在某些实施方案中,核苷酸序列为mRNA的形式并通过电穿孔递送至细胞中。

[0126] 为了体内递送表达载体,可以使用病毒转导。本领域技术人员可以采用本领域已知的多种病毒载体用于本公开,例如痘苗载体、腺病毒载体、慢病毒载体、痘病毒载体、腺相关病毒(AAV)载体、逆转录病毒载体和杂交病毒载体。在一些实施方案中,本文使用的病毒载体是重组AAV(rAAV)载体。可以使用任何合适的AAV血清型。例如,AAV可以是AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV8.2、AAV9、AAV.PHP.B、AAV.PHP.eB或AAVrh10,或者是新型血清型或假型诸如AAV2/8、AAV2/5、AAV2/6、AAV2/9或AAV2/6/9。在一些实施方案中,表达载体是AAV病毒载体并且通过重组AAV病毒体引入靶人类细胞中,所述重组AAV病毒体的基因组包含所述构建体,包括在两端具有AAV反向末端重复(ITR)序列,以

允许在生产系统(诸如昆虫细胞/杆状病毒生产系统或哺乳动物细胞生产系统)中产生AAV病毒体。可以对AAV进行工程化,使得其衣壳蛋白在人体中具有降低的免疫原性或增强的转导能力。可以使用本领域已知的方法来产生本文所述的病毒载体。可以使用任何合适的允许细胞类型或包装细胞类型来产生病毒颗粒。例如,哺乳动物(例如293)或昆虫(例如sf9)细胞可用作包装细胞系。

[0127] 任何类型的细胞均可被靶向用于本文所述的基因编辑方法。例如,细胞可以是真核细胞或原核细胞。在一些实施方案中,细胞是哺乳动物(例如人类)细胞或植物细胞。人类细胞可以包括,例如,T细胞、自然杀伤(NK)细胞、NK T细胞、α-βT细胞、γ-δT细胞、细胞毒性T淋巴细胞(CTL)、调节性T细胞、B细胞、人胚胎干细胞、肿瘤浸润淋巴细胞(TIL)或可分化为淋巴样细胞的多能干细胞(例如,诱导性多能干细胞(iPSC))。在一些实施方案中,系统可用于在多能干细胞分化为多种细胞类型之前对其进行修饰。例如,淋巴样细胞前体可在分化为淋巴样细胞类型(诸如调节性T细胞、效应T细胞、自然杀伤细胞等)之前进行修饰。特别地,包含多于一种Bxb1重组酶变体或直系同源重组酶变体的本公开的系统可用于制备同时具有多个整合、切除或倒位基因的细胞,包括多能细胞。在一些实施方案中,含有多于一种所提供的重组酶变体的系统可用于制备例如同种异体T细胞。

[0128] 对于农业应用,还考虑将蛋白质或核酸分子引入植物细胞中的任何方法,诸如农杆菌介导的T-DNA递送。

[0129] III.治疗应用

[0130] 本公开提供了对细胞DNA中的基因或DNA序列进行整合、切除和倒位的方法,所述方法包括将本文所述的Bxb1重组酶变体或直系同源重组酶变体系统递送至细胞(例如来自患者)。细胞可在患者体内(体内治疗),或者可对从患者中去除的细胞进行本文所述的方法,并且然后将经编辑的细胞递送至患者(离体治疗)。在一些实施方案中,在用于治疗之前,对细胞进行进一步离体操纵。术语"治疗"涵盖缓解症状、预防症状发作、减缓疾病进展、改善生活质量和提高生存率。在一些实施方案中,通过本文所述方法治疗的患者是哺乳动物,例如人类。

[0131] 在一些实施方案中,本公开的方法用于插入或切除与疾病相关的基因或调控序列,以恢复正常的基因表达或活性。在一些实施方案中,本公开的方法可靶向基因的特定等位基因,例如野生型或突变的等位基因。在某些实施方案中,等位基因可能与癌症相关。

[0132] 在一些实施方案中,患者患有癌症。在某些实施方案中,在基因编辑之前或之后对来自患者的细胞进行进一步修饰,以提供对化疗剂的抗性。接着可用化疗剂治疗患者,在一些实施方案中,其可使得经编辑的细胞比未编辑的细胞存活率更高。

[0133] 在一些实施方案中,患者患有自身免疫性病症。

[0134] 在一些实施方案中,患者患有常染色体显性疾病,诸如常染色体显性多囊肾病。

[0135] 在一些实施方案中,患者患有神经发育病症。

[0136] 在一些实施方案中,患者患有线粒体病症。

[0137] 在一些实施方案中,患者患有镰状细胞病、血友病(例如血友病A、B或C)、囊性纤维化、苯丙酮尿症、泰-萨克斯病(Tay-Sachs)、朊病毒病、色盲、溶酶体贮积病(例如法布里氏病(Fabry disease))、弗里德赖希共济失调(Friedreich's ataxia)、前列腺癌、β地中海贫血、亨廷顿氏病(Huntington's disease)、肾移植、炎症性肠病、多发性硬化症、肌萎缩侧索

硬化症或额颞叶痴呆。

[0138] 本公开还提供了一种药物组合物,其包含本文所述的基因编辑系统的元件,诸如Bxb1重组酶变体或直系同源重组酶变体,或编码所述元件的核苷酸序列(例如,在本文所述的病毒或非病毒载体中)。所述药物组合物还可包含药学上可接受的载剂,诸如水、盐水(例如磷酸盐缓冲盐水)、葡萄糖、甘油、蔗糖、乳糖、明胶、葡聚糖、白蛋白或果胶。此外,组合物可包含辅助物质,诸如润湿剂或乳化剂、pH缓冲剂、稳定剂或增强药物组合物有效性的其它试剂。药物组合物可含有递送媒介物,诸如脂质体、纳米胶囊、微粒、微球、脂质颗粒和囊泡。

[0139] 在一些实施方案中,本文所述的提供的重组酶变体可用于本文所述的治疗方法中,可用于本文所述的治疗中,或可用于制造用于本文所述的治疗的药物中。

[0140] 除非本文另有定义,否则与本公开结合使用的科学和技术术语应具有本领域普通技术人员通常理解的含义。下面描述了示例性方法和材料,但与本文所述的方法和材料类似或等效的方法和材料也可用于本公开的实践或测试中。如有冲突,则以本说明书(包括定义)为准。一般来说,与本文所述的神经病学、医学、医药化学以及细胞生物学结合使用的命名和其技术是本领域中众所周知和通常使用的。酶促反应和纯化技术根据制造商的说明书进行,如本领域中通常完成的或本文所述。此外,除非上下文另有要求,否则单数术语应包括复数并且复数术语应包括单数。在整个本说明书和实施方案中,词语“具有(have)”和“包含(comprise)”或变形诸如“具有(has)”、“具有(having)”、“包含(comprises)”或“包含(comprising)”将被理解成意指包括所陈述的整数或整数组,但不排除任何其它整数或整数组。本文提及的所有出版物和其它参考文献均以全文引用的方式并入。尽管本文引用了多份文件,但此引用并不构成对任何这些文件构成本领域公知常识的承认。本文中使用的术语“大约”或“约”在应用于一个或多个感兴趣的值时,是指与所述参考值相似的值。在某些实施方案中,所述术语是指在任一方向(大于或小于上)落入规定参考值的10%、9%、8%、7%、6%、5%、4%、3%、2%、1%或更小的范围内的值,除非另有说明或从上下文中明显看出。

[0141] 为了更好地理解本发明,列出了以下实施例。这些实施例仅用于说明的目的,而并不被理解为以任何形式限制本发明的范围。

[0142] 实施例

[0143] 实施例1:attB和attP质粒变体

[0144] 以下是野生型Bxb1重组酶共有靶序列(带下划线的序列5’-GGTTTGT为ZD域结合位点,带下划线的序列5’-ACNACNG为RD域结合位点,且中心二核苷酸位置以粗体显示为XX):

[0145] attP共有序列:

[0146] 5′-GGTTTGTNNNNNNNACNACNGNN**XX**NNCNGTNGTNNNNNNNACAAACC
CCAAACANNNNNNNTGNTGNCNN**XX**NNGNCANCANNNNNNNTGTTTGG-5′ (SEQ ID NO: 81)

[0147] attB共有序列:

[0148] 5′-GGTTTGTNNACNACNGNN**XX**NNCNGTNGTNNACAAACC
CCAAACANNTGNTGNCNN**XX**NNGNCANCANNTGTTTGG-5′ (SEQ ID NO: 82)

[0149] 为了简化对具有改变的靶DNA特异性的Bxb1重组酶变体的搜索,创建了更对称型式的Bxb1 attB靶位点(表4;参见野生型attB位点的SEQ ID NO:2)。这有利于对attB靶位点

的两半产生对称突变。由于野生型attP位点在ZD和RD域结合位点处已经是回文的(参见SEQ ID NO:3),因此无需创建回文型式。与attB质粒变体一样,结合域的单个核苷酸发生突变,以产生每个核苷酸都可用的构建体(表5),如针对以下位置9和10处的突变以粗体所显示(外部和内部下划线序列分别由ZD和RD域结合;中心二核苷酸为斜体)。表中的"SEQ"是指SEQ ID NO。

[0150] 位置10处的突变:

		SEQ ID NO
attB	GGTTTGTCGACGACGGCG*GT*CTCCGTCGTCAACAAACC	80
	GGTTTGTCG**C**CGACGGCG*GT*CTCCGTCG**G**CAACAAACC	252
	GGTTTGTCG**G**CGACGGCG*GT*CTCCGTCG**C**CAACAAACC	253
	GGTTTGTCG**T**CGACGGCG*GT*CTCCGTCG**A**CAACAAACC	254

[0151]

attP	GGTTTGTCTGGTCAACCACCGCGG*GT*CTCAGTGGTGTACGGTACAAACC	3
	GGTTTGTCTGGTCA**C**CCACCGCGG*GT*CTCAGTGG**G**GTACGGTACAAACC	255
	GGTTTGTCTGGTCA**G**CCACCGCGG*GT*CTCAGTGG**C**GTACGGTACAAACC	256
	GGTTTGTCTGGTCA**T**CCACCGCGG*GT*CTCAGTGG**A**GTACGGTACAAACC	257

[0152] 位置9处的突变:

attB	GGTTTGTCGACGACGGCG*GT*CTCCGTCGTCAACAAACC	80
	GGTTTGTCGA**A**GACGGCG*GT*CTCCGTC**T**TCAACAAACC	258
	GGTTTGTCGA**G**GACGGCG*GT*CTCCGTC**C**TCAACAAACC	259
	GGTTTGTCGA**T**GACGGCG*GT*CTCCGTC**A**TCAACAAACC	260

[0153]

attP	GGTTTGTCTGGTCAACCACCGCGG*GT*CTCAGTGGTGTACGGTACAAACC	3
	GGTTTGTCTGGTCAA**A**CACCGCGG*GT*CTCAGTG**T**TGTACGGTACAAACC	261
	GGTTTGTCTGGTCAA**G**CACCGCGG*GT*CTCAGTG**C**TGTACGGTACAAACC	262
	GGTTTGTCTGGTCAA**T**CACCGCGG*GT*CTCAGTG**A**TGTACGGTACAAACC	263

