(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113853440 A
(43)申请公布日 2021.12.28

(21)申请号 202080037176.8

(22)申请日 2020.05.22

(30)优先权数据
62/851,543 2019.05.22 US

(85)PCT国际申请进入国家阶段日
2021.11.18

(86)PCT国际申请的申请数据
PCT/GB2020/051253 2020.05.22

(87)PCT国际申请的公布数据
WO2020/234608 EN 2020.11.26

(71)申请人 牛津纳米孔科技公开有限公司
地址 英国牛津

(72)发明人 戴晓光 M•彭德尔顿

(74)专利代理机构 北京同立钧成知识产权代理有限公司 11205
代理人 严罗一 臧建明

(51)Int.Cl.
C12Q 1/6806(2018.01)
C12Q 1/6869(2018.01)

权利要求书2页 说明书15页 附图5页

(54)发明名称

用于检测细胞内一个或多个DNA分子内相互作用的方案

(57)摘要

一种用于检测细胞内一个或多个DNA分子内的元件之间的相互作用的方法,其中元件在一级DNA序列中不相邻,该方法包括:a)提供细胞,其中紧邻的一个或多个DNA分子内的元件是交联的;b)同时裂解细胞和机械地分裂细胞内的DNA分子;c)邻位连接一个或多个分裂DNA分子;d)逆转连接DNA分子中的交联;e)对连接DNA分子进行测序;以及f)分析测序数据以检测细胞内一个或多个DNA分子内元件之间的相互作用。

CN 113853440 A

1.一种用于检测细胞内一个或多个DNA分子内元件之间的相互作用的方法,其中所述元件在一级DNA序列中不相邻,所述方法包括:

a)提供细胞,其中紧邻的一个或多个DNA分子内的元件是交联的;

b)同时裂解所述细胞和机械地分裂所述细胞内的所述DNA分子;

c)邻位连接所述一个或多个分裂DNA分子;

d)逆转所述连接DNA分子中的交联;

e)对所述连接DNA分子进行测序;以及

f)分析测序数据以检测所述细胞内所述一个或多个DNA分子内元件之间的相互作用。

2.根据权利要求1所述的方法,其中所述细胞被裂解并且通过珠磨分裂所述DNA分子。

3.一种用于检测细胞或细胞核内一个或多个DNA分子内元件之间的相互作用的方法,其中所述元件在一级DNA序列中不相邻,所述方法包括:

a)提供细胞或细胞核,其中紧邻的一个或多个DNA分子内的元件是交联的;

b)通过珠磨机械地分裂所述细胞内的所述DNA分子;

c)邻位连接所述一个或多个分裂DNA分子;

d)逆转所述连接DNA分子中的交联;

e)对所述连接DNA分子进行测序;以及

f)分析测序数据以检测所述细胞内所述一个或多个DNA分子内元件之间的相互作用。

4.根据权利要求1至3中任一项所述的方法,还包括交联所述DNA分子和/或DNA相互作用蛋白质的初始步骤。

5.根据权利要求4所述的方法,其中通过用甲醛和/或戊二酸二琥珀酰亚胺酯(DSG)处理所述细胞来实现交联。

6.根据前述权利要求中任一项所述的方法,还包括在步骤(c)之前使步骤(b)中产生的所述DNA片段平末端化的步骤。

7.根据前述权利要求中任一项所述的方法,还包括在步骤(c)或(d)之后将接头添加到所述DNA片段的末端的步骤。

8.根据权利要求7所述的方法,其中所述接头是测序接头。

9.根据权利要求8所述的方法,其中所述测序接头是PCR测序接头。

10.根据前述权利要求中任一项所述的方法,还包括在步骤(d)之后纯化所述DNA片段的步骤。

11.根据前述权利要求中任一项所述的方法,还包括在步骤(b)、(c)或(d)之后选择所需大小的DNA片段。

12.根据前述权利要求中任一项所述的方法,还包括在步骤(e)之前富集一个或多个关注的DNA分子的步骤。

13.根据权利要求12所述的方法,其中通过将一种或多种标记的寡核苷酸与所述DNA分子内的一个或多个关注的区域杂交并选择标记的DNA分子来富集关注的DNA片段。

14.根据权利要求13所述的方法,其中所述寡核苷酸用亲和标签标记,并且通过与所述亲和标签的结合配偶体结合来选择标记的DNA分子。

15.根据前述权利要求中任一项所述的方法,其中测序通过基于纳米孔的方法进行。

16.根据前述权利要求中任一项所述的方法,其中所述DNA分子包含染色体序列和/或

染色体外序列。

17.根据前述权利要求中任一项所述的方法,其中所述分析所述测序数据的步骤包括识别来自所述一个或多个DNA分子内不同元件的串联序列,从而检测一个或多个DNA元件中的相互作用元件。