[0154] 表4.变体attB序列

[0155]

SEQ	描述	序列
4	Bxb1_AttB_回文	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACGACGGCGGTCTCCGTCGTCA ACAAACCCCGGCAGATAAAAGTACCCAGAACCAGA GCCTACAATACCGGCCCTGGGAATATAAG
5	Bxb1_AttB_tGTTTGTCGACGAC	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAtGTTTGTCGACGACGGCGGTCTCCGTCGTCAA CAAACaCCGGCAGATAAAAGTACCCAGAACCAGAGC

[0156]

SEQ	描述	序列
	GGCGGT	CTACAATACCGGCCCTGGGAATATAAG
6	Bxb1_AttB_cGTTTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAcGTTTGTCGACGACGGCGGTCTCCGTCGTCAACAAACgCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
7	Bxb1_AttB_aGTTTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAaGTTTGTCGACGACGGCGGTCTCCGTCGTCAACAAACtCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
8	Bxb1_AttB_GtTTTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGtTTTGTCGACGACGGCGGTCTCCGTCGTCAACAAAaCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
9	Bxb1_AttB_GcTTTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGcTTTGTCGACGACGGCGGTCTCCGTCGTCAACAAAgCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
10	Bxb1_AttB_GaTTTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGaTTTGTCGACGACGGCGGTCTCCGTCGTCAACAAAtCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
11	Bxb1_AttB_GGgTTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGgTTGTCGACGACGGCGGTCTCCGTCGTCAACAAcCCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
12	Bxb1_AttB_GGaTTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGaTTGTCGACGACGGCGGTCTCCGTCGTCAACAAtCCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
13	Bxb1_AttB_GGcTTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGcTTGTCGACGACGGCGGTCTCCGTCGTCAACAAgCCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
14	Bxb1_AttB_GGTgTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTgTGTCGACGACGGCGGTCTCCGTCGTCAACAcACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
15	Bxb1_AttB_GGTaTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTaTGTCGACGACGGCGGTCTCCGTCGTCAACAtACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
16	Bxb1_AttB_GGTcTGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTcTGTCGACGACGGCGGTCTCCGTCGTCAACAgACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
17	Bxb1_AttB_GGTTgGTCGACGA	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTgGTCGACGACGGCGGTCTCCGTCGTCAACcAACCCCGGCAGATAAAAGTACCCAGAACCAGAG

[0157]

SEQ	描述	序列
	CGGCGGT	CCTACAATACCGGCCCTGGGAATATAAG
18	Bxb1_AttB_GGTTaGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTaGTCGACGACGGCGGTCTCCGTCGTCAACtAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
19	Bxb1_AttB_GGTTcGTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTcGTCGACGACGGCGGTCTCCGTCGTCAACgAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
20	Bxb1_AttB_GGTTTtTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTtTCGACGACGGCGGTCTCCGTCGTCAAaAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
21	Bxb1_AttB_GGTTTcTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTcTCGACGACGGCGGTCTCCGTCGTCAAgAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
22	Bxb1_AttB_GGTTTaTCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTaTCGACGACGGCGGTCTCCGTCGTCAAtAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
23	Bxb1_AttB_GGTTTGgCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTGgCGACGACGGCGGTCTCCGTCGTCAcCAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
24	Bxb1_AttB_GGTTTGaCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTGaCGACGACGGCGGTCTCCGTCGTCAtCAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
25	Bxb1_AttB_GGTTTGcCGACGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTGcCGACGACGGCGGTCTCCGTCGTCAgCAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
26	Bxb1_AttB_GGTTTGTCGtCGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTGTCGtCGACGGCGGTCTCCGTCGaCAACAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
27	Bxb1_AttB_GGTTTGTCGgCGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTGTCGgCGACGGCGGTCTCCGTCGcCAACAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
28	Bxb1_AttB_GGTTTGTCGcCGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTGTCGcCGACGGCGGTCTCCGTCGgCAACAAACCCCGGCAGATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
29	Bxb1_AttB_GGTTTGTCGAgGACGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATAGGGGAGGTTTGTCGAgGACGGCGGTCTCCGTCcTCAACAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG

[0158]

SEQ	描述	序列
	CGGCGGT	CCTACAATACCGGCCCTGGGAATATAAG
30	Bxb1_AttB_GGTTTG TCGAtGA CGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGAtGACGGCGGTCTCCGTCaTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
31	Bxb1_AttB_GGTTTG TCGAaGA CGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGAaGACGGCGGTCTCCGTCtTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
32	Bxb1_AttB_GGTTTG TCGACtA CGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACtACGGCGGTCTCCGTaGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
33	Bxb1_AttB_GGTTTG TCGACcA CGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACcACGGCGGTCTCCGTgGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
34	Bxb1_AttB_GGTTTG TCGACaA CGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACaACGGCGGTCTCCGTtGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
35	Bxb1_AttB_GGTTTG TCGACGt CGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACGtCGGCGGTCTCCGaCGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
36	Bxb1_AttB_GGTTTG TCGACGg CGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACGgCGGCGGTCTCCGcCGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
37	Bxb1_AttB_GGTTTG TCGACGc CGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACGcCGGCGGTCTCCGgCGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
38	Bxb1_AttB_GGTTTG TCGACGA gGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACGAgGGCGGTCTCCcTCGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
39	Bxb1_AttB_GGTTTG TCGACGA tGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACGAtGGCGGTCTCCaTCGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG
40	Bxb1_AttB_GGTTTG TCGACGA aGGCGGT	ATGCTTCTCTGCATCCGTCTGCTAAATCCGCGTGATA GGGGAGGTTTGTCGACGAaGGCGGTCTCCtTCGTCAA CAAACCCCGGCAGATAAAAGTACCCAGAACCAGAG CCTACAATACCGGCCCTGGGAATATAAG

[0159] 表5.变体attP序列

[0160]

SEQ	描述	序列
41	AttP_TCCAC	GGTTTGTCTGGTCATCCACCGCGGTCTCAGTGGAGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
42	AttP_CCCAC	GGTTTGTCTGGTCACCCACCGCGGTCTCAGTGGGGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
43	AttP_GCCAG	GGTTTGTCTGGTCAGCCACCGCGGTCTCAGTGGCGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
44	AttP_AACAC	GGTTTGTCTGGTCAAACACCGCGGTCTCAGTGTTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
45	AttP_ATCAC	GGTTTGTCTGGTCAATCACCGCGGTCTCAGTGATGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
46	AttP_AGCAC	GGTTTGTCTGGTCAAGCACCGCGGTCTCAGTGCTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
47	AttP_ACTAC	GGTTTGTCTGGTCAACTACCGCGGTCTCAGTAGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
48	AttP_ACAAC	GGTTTGTCTGGTCAACAACCGCGGTCTCAGTTGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
49	AttP_ACGAC	GGTTTGTCTGGTCAACGACCGCGGTCTCAGTCGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
50	AttP_ACCTC	GGTTTGTCTGGTCAACCTCCGCGGTCTCAGAGGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
51	AttP_ACCCC	GGTTTGTCTGGTCAACCCCCGCGGTCTCAGGGGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
52	AttP_ACCGC	GGTTTGTCTGGTCAACCGCCGCGGTCTCAGCGGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
53	AttP_ACCAG	GGTTTGTCTGGTCAACCAGCGCGGTCTCACTGGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
54	AttP_ACCAT	GGTTTGTCTGGTCAACCATCGCGGTCTCAATGGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
55	AttP_ACCAA	GGTTTGTCTGGTCAACCAACGCGGTCTCATTGGTGTACGGTACAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
56	AttP_CGTTTGT	CGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACGCAATAAAAGTACCCAGAACCAGA

[0161]

SEQ	描述	序列
		GCCTACAATACCGGCCCTGGGAATATAAG
57	AttP_TGTTTGT	TGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACACAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
58	AttP_AGTTTGT	AGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACTCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
59	AttP_GCTTTGT	GCTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAAGCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
60	AttP_GTTTTGT	GTTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAAACCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
61	AttP_GATTTGT	GATTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAATCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
62	AttP_GGATTGT	GGATTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAATCCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
63	AttP_GGCTTGT	GGCTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAGCCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
64	AttP_GGGTTGT	GGGTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAACCCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
65	AttP_GGTCTGT	GGTCTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAGACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
66	AttP_GGTGTGT	GGTGTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACACACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
67	AttP_GGTATGT	GGTATGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACATACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
68	AttP_GGTTAGT	GGTTAGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACTAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
69	AttP_GGTTCGT	GGTTCGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACGAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
70	AttP_GGTTGGT	GGTTGGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACCAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
71	AttP_GGTTTTT	GGTTTTTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTAAAAACCCAATAAAAGTACCCAGAACCAGAGCCTACAATACCGGCCCTGGGAATATAAG
72	AttP_GGTT	GGTTTATCTGGTCAACCACCGCGGTCTCAGTGGTGT

[0162]

SEQ	描述	序列
	TAT	ACGGTATAAACCCAATAAAAGTACCCAGAACCAGA GCCTACAATACCGGCCCTGGGAATATAAG
73	AttP_GGTT TCT	GGTTTCTCTGGTCAACCACCGCGGTCTCAGTGGTGT ACGGTAGAAACCCAATAAAAGTACCCAGAACCAGA GCCTACAATACCGGCCCTGGGAATATAAG
74	AttP_GGTT TGA	GGTTTGACTGGTCAACCACCGCGGTCTCAGTGGTGT ACGGTTCAAACCCAATAAAAGTACCCAGAACCAGA GCCTACAATACCGGCCCTGGGAATATAAG
75	AttP_GGTT TGC	GGTTTGCCTGGTCAACCACCGCGGTCTCAGTGGTGT ACGGTGCAAACCCAATAAAAGTACCCAGAACCAGA GCCTACAATACCGGCCCTGGGAATATAAG
76	AttP_GGTT TGG	GGTTTGGCTGGTCAACCACCGCGGTCTCAGTGGTGT ACGGTCCAAACCCAATAAAAGTACCCAGAACCAGA GCCTACAATACCGGCCCTGGGAATATAAG

[0163] 实施例2:初始丝氨酸扫描筛选

[0164] 具有丝氨酸扫描突变的单个构建体是由表达由CMV启动子驱动的Bxb1重组酶的pVax载体生成的。Bxb1重组酶构建体在C末端包含氨基酸序列GSGSGSHHHHHHGSGPKKKRKV(SEQ ID NO:249)。在非表达质粒中生成与SEQ ID NO:2-76匹配的单个构建体。将构建体组汇集在一起,使得特定位点处的所有四个碱基的浓度相等(例如,将具有attP SEQ ID NO:3、41、42和43的质粒汇集在一起并将具有SEQ ID NO:3、44、45和46的质粒汇集在一起;将具有attB SEQ ID NO:4、5、6和7的质粒汇集在一起并将具有SEQ ID NO:4、8、9和10的质粒汇集在一起)。

[0165] K562细胞按照制造商的方案进行转染。用200ng的Bxb1表达质粒、总共20ng的attB质粒和总共1600ng的attP质粒转染2E5细胞。attB质粒在特定位点各有5ng汇集的核苷酸突变体。例如,当测试碱基24(图2)时,将SEQ ID NO:4、5、6和7汇集在一起,使得在总共20ng的attB质粒中各递送5ng。在同一个孔中,将SEQ ID NO:3、56、57和58汇集在一起,使得在总共1600ng的attP质粒中各递送400ng。

[0166] 具有突变M244S、L245S、G246S、Y247S、A248S、L250S、N251S、G252S、K253S、T254S、V255S、R256S、D257S、D258S、D259S、G260S、A262S、K284S、T285S、S286A、R287S、A288S、K289S、P290S、A291S、V292S、S293A、T294S、P295S、S296A、L297S、L298S、L299S、R300S、V301S、A311S、Y312S、K313S、F314S、A315S、G316S、G317S、G318S、R319S、K320S、H321S、P322S、R323S、Y324S、R325S、C326S、R327S、S328A、M329S、G330S、F331S、P332S、K333S、H334S、C335S、E445S、Q446S、D447S、A449S、A450S、K451S、Y452S、T453S、W454S、L455S、R456S、M458S、N459S和V460S的单个表达质粒在认为被ZD域识别的所有七个核苷酸的文库中进行转染。

[0167] 具有突变I137S、K138S、E139S、R140S、N141S、R142S、S143A、A144S、A145S、H146S、F147S、N148S、I149S、R150S、A151S、G152S、K153S、Y154S、R155S、G156S、S157A、L158S、P159S、P160S、H195S、E196S、P197S、L198S、H199S、L200S、V201S、E229S、W230S、S231A、A232S、T233S、A234S、L235S、K236S、R237S、S238A、M239S、M244S、L245S、G246S、Y247S、A248S、T249S、L250S、N251S、G252S、K253S、T254S、V255S、R256S、D257S、D258S、D259S、

G260S、A262S、K284S、T285S、S286A、R287S、A288S、K289S、P290S、A291S、V292S、S293A、T294S、P295S、S296A、L297S、L298S、L299S、R300S、V301S、A311S、Y312S、K313S、F314S、A315S、G316S、G317S、G318S、R319S、K320S、H321S、P322S、R323S、Y324S、R325S、C326S、R327S、S328A、M329S、G330S、F331S、P332S、K333S、H334S和C335S的单个表达质粒在跨越认为被RD域识别的五个碱基的核苷酸突变体文库中进行转染(参见图2)。

[0168] 在37℃下培育三天后,转染细胞按照制造商说明书使用Lucigen的QuickExtract™(目录号:QE09050)来收获。使用Invitrogen的Accuprimer™Taq DNA聚合酶,高保真度(目录号:12346094)对提取的DNA进行聚合酶链式反应(PCR),如下表6中所示。

[0169] 表6.NGS PCR1设置

[0170]

试剂	1× (µl)	MM (µl)
10× AccuPrime™缓冲液 II	2.5	275
10 µM fw 寡核苷酸	1	110
10 µM 反向寡核苷酸	1	110
AccuPrime™ Taq (HiFi)	0.2	22
H_2O	17.3	1903
gDNA	3	
总计	**25**	**2420**
反应: 110	分装 22.0 µl MM	
	添加 3.0 µl gDNA	

[0171] 使用下表7中所示的引物进行PCR,其中P1和P2分别是指引物1和2。

[0172] 表7.NGS PCR1引物

[0173]

	SEQ ID NO	
P1	250	ACACGACGCTCTTCCGATCTNNNNATGCTTCTCTGCATCCGTCT
P2	251	GACGTGTGCTCTTCCGATCTAGGGAATAGATGACATCAAAGAGAA

[0174] PCR使用下表8中的温度循环方案进行。时间的格式为分:秒。

[0175] 表8.NGS PCR1程序

[0176]

温度	时间	
95℃	5:00	
95℃	0:30	30×
55℃	0:30	
68℃	0:40	
68℃	10:00	
4℃	∞	

[0177] 将下一代测序(NGS)PCR1的产物用作NGS PCR2的模板,如下表9中所示。

[0178] 表9.NGS PCR2设置

[0179]

试剂	1× (μl)	MM (μl)
2× Phusion™ HiFi PCR MM，带 HF 缓冲液	10	1100
10 μM 正向寡核苷酸(N5xx)	1	110
10 μM 反向寡核苷酸(索引)	1	
H_2O	6.5	715
gDNA	1.5	
总计	**20**	**1925**
反应：110	分装 17.5 μl MM	
	添加 1 μl 索引寡核苷酸	
	添加 1.5 μl PCR #1	

[0180] PCR使用下表10中的温度循环方案进行。时间的格式为分:秒。

[0181] 表10.NGS PCR2程序

[0182]

温度	时间	
98℃	0:30	
98℃	0:10	12×
60℃	0:30	
72℃	0:40	
72℃	10:00	
4℃	∞	

[0183] 所用的引物是通用引物,可添加Illumina测序仪所需的DNA序列。NGS PCR2产物经纯化并在Illumina测序仪上进行测序。

[0184] 关于汇集在一起的核苷酸位点处的特异性变化对序列进行分析。例如,对于用核苷酸21测试的样品,检查并记录了与野生型Bxb1重组酶所见的百分比不同的整合事件(图4)。

[0185] 实施例3:已鉴定残基的饱和诱变

[0186] 对在实施例2的丝氨酸扫描期间表现出不同整合概况的氨基酸进行饱和诱变。这些差异范围包括消除整合、放宽特异性、增加特异性或改变特异性。选择进行饱和诱变的位点包括F147、N148、I149、Y154、R155、G156、L158、P197、L198、W230、S231、A232、T233、R237、D257、E309、Y312、F314、A315、G316、G318、R323、Y324、R325、C326和C335。K562转染和NGS分析按照实施例2中描述的相同方式进行(图4)。

[0187] Bxb1重组酶的RD域中的饱和诱变结果的总结示于下表11中。“特异性转变”是指在给定位置处给定突变体向非WT核苷酸的最大转变。例如,如果WT Bxb1重组酶在位置9处具有以下偏好:4.5% A、85.6% C、1.0% G和8.9% T,并且如果R237A突变体在位置9处具有以下偏好:23.6% A、1.9% C、17.1% G和57.4% T,则向非WT核苷酸的最大转变为T(57.4%-8.9%＝48.5%变化);R237A在位置9处的特异性转变为48.5%。“平均%特异性转变”是所有突变体在给定氨基酸位置处的特异性转变的平均值。“%TI”是指成功靶向整合的百分比;其计算为具有与供体序列整合一致的序列读段占总序列读段的百分比。在突变

体的%TI中,一个类(例如R237X)是该类所有成员的%TI的平均值(例如R237A、R237C、R237D等的平均%TI),WT是四次重复的平均值。在下表11和12中,突变类(例如R237X)的"%TI"是该类所有成员的平均%TI(例如R237A、R237C、R237D等的平均%TI),并且WT是四次重复的平均值。

[0188] 表11.Bxb1 RD域饱和诱变结果

[0189]

			平均%特异性转变				
	感兴趣的突变体	%TI	A	C	N	A	C
WT	NA	77.4	0.0	0.0	0.0	0.0	0.0
F147X	3	33.1	8.4	10.7	4.6	9.4	18.9
N148X	2	61.8	2.5	2.7	1.1	2.7	4.5
I149X	0	8.4	0.0	0.0	0.0	0.0	0.0
Y154X	0	8.3	0.0	0.0	0.0	0.0	0.0
R155X	0	15.7	0.0	0.0	0.0	0.0	0.3
G156X	0	0.1	0.0	0.0	0.0	0.0	0.0
L158X	5	50.6	0.0	0.0	5.3	9.4	5.3
P197X	3	76.8	2.2	2.3	0.7	1.9	2.5
L198X	0	59.6	0.0	0.0	0.0	0.0	0.0
W230X	0	39.5	0.4	0.0	1.0	0.0	0.0
S231X	7	50.2	11.4	0.0	0.0	0.0	0.0

[0190]

A232X	0	70.9	0.2	0.0	0.9	0.5	1.3
T233X	6	64.4	8.8	2.1	0.0	2.4	5.1
R237X	19	28.4	0.0	47.6	0.0	0.0	0.0
D257X	0	74.5	5.4	3.9	3.3	4.1	6.1

[0191] Bxb1重组酶的ZD域中的饱和诱变结果的总结示于下表12中。

[0192] 表12.Bxb1 ZD域饱和诱变结果

[0193]

取代	感兴趣的突变体	%TI	平均%特异性转变						
			G	G	T	T	T	G	T
WT	NA	82.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E309X	0	79.9	1.0	0.5	0.2	0.7	1.4	0.5	4.0
Y312X	0	25.5	0.0	0.0	0.1	0.8	0.6	0.8	8.3
F314X	15	35.7	1.4	1.2	18.5	59.8	7.5	13.8	8.7
A315X	18	82.6	4.6	5.5	8.3	6.7	1.1	2.8	5.1
G316X	4	67.0	5.0	25.8	11.8	24.9	0.1	0.2	0.0
G318X	2	79.1	7.7	2.4	3.9	4.7	2.5	0.6	1.2
R323X	2	8.2	4.3	4.5	5.2	3.6	0.2	0.6	0.0
Y324X	0	41.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0
R325X	9	6.5	0.0	0.0	0.4	0.4	30.4	4.6	19.1
C326X	0	3.3	0.5	0.0	8.2	0.0	0.0	0.0	0.0
C335X	0	5.9	0.3	0.0	7.9	10.2	0.0	0.0	0.0

[0194] 并列实验中位置10处的S231F和WT Bxb1的详细结果示于图6A中,并且位置19和21处的F314G和G316Y的详细结果示于图8A中。

[0195] 实施例4:Bxb1介导的人类细胞内源性位点处的靶向整合

[0196] Bxb1变体的性能在一组内源性Bxb1位点处进行了表征,其中野生型Bxb1能够实现可检测的整合活性(图5)。

[0197] 人类基因组中的Bxb1重组酶伪attB靶序列如下表13中所示;伪attB位点在5’和3’端两侧有五个核苷酸,并且中心二核苷酸位置以粗体显示。

[0198] 表13.内源性Bxb1伪attB靶序列

[0199]