18.根据前述权利要求中任一项所述的方法,其中所述元件是染色体内的基因座。

用于检测细胞内一个或多个DNA分子内相互作用的方案

技术领域

[0001] 本发明总体上涉及一种用于检测细胞内一个或多个DNA分子内元件之间的相互作用的方法,其中元件在一级DNA序列中不相邻。

背景技术

[0002] 当前需要的是能够提供可改进对基因组空间组织的理解的信息的技术。发展速度缓慢的现有技术包括引入了偏差的步骤。这些偏差阻止整个基因组序列中DNA相互作用的解析。

发明内容

[0003] 本发明人已经指出了用于检测和解析细胞内一个或多个DNA分子内元件之间的相互作用的新型方法。具体地,这些方法能够检测在一级DNA序列中不相邻的相互作用元件。相互作用信息可以提供对支撑基因组层次组织的构象特征的理解。具体地,这些构象特征包括整个染色体领域、大规模活性和抑制区室、拓扑相关域、核纤层蛋白相关域、核仁相关域以及相同或不同染色体内一个或多个元件之间的单独循环相互作用。当应用于异质宏基因组样品时,这些方法还可以提供同样的信息,尤其是识别细菌染色体与其质粒之间的相互作用。

[0004] 这些方法的一项关键优势在于,同时进行的细胞机械分裂和交联DNA机械分裂的步骤意味着:由于方案中的步骤数量较少,因此引发错误的能力降低。单步式细胞裂解和DNA分裂使得这些方法得到简化和精简。此外,重要的是,这些方法提供了基于测序的元件相互作用信息,该信息与序列无关,因而不受限制酶靶向等偏差的影响;这些方法不受DNA碱基化学修饰或DNA序列可及性的影响。特别地,本领域中的类似方法在分裂步骤中利用限制酶。该方法不受限制酶基序的浓度过低或过高的基因组区域的影响。相反,根据本方法的DNA机械分裂能够在不牺牲分辨率的情况下最大程度地绘制测序数据。

[0005] 因此,本文提供了一种用于检测细胞内一个或多个DNA分子内的元件之间相互作用的方法,其中这些元件在DNA一级序列中不相邻,该方法包括:a)提供一种细胞,其中一个或多个DNA分子内的元件靠近的DNA分子是交联的;b)同时裂解细胞并机械裂解细胞内的DNA分子;c)邻近连接一个或多个片段化的DNA分子;d)逆转连接的DNA分子中的交联;e)对连接的DNA分子进行测序;f)分析测序数据以检测细胞内一个或多个DNA分子内元素之间的相互作用。

[0006] 还提供了一种用于检测细胞或细胞核内一个或多个DNA分子内元件之间相互作用的方法,其中元件在DNA一级序列中不相邻,该方法包括:a)提供细胞或细胞核,其中一个元件内的元件或多个相邻的DNA分子交联;b)通过珠跳动使细胞内的DNA分子机械断裂;c)邻近连接一个或多个片段化的DNA分子;d)逆转连接的DNA分子中的交联;e)对连接的DNA分子进行测序;f)分析测序数据以检测细胞内一个或多个DNA分子内元素之间的相互作用。

附图说明

[0007] 应理解的是,附图仅用于说明本发明的特定实施例而并不旨在是限制性的。

[0008] 图1示出了如何使用本公开的方法来提供关于一个或多个DNA分子内元件之间的相互作用的信息的实施例。该实施例显示了样品管中完整细胞的异质混合物。然后,将交联剂应用于完整细胞以交联细胞内的分子。特别地,该交联剂可以使DNA分子与相互作用的DNA分子、DNA分子与相互作用的蛋白质和/或蛋白质与相互作用的蛋白质交联。圆形虚线代表细胞、细胞核或任何其他囊泡。标有“基因组A”、“质粒A”和“基因组B”的线条代表DNA分子。较小的重叠灰色圆圈代表彼此相互作用并同时与DNA分子相互作用的蛋白质。因此,该示意图中的蛋白质“桥接”了一个或多个DNA分子内的相互作用元件。由于应用了交联剂,这些蛋白质-蛋白质和蛋白质-DNA相互作用会发生交联。通过“珠磨”的机械过程对细胞和DNA分子进行分裂。然后,将交联DNA分子的分裂末端连接到彼此接近的分裂末端。在示例性附图的顶部图中,在各个分子内元件的情况下,分裂基因组A DNA分子连接到分裂的质粒ADNA分子,其中各个分子内的元件,由此意味着彼此紧邻的各个交联DNA分子的片段末端被连接。该实施例表明,然后可以逆转交联并且可以纯化连接的DNA分子。在该实施例中,来自顶部图的经过纯化的连接DNA分子代表基因组A和质粒A序列的串联序列,这表明:这些序列内的元件在它们所源自的原始细胞中彼此相互作用。该实施例进一步表明,经过纯化的DNA分子随后可以通过聚合酶链反应(PCR)进行大小选择和扩增,然后再进行测序文库制备方案(例如,掺入一个或多个接头、前导序列和/或发夹环)和测序。