SEQ	位点	Bxb1 靶标
137	s5-1	CCATTCCTTCTTCTACCACAGCG**TT**CTCAATTGGGTAAGCA TCAAAAT
138	s5-2	TTTAGTTGCTCCTGACAGCAGTG**GG**AGCTATTGTCTAAGAG ATACAAA
139	s5-4	GGAGTTGAGTCCTGACTGCAGCA**CA**TGCAGGTGAGGAAGC AGTAGAAG
140	s5-6	TACTGGGGATTCCCATAACCGTG**CA**CTCAGCTGCGGGAAG GCAGGAGA
141	s5-8	GAGTAGTGTATCTGACACCAGCA**CA**CTCTGGTGTCAGAATC TTGGGTT
142	s5-10	ATAGTGAAACTTACAACACAGAG**TT**TCCAGTTGTTCAATCT CTCTTGT
143	s5-11	TTGAAGCCCCTTCTCCTACAGAG**CA**AGCAGCAGGGTAAATT CTGTATT
144	s5-14	TGGTGCAATCTTGGCTCACTGCAA**CC**TCTGCAGTTGAAGCA ATTCTCC
145	s5-15	ATTTGTGGTACTATACTACAGAG**AA**TGCTGATGTTAAAGAG CTAATTT
146	s5-16	GCACTCCAGCCTGGGCAACAGAG**CA**CTCTGTTGTCAAAAA AAAAAAAA
147	s5-17	ACACCTGGATCTCCACCCCTGTG**TT**TGCTGTTGTCTAAGTTT TACATT
148	s5-18	CCGTCATGTTCCTCACCACAGCT**AA**CCCAGTCTTCAAAACT CACTTAA
149	s5-19	GGCCTCCTGCTTGCTCAGCTGAG**AC**GGCAGAGGCTGAAGCT GCACGGG
150	s5-21	CTTCCTTCTTTTAAACCACTGAT**CA**TGCAGAAGTATAAGCT CACCAGT
151	s5-24	GTGGCAGAAATAGTACAACAGCT**AA**CGCAGATGTTAAAAA TCAAATGG
152	s1-1	CAGTGTGCTTGTTGCCTACAGCC**TC**TGCAGAAGTTCACAAA CTTGCAC
153	s1-5	ATTAGGGGTTGGTGGCAATGGAG**AT**CTCAGTGGTGAAAAA TCACTTCT
154	s1-10	ATGTGGTTTTGTTCACTGTAGAA**TC**CTCAGTGGTTAGCACA ATGTTAA
155	s1-11	CATTTCTCTTGATGACTGCAGAG**TA**TTCCATTGTTGACAAA TCCATTT
156	s1-17	GCTATGGGTTTTTAAACACTGAC**TT**GCCTGGTGTGAGGATC CATGCTT
157	s1-25	GTGTTCTGATACTTGCAACTGAA**CA**CACTGTAGTCCACACA CCACTAT
158	s1-28	GGGGGTTGCTTCTCACCGTAGAT**GA**CCCTCTTGTCCACAAA CCACAAC
159	s1-31	CAGCCTGGCTCCCAACCTCCGTC**AC**TGCCGTTGCAAACAAA CCAAACC

[0200]

160	s1-39	TTTCAAAGATCCTGACCTCCGAA**GA**AGCTGGAGTTGAGATC ATACTGT
161	s1-41	CGAGCTTGCTCCTCACAGCAGAG**GA**CTCTGAAGACAGGAG CATCATGA
162	s1-43	CCTTGCAGATCTATACGGCTGAC**TA**CACAGTGGTGAGAACC ATAGTCT
163	s3-28	ACCTTTGTTTGGCCCCAACTGCC**TC**TGCCACTGTCGACACA CTTGGGA
164	s3-41	GATGCATGATTCTAGGCACTGGG**AA**CACAGTGGTGATCAA ACCAGTCA

[0201] 通过将下表14中所示的靶位点特异性序列克隆至标准质粒DNA骨架中来产生单个供体构建体。每个供体DNA序列由Bxb1 attP序列(其中靶位点与中心二核苷酸匹配)以及靶位点特异性引物结合位点组成。attP序列和引物结合位点位于额外基因组序列的两侧,所述序列被部分随机化以形成标签序列,从而通过下一代测序检测靶向整合事件。

[0202] 表14.靶位点特异性供体序列

[0203]

SEQ	位点	供体
165	s5-1	GGTTTGTCTGGTCAACCACCGCGTTCTCAGTGGTGTACG GTACAAACCTTCGTTTCAACTGACCAATTCACTCCAAAT GTAACA
166	s5-2	GGTTTGTCTGGTCAACCACCGCGGGCTCAGTGGTGTAC GGTACAAACCAAGAGGAAAAGTATACCTGCAGCAAGT GATAT
167	s5-4	GGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTAC GGTACAAACCATTGCTGCCGTCTGTCCCCTTCCCATCCC TTCTGACATGT
168	s5-6	GGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTAC GGTACAAACCAGCCCGAACTTAGCTTCCCTCATCGCCT A
169	s5-8	GGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTAC GGTACAAACCATATCGTTTGCTACTCAAACTGTGTGGTC CT
170	s5-10	GGTTTGTCTGGTCAACCACCGCGTTCTCAGTGGTGTACG GTACAAACCTATCTTCCTTTTCTGTGGAAGGGAAAACTT
171	s5-11	GGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTAC GGTACAAACCAAGCAGCTCGGAGCCTCTCTGTCACTTG CTCTTTAGA

[0204]

172	s5-14	GGTTTGTCTGGTCAACCACCGCGACCTCAGTGGTGTAC GGTACAAACCGCCTGTCACCTCCTGAGTACCTGGGATT ACAGGCGCACAACACTACTCT
173	s5-15	GGTTTGTCTGGTCAACCACCGCGAACTCAGTGGTGTAC GGTACAAACCTCAGGAGGTGTTTTTCCTTCCTTTTGAAA TCAGATAGCAGGAAA
174	s5-16	GGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTAC GGTACAAACCAAGAGCACAGTGATTATGATTATTGTTT CTGAGTGATGGCA
175	s5-17	GGTTTGTCTGGTCAACCACCGCGTTCTCAGTGGTGTACG GTACAAACCTAATGTCCGCACTTCCTAAGTTTTGGTTTC CAGAGCCACTAGTAA
176	s5-18	GGTTTGTCTGGTCAACCACCGCGAACTCAGTGGTGTAC GGTACAAACCTCTATTCGCACTCAGGGAGGCCTTCCCT GATCTCCAGGGCCAGATTAAGTA
177	s5-19	GGTTTGTCTGGTCAACCACCGCGACCTCAGTGGTGTAC GGTACAAACCCGGCCAGTCAGTCACTGTCAGGTGAGCA CCCGCGAGGGTCCCGGGGAGGTACGGGTGAC
178	s5-21	GGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTAC GGTACAAACCACCTTCACGTCCTAACCAGACCATGAGC TCCAGGGGTAAAATC
179	s5-24	GGTTTGTCTGGTCAACCACCGCGAACTCAGTGGTGTAC GGTACAAACCCATGTAAAACCAGCTTATTGCCAACAGA GTTATGGA
180	s1-1	GGTTTGTCTGGTCAACCACCGCGTCCTCAGTGGTGTACG GTACAAACCAGAATCAGGATAAGGCTCCTCTAGGTACA TGTGGCTAAGCTGC
181	s1-5	GGTTTGTCTGGTCAACCACCGCGATCTCAGTGGTGTAC GGTACAAACCGGTCCTCCATCAAGCGCTGAGGCACAGT GAATGTACTCAGAGGG
182	s1-10	GGTTTGTCTGGTCAACCACCGCGTCCTCAGTGGTGTACG GTACAAACCACACAATAGCGACATGGAGCAGGTATCA GTTTACGTTT
183	s1-11	GGTTTGTCTGGTCAACCACCGCGTACTCAGTGGTGTAC GGTACAAACCAACCACCTGATTGTGATGGATGTTTAGG TTGTTTCCAATCTTTTGCTATTAAAAAGAATGTGACAGT GAGAAACTTTGTGTATAGTGTCATTTTGCAC
184	s1-17	GGTTTGTCTGGTCAACCACCGCGTTCTCAGTGGTGTACG GTACAAACCAGCATCTTTCAAATAGCACCTCATTTTATC CTGAAGACCCAG
185	s1-25	GGTTTGTCTGGTCAACCACCGCGCACTCAGTGGTGTAC GGTACAAACCCAAAACTGGTAGAATCTCAGTGACCTTA AAGATTGCCTGGG
186	s1-28	GGTTTGTCTGGTCAACCACCGCGGACTCAGTGGTGTAC GGTACAAACCCATCTCCCTTTTTCCGTGCTGATTAGGAC TGGGGCAGTGGGGTTCGTGTATTTGTGACCTGG

[0205]

187	s1-31	GGTTTGTCTGGTCAACCACCGCGACCTCAGTGGTGTAC GGTACAAACCGCAGTGTCCACCTGGCGGAACCTTTCTC TCCACTATCCA
188	s1-39	GGTTTGTCTGGTCAACCACCGCGGACTCAGTGGTGTAC GGTACAAACCTAGTAAATAGGATCAAAGGAATAGAGT CTGATGCC
189	s1-41	GGTTTGTCTGGTCAACCACCGCGGACTCAGTGGTGTAC GGTACAAACCCTGTGACTACATTTAGTGAGCAGGTGGA ATGAACAA
190	s1-43	GGTTTGTCTGGTCAACCACCGCGTACTCAGTGGTGTAC GGTACAAACCGGATAAGTTTCAGTAAAGGGACTACTGC AGGATTTGGTGGTA
191	s3-28	GGTTTGTCTGGTCAACCACCGCGTCCTCAGTGGTGTACG GTACAAACCTGCGTTGTCTGAGCTCCTCAGCACTCCCTC CCCAACTTCCTAT
192	s3-41	GGTTTGTCTGGTCAACCACCGCGAACTCAGTGGTGTAC GGTACAAACCAGTCACCAATCTGGAGGGGATGAGGTG AGGGAGGGATAGACAATA

[0206] K562细胞按照制造商的方案进行转染。用总共200ng的Bxb1表达质粒(200ng的一个Bxb1表达质粒,或100ng的Bxb1变体表达质粒与100ng的WT Bxb1表达质粒混合),以及1600ng靶位点特异性attP供体来转染质粒2E5细胞。在大多数实验中,Bxb1变体的TI活性数据是使用50％ Bxb1变体表达质粒和50％ WT Bxb1表达质粒的混合物获得的。一个例外是图11左图中显示的数据,其使用了100％(200ng)Bxb1 D257K表达质粒或100％ WT Bxb1表达质粒。另外,除了涉及D257K Bxb1变体的实验外,还使用了Bxb1变体和WT Bxb1的50％混合物。这是因为,具有改变的靶向偏好的Bxb1变体通常仅显示对具有表1中所示改变的半位点的改善的靶向性,并且可能比WT Bxb1更差地靶向供体中的attP位点和/或另一个半位点。变体Bxb1和WT Bxb1的50％混合物应允许Bxb1变体结合其偏好的半位点,并允许WT Bxb1结合供体和Bxb1变体不偏好的半位点。

[0207] 在37℃下培育三天后,转染细胞按照制造商说明书使用Lucigen的QuickExtract™(目录号:QE09050)来收获。使用Invitrogen的Accuprimer™Taq DNA聚合酶,高保真度(目录号:12346094)对提取的DNA进行聚合酶链式反应(PCR),如表6中所示。

[0208] 使用表15中所示的正向引物进行靶位点特异性PCR。

[0209] 表15.靶位点特异性正向引物序列

[0210]

SEQ	位点	F-引物
193	s5-1	ACACGACGCTCTTCCGATCTNNNNTTCTTTCATTTCGTGTGAGGGTC
194	s5-2	ACACGACGCTCTTCCGATCTNNNNACCCTGTGTTAGAAGCAGTAAC
195	s5-4	ACACGACGCTCTTCCGATCTNNNNATGTGTTTGGAAGCTGAGCATAG
196	s5-6	ACACGACGCTCTTCCGATCTNNNNCGATGTGTTCCTCCATGAAGCA
197	s5-8	ACACGACGCTCTTCCGATCTNNNNGCATAGCTGAGTAATACATTGCATG
198	s5-10	ACACGACGCTCTTCCGATCTNNNNATGGGCAGCATTCATTAAGTAAAT
199	s5-11	ACACGACGCTCTTCCGATCTNNNNCAATTAGTTGGCTGTATAATTTGG
200	s5-14	ACACGACGCTCTTCCGATCTNNNNATTTCACATTGATTGATTTTCTTGTT
201	s5-15	ACACGACGCTCTTCCGATCTNNNNACTGCCACCTACATTCAACAAAT
202	s5-16	ACACGACGCTCTTCCGATCTNNNNAGAGGGCTGAGGTTGCAATG
203	s5-17	ACACGACGCTCTTCCGATCTNNNNTGCAGATAATTCTTATGCAAAACAGAAC
204	s5-18	ACACGACGCTCTTCCGATCTNNNNGCAGAAAGTTAGTATGACACAGAC
205	s5-19	ACACGACGCTCTTCCGATCTNNNNGACCAGACAGTTCCAGATACATTCCA
206	s5-21	ACACGACGCTCTTCCGATCTNNNNTTACTCTCTCTCCTAGACTCAAGC
207	s5-24	ACACGACGCTCTTCCGATCTNNNNTGCTAGTATTTGTCTGCACTTTCT
208	s1-1	ACACGACGCTCTTCCGATCTNNNNCGTGGGCCAGTTTTAAATGTCA
209	s1-5	ACACGACGCTCTTCCGATCTNNNNGCTGAACATTAGCTCCATATAT
210	s1-10	ACACGACGCTCTTCCGATCTNNNNAGTCTCTCATTTACCAGTTTTGCTT
211	s1-11	ACACGACGCTCTTCCGATCTNNNNGGATTATCCCATATCCAGACATA
212	s1-17	ACACGACGCTCTTCCGATCTNNNNAGCCCAGAGTTAACCAAGCTAC

[0211]

213	s1-25	ACACGACGCTCTTCCGATCTNNNNTACAACTAAAGCACCAATGGCTC
214	s1-28	ACACGACGCTCTTCCGATCTNNNNGTCAGAAGTCCCAGATGTGCT
215	s1-31	ACACGACGCTCTTCCGATCTNNNNGGAATCCGCCTCCTGACG
216	s1-39	ACACGACGCTCTTCCGATCTNNNNTGATGTTAGAGATCAGCCTGTAC
217	s1-41	ACACGACGCTCTTCCGATCTNNNNAGCCATTTCCTTCCTAGCAAATT
218	s1-43	ACACGACGCTCTTCCGATCTNNNNCAATACAGGCACAATCCCCTTATT
219	s3-28	ACACGACGCTCTTCCGATCTNNNNTGAAGGGAAAGGTCTGGCTTTA
220	s3-41	ACACGACGCTCTTCCGATCTNNNNCAGCAGTCGATGTGGGAAC

[0212] 使用下表16中所示的反向引物进行靶位点特异性PCR。

[0213] 表16.靶位点特异性反向引物序列

[0214]

SEQ	位点	R-引物
221	s5-1	GACGTGTGCTCTTCCGATCTTGTTACATTTGGAGTGAATTGGTCA
222	s5-2	GACGTGTGCTCTTCCGATCTATATCACTTGCTGCAGGTATACTTTT
223	s5-4	GACGTGTGCTCTTCCGATCTACATGTCAGAAGGGATGGGAAG
224	s5-6	GACGTGTGCTCTTCCGATCTTAGGCGATGAGGGAAGCTAAGT
225	s5-8	GACGTGTGCTCTTCCGATCTAGGACCACACAGTTTGAGTAGC
226	s5-10	GACGTGTGCTCTTCCGATCTAAGTTTTCCCTTCCACAGAAA
227	s5-11	GACGTGTGCTCTTCCGATCTTCTAAAGAGCAAGTGACAGAGAGG
228	s5-14	GACGTGTGCTCTTCCGATCTAGAGTAGTGTTGTGCGCCTG
229	s5-15	GACGTGTGCTCTTCCGATCTTTTCCTGCTATCTGATTTCAAAAGGA
230	s5-16	GACGTGTGCTCTTCCGATCTTGCCATCACTCAGAAACAATAATCA
231	s5-17	GACGTGTGCTCTTCCGATCTTTACTAGTGGCTCTGGAAACCA
232	s5-18	GACGTGTGCTCTTCCGATCTTACTTAATCTGGCCCTGGAGATC

[0215]

233	s5-19	GACGTGTGCTCTTCCGATCTGTCACCCGTACCTCCCCG
234	s5-21	GACGTGTGCTCTTCCGATCTGATTTTACCCCTGGAGCTCATG
235	s5-24	GACGTGTGCTCTTCCGATCTTCCATAACTCTGTTGGCAATAAG
236	s1-1	GACGTGTGCTCTTCCGATCTGCAGCTTAGCCACATGTACCTA
237	s1-5	GACGTGTGCTCTTCCGATCTCCCTCTGAGTACATTCACTG
238	s1-10	GACGTGTGCTCTTCCGATCTAAACGTAAACTGATACCTGCTCC
239	s1-11	GACGTGTGCTCTTCCGATCTGTGCAAAATGACACTATACACAAAG
240	s1-17	GACGTGTGCTCTTCCGATCTCTGGGTCTTCAGGATAAAATGAGG
241	s1-25	GACGTGTGCTCTTCCGATCTCCCAGGCAATCTTTAAGGTCAC
242	s1-28	GACGTGTGCTCTTCCGATCTCCAGGTCACAAATACACGAAC
243	s1-31	GACGTGTGCTCTTCCGATCTTGGATAGTGGAGAGAAAGGTTCC
244	s1-39	GACGTGTGCTCTTCCGATCTGGCATCAGACTCTATTCCTT
245	s1-41	GACGTGTGCTCTTCCGATCTTTGTTCATTCCACCTGCTCACT
246	s1-43	GACGTGTGCTCTTCCGATCTTACCACCAAATCCTGCAGTAGT
247	s3-28	GACGTGTGCTCTTCCGATCTATAGGAAGTTGGGGAGGGAGTG
248	s3-41	GACGTGTGCTCTTCCGATCTTATTGTCTATCCCTCCCTCACCT

[0216] PCR使用表8中的温度循环方案进行;时间格式为分钟:秒。将下一代测序(NGS)PCR1的产物用作NGS PCR2的模板,如下表9中所示。PCR使用表10中的温度循环方案进行;时间格式为分钟:秒。所用的引物是通用引物,可添加Illumina测序仪所需的DNA序列。NGS PCR2产物经纯化并在Illumina测序仪上进行测序。另请参阅Miller等人,Nat Biotechnol.(2019)37(8):945-52以了解有关此基于测序的测定的进一步描述。

[0217] 在一个示例性实验中,确定了位置10的Bxb1变体S231F的靶标偏好转变。相对于WT Bxb1重组酶,S231F变体表现出对C和T碱基改善的靶向性(图6A),从而使得内源性人类靶位点s5-1、s5-11和s1-41处的靶向整合得到改善(图6B-C)。图7显示了不同Bxb1变体在结合RD域的位置(例如位置7、9和10)处的靶标偏好变化。

[0218] 在其它示例性实验中,确定了位置19和21处的Bxb1变体F314G和G316Y的靶标偏好转变。相对于WT Bxb1,F314G变体表现出对位置19处的T改善的靶向性,并且相对于WT Bxb1,G316Y变体表现出对位置21处的G改善的靶向性(图8A)。发现F314G和G316Y变体分别在s5-16和s3-28内源性人类靶位点处具有改善的靶向整合(图8B-C)。图9和图10显示了各种Bxb1变体在结合ZD域的位置(例如位置19、21、22、23和24)处的靶标偏好变化。

[0219] 相对于WT Bxb1,此处描述的Bxb1变体还可表现出在多种不同的内源性人类靶位点处改善的靶向整合。例如,D257K Bxb1变体表现出在人类靶位点(包括s1-10、s1-39、s3-28、s3-41、s5-10、s5-11、s5-14、s5-15、s5-16和s5-17)处改善的靶向整合(图11)。

[0220] 序列表

[0221] 本文公开的序列列于下表中(SEQ:SEQ ID NO)。

[0222]

SEQ	描述
1	野生型 Bxb1 重组酶示例性氨基酸序列
2	野生型 attB 序列
3	野生型 attP 序列
4-40	表 4 中的变体 attB 序列
41-76	表 5 中的变体 attP 序列
77	野生型 φC31 整合酶示例性氨基酸序列
78	野生型 Pa557 重组酶示例性氨基酸序列
79	野生型 Pa570 重组酶示例性氨基酸序列
80	对称 attB 序列
81	共有 attP 序列
82	共有 attB 序列
83-130	表 3 中的内源性人类 attP 和 attB 样序列
131	野生型 LI 整合酶示例性氨基酸序列
132	野生型 A118 整合酶示例性氨基酸序列

133	野生型 TP901 重组酶示例性氨基酸序列
134	共有 attB 序列
135	共有 attB 序列
136	共有 attB 序列
137-164	内源性 Bxb1 伪 attB 靶序列
165-192	靶位点特异性供体序列
193-220	靶位点特异性正向引物序列
221-248	靶位点特异性反向引物序列
249	Bxb1 重组酶构建体 C 末端氨基酸序列
250-251	表 7 中的 NGS PCR1 引物
252-254	位置 10 处有突变的对称 attB 序列
255-257	位置 10 处有突变的 attP 序列
258-260	位置 9 处有突变的对称 attB 序列
261-263	位置 9 处有突变的 attP 序列
264-267	内源性 Bxb1 伪 attB 靶序列
268-269	对称 attB 序列
270-271	内源性 Bxb1 伪 attB 靶序列

[0223]

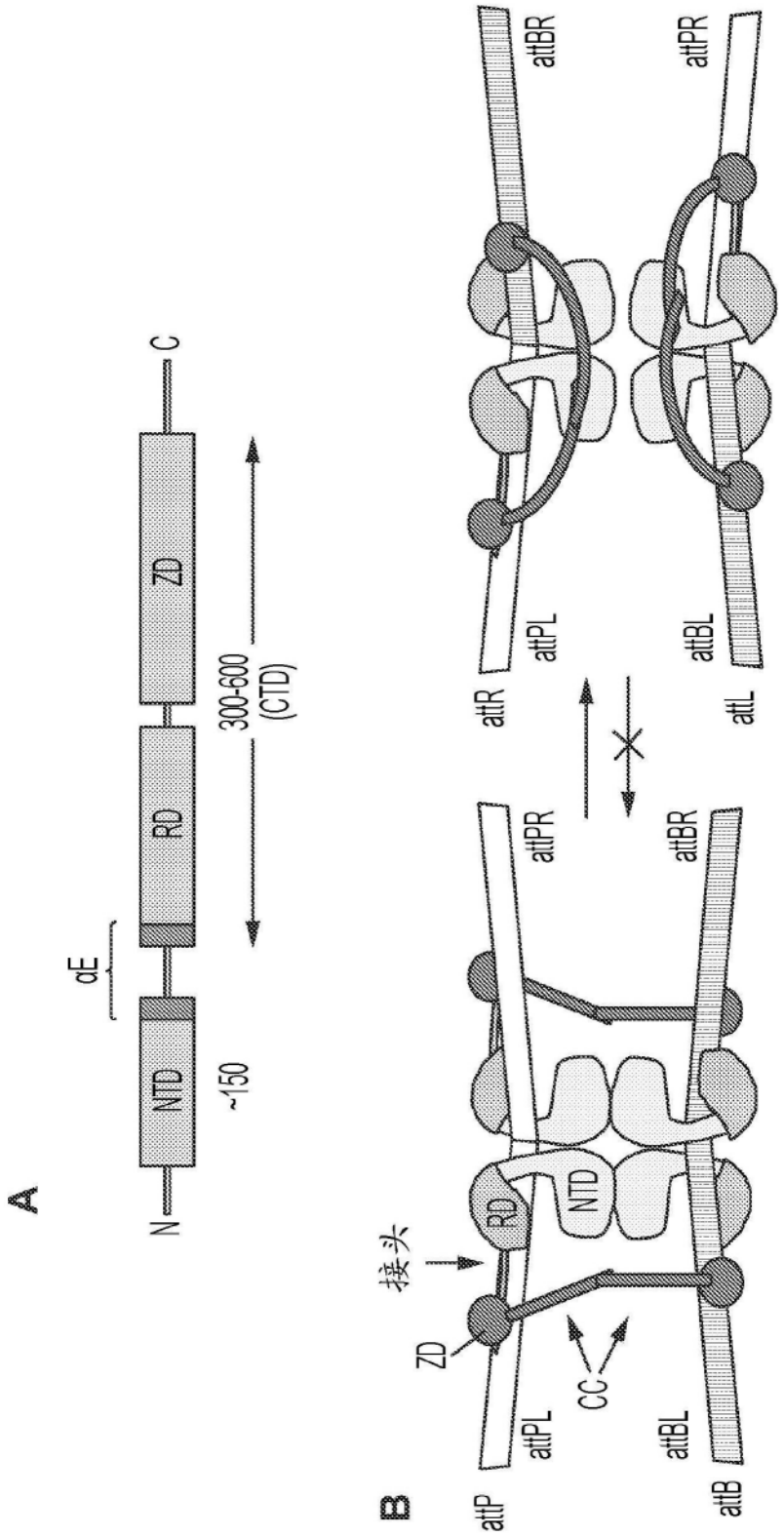

A
N
NTD
~150
αE
RD
ZD
C
300-600
(CTD)
B
attP
attPL
attB
attBL
attR
attL
attPR
attBR
接头
ZD
CC
RD
NTD

图1

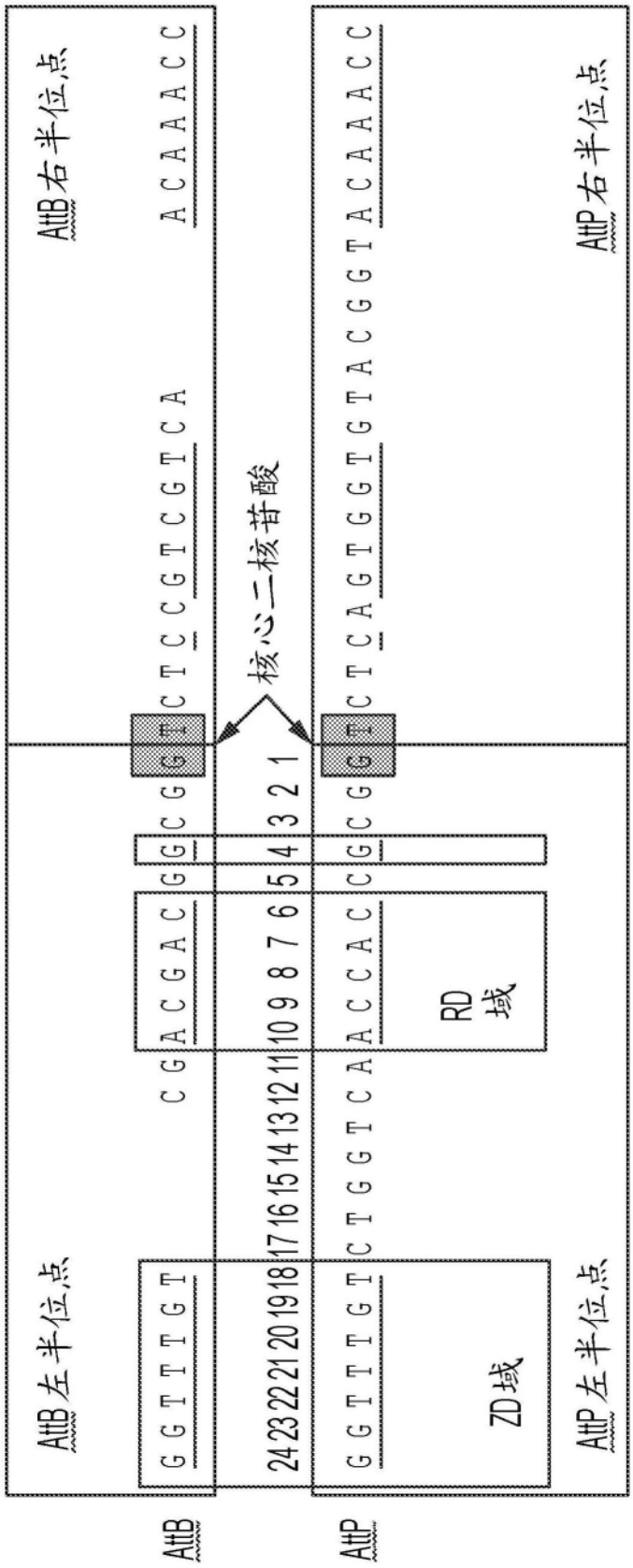

AttB左半位点
AttB右半位点
AttP左半位点
AttP右半位点
核心二核苷酸
RD域
ZD域
AttB
AttP

图2

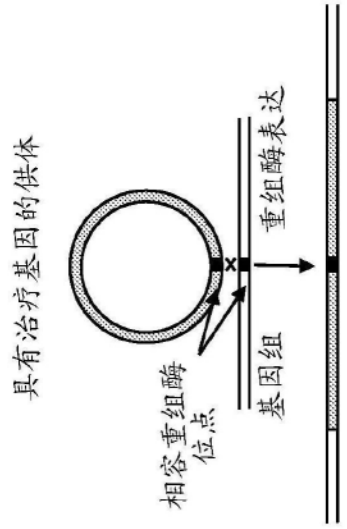


图3A

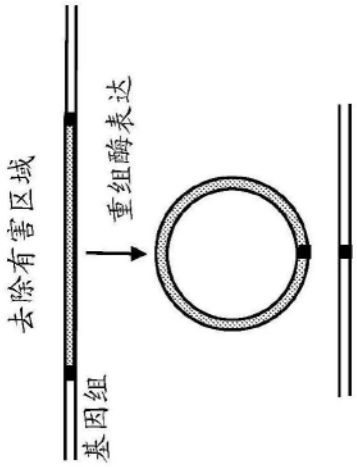


图3B

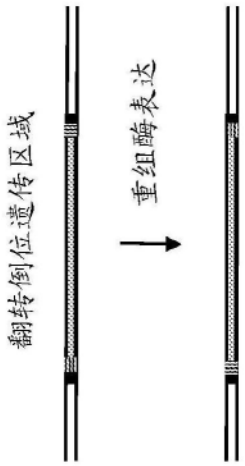

翻转倒位遗传区域
重组酶表达

图3C

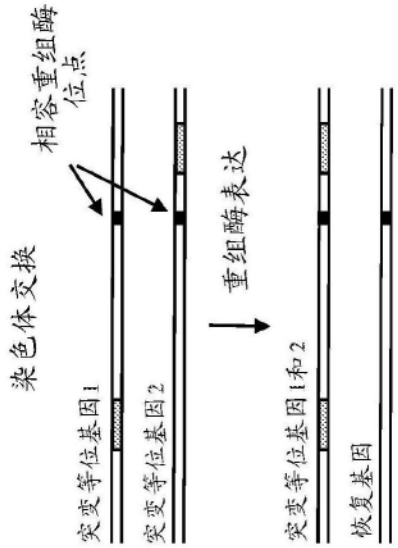

染色体交换
相容重组酶位点
突变等位基因1
突变等位基因2
重组酶表达
突变等位基因1和2
恢复基因

图3D

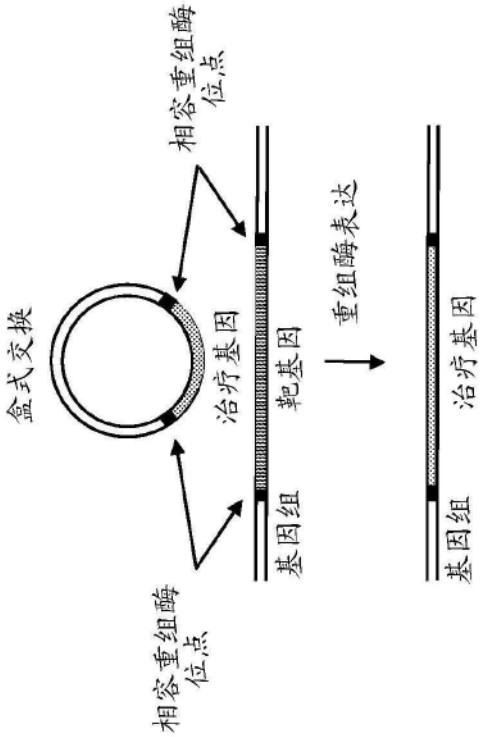


图3E

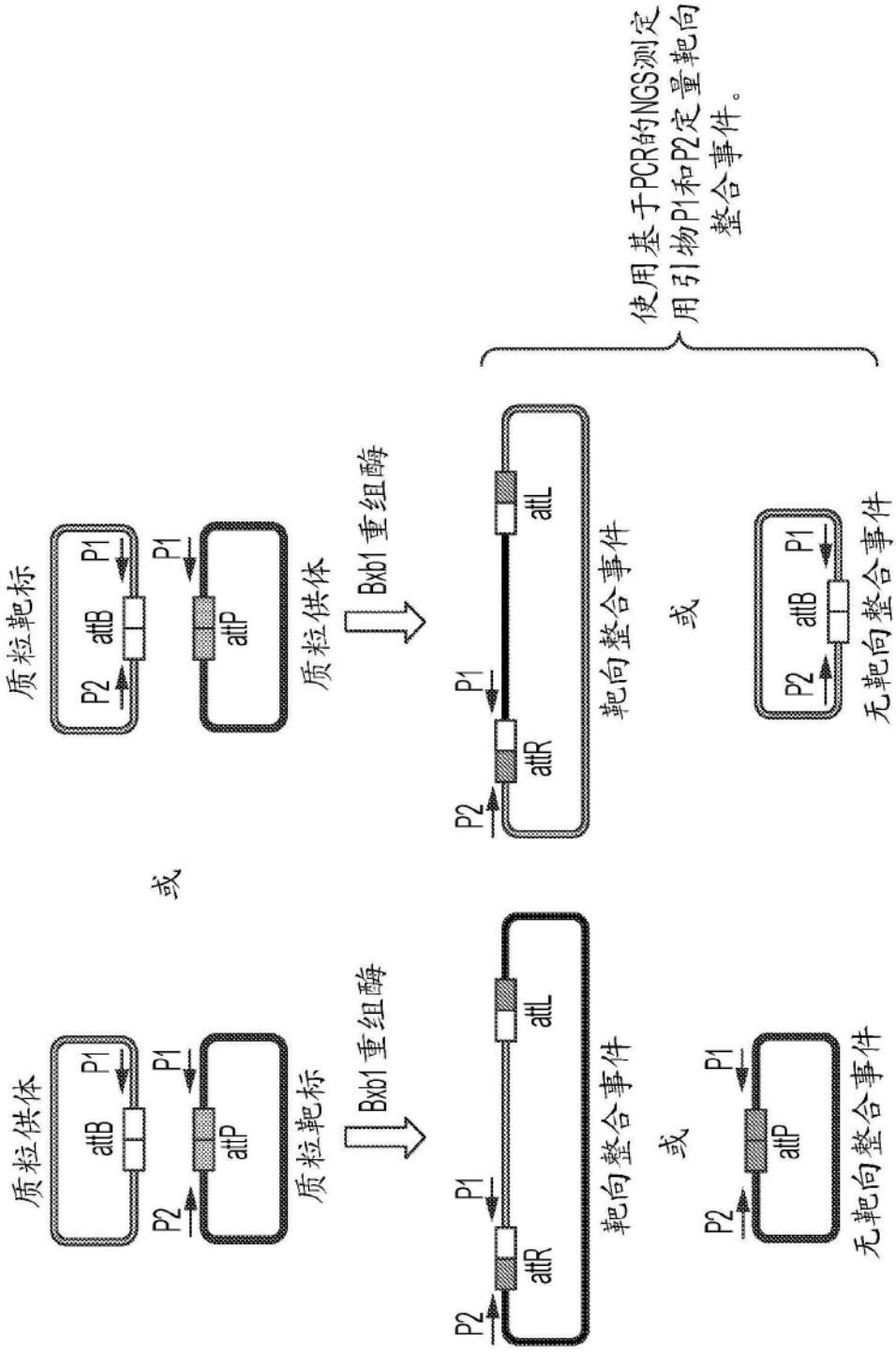

质粒靶标
P2
attB
P1
P1
attP
质粒供体
Bxb1 重组酶
P1
attL
P2
attR
靶向整合事件
或
P2
attB
P1
无靶向整合事件
使用基于PCR的NGS测定
用引物P1和P2定量靶向
整合事件。
质粒供体
attB
P1
P1
attP
P2
质粒靶标
Bxb1 重组酶
P1
attL
P2
attR
靶向整合事件
或
P1
P2
attP
无靶向整合事件

图4

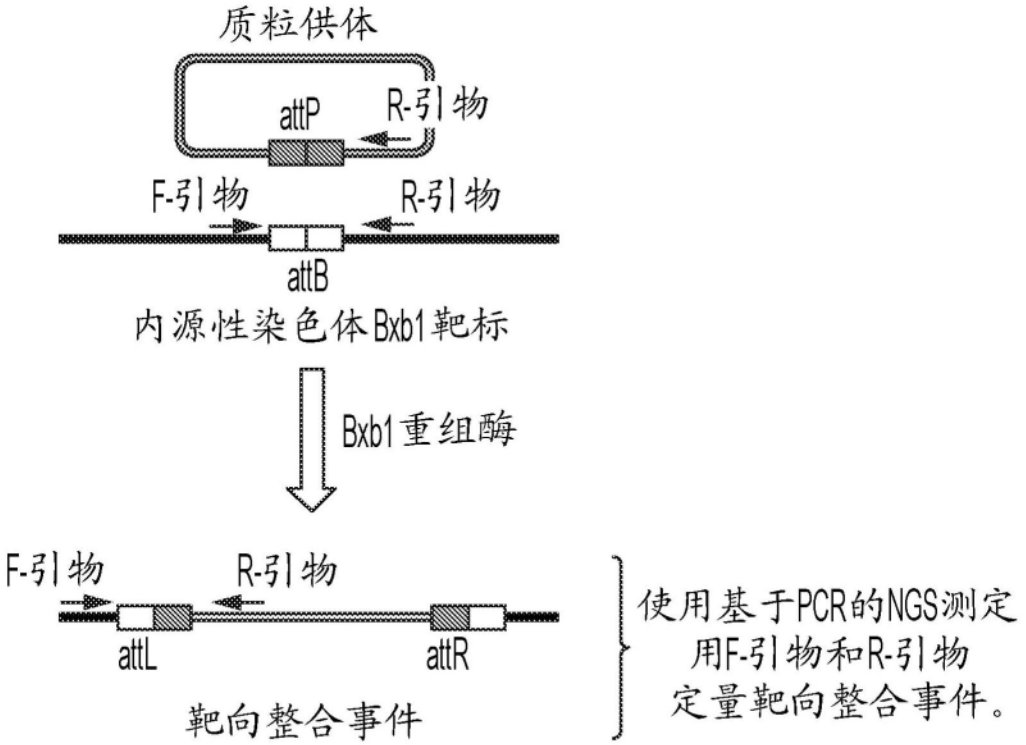


图5

	位置10(wt Bxb1靶标中的A)			
	%A	%C	%G	%T
WT	85.1	0.8	10.0	4.1
S231F	62.7	12.1	5.2	19.9

图6A

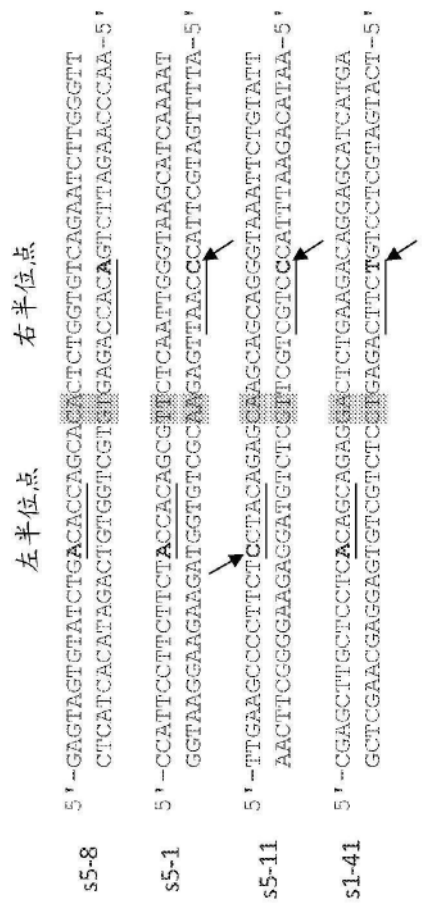
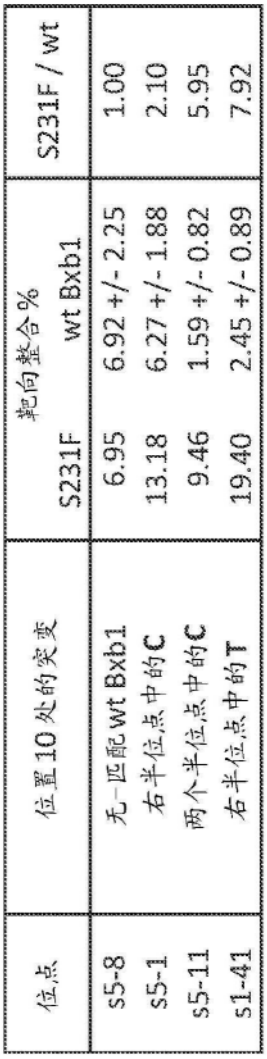

左半位点
右半位点
5'-GAGTAGTGTATCTGACACCAGCACCTCTGGTGTCAGAATCTTGGGTT
CTCATCACATAGACTGTGGTCGTGGAGACCACAGTCTTAGAACCCAA-5'
s5-8
5'-CCATTCCTTCTTCTACCACAGCGTTCTCAATTGGGTAAGCATCAAAAT
GGTAAGGAAGAAGATGGTGTCGCAAGAGTTAACCCATTCGTAGTTTTA-5'
s5-1
5'-TTGAAGCCCCTTCTCCTACAGAGCAAGCAGCAGGGTAAATTCTGTATT
AACTTCGGGGAAGAGGATGTCTCGTTCGTCGTCCCATTTAAGACATAA-5'
s5-11
5'-CGAGCTTTGCTCCTCACAGCAGAGACTCTGAAGACAGGAGCATCATGA
GCTCGAACGAGGAGTGTCGTCTCGAGACTTCTGTCCTCGTAGTACT-5'
s1-41

图6B



位点	位置10处的突变	靶向整合%		S231F / wt
		S231F	wt Bxb1	
s5-8	无-匹配wt Bxb1	6.95	6.92 +/- 2.25	1.00
s5-1	右半位点中的C	13.18	6.27 +/- 1.88	2.10
s5-11	两个半位点中的C	9.46	1.59 +/- 0.82	5.95
s1-41	右半位点中的T	19.40	2.45 +/- 0.89	7.92

图6C

位点	位置	半位点中的突变 左	半位点中的突变 右	Bxb1 突变体	%TI 突变体	%TI wt Bxb1	突变体/wt
s5-4	7	G	C	L158D	0.35	0.26 +/- 0.02	1.33
s5-4	7	G	C	T233F	0.66	0.26 +/- 0.02	2.51
s5-4	7	G	C	T233Y	0.88	0.26 +/- 0.02	3.34
s5-18	9	wt Bxb1	A	R237A	0.78	0.25 +/- 0.06	3.15
s5-18	9	wt Bxb1	A	R237C	0.51	0.25 +/- 0.06	2.06
s5-18	9	wt Bxb1	A	R237K	0.59	0.25 +/- 0.06	2.38
s5-18	9	wt Bxb1	A	R237N	0.57	0.25 +/- 0.06	2.30
s5-6	9	T	wt Bxb1	R237N	0.79	0.26 +/- 0.13	3.04
s3-28	10	C	wt Bxb1	N148T	0.41	0.25 +/- 0.05	1.62
s5-4	10	wt Bxb1	T	P197H	0.58	0.26 +/- 0.02	2.20
s3-28	10	C	wt Bxb1	S231F	0.42	0.25 +/- 0.05	1.66
s5-1	10	wt Bxb1	C	S231F	13.18	6.27 +/- 1.88	2.10
s5-11	10	C	C	S231F	9.46	1.59 +/- 0.82	5.95
s5-4	10	wt Bxb1	T	S231F	1.60	0.26 +/- 0.02	6.08
s5-19	10	T	G	S231F	23.59	2.11 +/- 0.56	11.18
s1-41	10	wt Bxb1	T	S231F	19.40	2.45 +/- 0.89	7.92
s5-4	10	wt Bxb1	T	S231G	0.76	0.26 +/- 0.02	2.89
s5-4	10	wt Bxb1	T	S231R	0.43	0.26 +/- 0.02	1.63
s1-41	10	wt Bxb1	T	S231R	5.77	2.45 +/- 0.89	2.36
s5-11	10	C	C	S231Y	6.00	1.59 +/- 0.82	3.77
s3-28	10	C	wt Bxb1	S231V	0.40	0.25 +/- 0.05	1.58
s5-4	10	wt Bxb1	T	S231V	0.57	0.26 +/- 0.02	2.16
s5-4	10	wt Bxb1	T	T233F	0.66	0.26 +/- 0.02	2.51
s1-1	10	C	wt Bxb1	T233K	0.10	0.04 +/- 0.02	2.43
s1-5	10	G	wt Bxb1	T233R	0.09	0.03 +/- 0.02	2.94
s3-28	10	C	wt Bxb1	T233W	0.39	0.25 +/- 0.05	1.54
s5-4	10	wt Bxb1	T	T233W	1.55	0.26 +/- 0.02	5.89
s5-19	10	T	G	T233W	5.21	2.11 +/- 0.56	2.47
s5-4	10	wt Bxb1	T	T233Y	0.88	0.26 +/- 0.02	3.34

图7

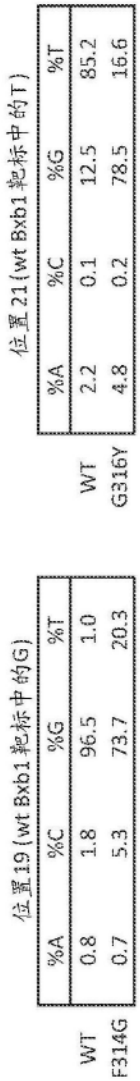

位置21(wt Bxb1靶标中的T)
%A %C %G %T
WT 2.2 0.1 12.5 85.2
G316Y 4.8 0.2 78.5 16.6
位置19(wt Bxb1靶标中的G)
%A %C %G %T
WT 0.8 1.8 96.5 1.0
F314G 0.7 5.3 73.7 20.3

图8A

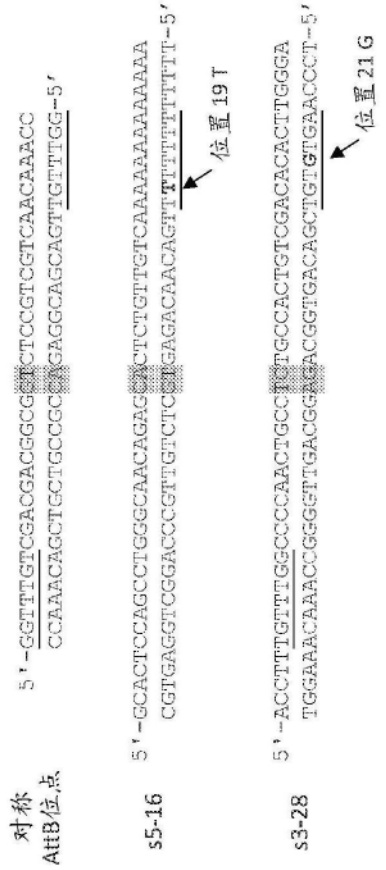

左半位点
右半位点
对称
AttB位点
5'-GGTTTGTCGACGACGGCGGTCTCCGTCGTCAACAAACC
CCAAACAGCTGCTGCCGCCAGAGGCAGCAGTTGTTTGG-5'
5'-GCACTCCAGCCTGGGCAACAGACGACTCTTGTTGTCAAAAAAAAAAAAA
CGTGAGGTCGGACCCGTTGTCTGCTGAGACAACAGTTTTTTTTTTTTTT-5'
位置19 T
s5-16
5'-ACCTTTGTTTGGCCCCAACTGCCTTGCCACTGTCGACACACTTGGGA
TGGAAACAAACCGGGGTTGACGGAACGGTGACAGCTGTGTGAACCCT-5'
位置21 G
s3-28

图8B

位点	位置19处的突变	靶向整合%		F314G / wt
		F314G	wt Bxb1	
s5-16	右半位点中的T	3.11	1.09 +/- 0.28	2.85

位点	位置21处的突变	靶向整合%		G316Y / wt
		G316Y	wt Bxb1	
s3-28	右半位点中的G	0.45	0.25 +/- 0.05	1.78

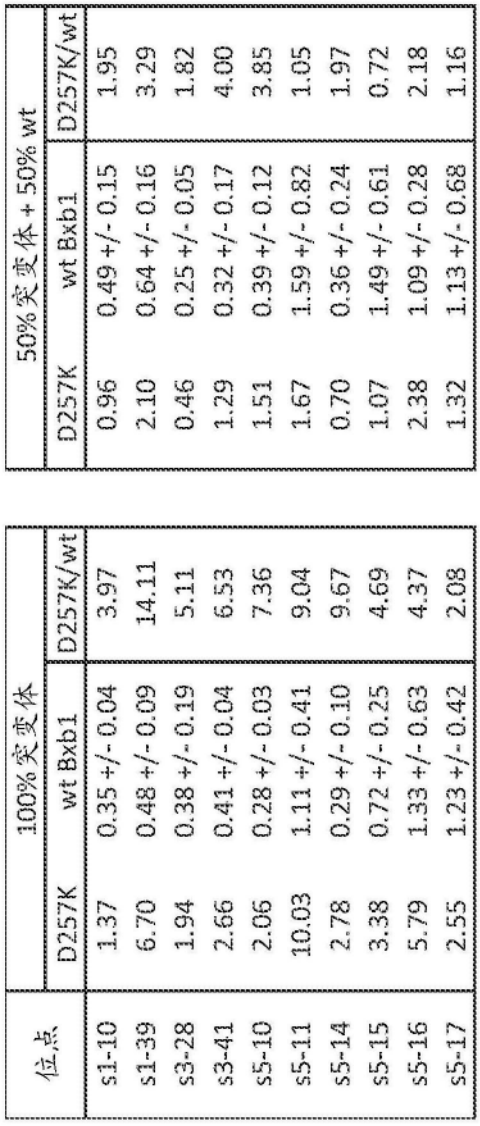
图8C

位点	位置	半位点中的突变 左	右	Bxb1 突变体	%TI 突变体	wt Bxb1	突变体/wt
s5-24	19		T	F314A	0.21	0.12 +/- 0.03	1.75
s5-6	19		T	F314C	0.66	0.26 +/- 0.13	2.54
s5-16	19		T	F314C	2.21	1.09 +/- 0.28	2.03
s1-5	19	wt Bxb1	T	F314C	0.15	0.03 +/- 0.02	4.97
s5-6	19		T	F314G	1.84	0.26 +/- 0.13	7.08
s5-16	19		T	F314G	3.11	1.09 +/- 0.28	2.85
s5-21	19	T		F314N	1.74	0.85 +/- 0.15	2.04
s5-21	19	T		F314Q	1.78	0.85 +/- 0.15	2.09
s5-6	19		T	F314Q	0.71	0.26 +/- 0.13	2.73
s5-16	19		T	F314Q	3.21	1.09 +/- 0.28	2.94
s1-5	19	wt Bxb1	T	F314Q	0.12	0.03 +/- 0.02	4.19
s5-6	19		T	F314S	2.03	0.26 +/- 0.13	7.81
s1-5	19	wt Bxb1	T	F314S	0.14	0.03 +/- 0.02	4.66
s5-6	19		T	F314T	2.40	0.26 +/- 0.13	9.23
s5-16	19		T	F314T	2.27	1.09 +/- 0.28	2.08
s1-5	19	wt Bxb1	T	A315F	0.12	0.03 +/- 0.02	3.92
s5-6	19		T	A315I	0.85	0.26 +/- 0.13	3.27
s1-5	19	wt Bxb1	T	A315I	0.11	0.03 +/- 0.02	3.61
s5-16	19		T	A315S	2.79	1.09 +/- 0.28	2.56
s1-5	19	wt Bxb1	T	A315S	0.08	0.03 +/- 0.02	2.60
s5-24	19		T	A315W	0.24	0.12 +/- 0.03	2.00
s5-16	19		T	A315Y	2.11	1.09 +/- 0.28	1.94
s5-16	19		T	G318R	2.17	1.09 +/- 0.28	1.99

图9

位点	位置	半位点中的突变 左	右	Bxb1 突变体	%TI 突变体	wt Bxb1	突变体/wt
s5-6	21		C	F314E	1.06	0.26 +/- 0.13	4.08
s3-28	21	wt Bxb1	G	G316Y	0.45	0.25 +/- 0.05	1.78
s1-5	22	G	A	A315F	0.12	0.03 +/- 0.02	3.92
s1-5	22	G	A	A315H	0.16	0.03 +/- 0.02	5.27
s1-5	22	G	A	A315N	0.13	0.03 +/- 0.02	4.32
s5-24	22		A	A315W	0.24	0.12 +/- 0.03	2.00
s1-5	22	G	A	G316A	0.11	0.03 +/- 0.02	3.82
s1-5	22	G	A	G316C	0.14	0.03 +/- 0.02	4.86
s1-11	23	T	A	A315F	0.11	0.04 +/- 0.01	3.19
s1-11	23	T	A	A315H	0.35	0.04 +/- 0.01	9.87
s1-11	23	T	A	A315W	0.12	0.04 +/- 0.01	3.30
s1-11	23	T	A	A315Y	0.30	0.04 +/- 0.01	8.28
s5-21	23	T		G316C	1.47	0.85 +/- 0.15	1.72
s1-10	23	T	T	G316C	1.57	0.49 +/- 0.15	3.20
s1-11	23	T	A	G316C	0.08	0.04 +/- 0.01	2.35
s1-11	23	T	A	G316T	0.18	0.04 +/- 0.01	4.98
s1-11	24	C	wt Bxb1	A315F	0.11	0.04 +/- 0.01	3.19
s1-11	24	C	wt Bxb1	A315Y	0.30	0.04 +/- 0.01	8.28

图10



位点	100%突变体			50%突变体 + 50% wt		
	D257K	wt Bxb1	D257K/wt	D257K	wt Bxb1	D257K/wt
s1-10	1.37	0.35 +/- 0.04	3.97	0.96	0.49 +/- 0.15	1.95
s1-39	6.70	0.48 +/- 0.09	14.11	2.10	0.64 +/- 0.16	3.29
s3-28	1.94	0.38 +/- 0.19	5.11	0.46	0.25 +/- 0.05	1.82
s3-41	2.66	0.41 +/- 0.04	6.53	1.29	0.32 +/- 0.17	4.00
s5-10	2.06	0.28 +/- 0.03	7.36	1.51	0.39 +/- 0.12	3.85
s5-11	10.03	1.11 +/- 0.41	9.04	1.67	1.59 +/- 0.82	1.05
s5-14	2.78	0.29 +/- 0.10	9.67	0.70	0.36 +/- 0.24	1.97
s5-15	3.38	0.72 +/- 0.25	4.69	1.07	1.49 +/- 0.61	0.72
s5-16	5.79	1.33 +/- 0.63	4.37	2.38	1.09 +/- 0.28	2.18
s5-17	2.55	1.23 +/- 0.42	2.08	1.32	1.13 +/- 0.68	1.16

图11