[0009] 图2示出了示例性的生物信息学分析工作流程,借此通过图1中实施例描述的方法获得序列读数,其中通过基于纳米孔的方法执行获得测序读数的测序步骤。在图2中,这些测序读数被称为“纳米孔MetaPore-C读数”。以MetaPore-C读数(串联序列)为例的测序读数与参考基因组序列进行局部比对。从图2的左下图可以看出,单个测序读数的区域可能与存在于一个以上的物种/基因组中的相同序列对齐。因此,对通过每个单独的MetaPore-C序列读数的比对路径进行优化,以解析测序读数所对齐的最可能的物种。该实施例进一步表明,基因组序列可以被隔离到合适长度(以bp为单位)的“分箱”中,并且可以将比对的MetaPore-C测序读数分配给所述分箱。然后,基于所分配的读数在MetaPore-C测序读数中彼此相邻的频率,分配给分箱的读数的数量可以用于制表接触图(热图)。

[0010] 图3示出了源自示例性方法的数据,这些方法显示了益生菌样品中染色体内和染色体外接触(相互作用)的识别。A示出了包含在初始益生菌样品内的15种已知细菌菌株的表,对该初始益生菌样品执行图1中描绘的方法,以确定细胞内一个或多个DNA分子内元件之间的相互作用。B和C示出了代表根据图2中列出的生物信息学工作流程制备的A的15种细菌菌株中每一种的接触图。D示出了平均核苷酸密度热图,用于表示A的15个细菌菌株(和它们的相关质粒)之间的基因组相似程度。E和F示出了表示每个细菌DNA分子的接触数量和接触类型的条形图。

具体实施方式

[0011] 应理解,所公开的产品和方法的不同应用可以根据本领域的特定需要而定制。还应理解,本文所用的术语仅出于描述本发明的特定实施例的目的,而不打算是限制性的。

[0012] 另外,如在本说明书和所附权利要求中所使用的,除非内容另外明确指明,否则单

数形式的“一种(a)”、“一个(an)”以及“所述(the)”均包括复数对象。因此,例如,提及“多核苷酸(apolynucleotide)”包括两个或更多个多核苷酸,提及“跨膜孔(a transmembrane pore)”包括两个或更多个孔,等等。

[0013] 本文引用的所有出版物、专利和专利申请,无论是上文还是下文,均通过引用整体并入本文。

[0014] 方法

[0015] 提供了一种用于检测细胞内一个或多个DNA分子内的元件之间相互作用的方法,其中所述元件在初级DNA序列中不相邻,该方法包括:a)提供一种细胞,其中一个或多个DNA分子内的元件靠得很近是交联的;b)同时裂解细胞并机械裂解细胞内的DNA分子;c)邻近连接一个或多个片段化的DNA分子;d)逆转连接的DNA分子中的交联;e)对连接的DNA分子进行测序;f)分析测序数据以检测细胞内一个或多个DNA分子内元素之间的相互作用。

[0016] 还提供了一种用于检测细胞或细胞核内一个或多个DNA分子内元件之间相互作用的方法,其中元件在DNA一级序列中不相邻,该方法包括:a)提供细胞或细胞核,其中一个元件内的元件或多个相邻的DNA分子交联;b)通过珠跳动使细胞内的DNA分子机械断裂;c)邻近连接一个或多个片段化的DNA分子;d)逆转连接的DNA分子中的交联;e)对连接的DNA分子进行测序;f)分析测序数据以检测细胞内一个或多个DNA分子内元素之间的相互作用。

[0017] 该方法可用于例如获得与细胞中一个或多个DNA分子(例如基因组)的空间组织相关的信息。特别地,该方法可以提供与细胞中基因组的层次组织有关的信息。可通过本方法解析的支持基因组层次组织的示例性构象特征包括整个染色体领域、大规模活性和抑制区室、拓扑相关域、核纤层蛋白相关域、核仁相关域以及相同或不同染色体内一个或多个元件之间的单独循环相互作用。当应用于异质宏基因组样品时,这些方法还可以提供同样的信息,尤其是识别细菌染色体与其质粒之间的相互作用。

[0018] 元件之间的相互作用

[0019] 在本文所述的任何方法中,“元件”可以指一个或多个核酸分子内任何大小的核苷酸序列的一部分。核酸分子可以是脱氧核糖核酸(DNA)或核糖核酸(RNA)。核酸分子可以包括与一条DNA链杂交的一条RNA链。

[0020] 核酸分子优选地是DNA、RNA或者DNA或RNA杂交体,最优选地是DNA。核酸分子可以是双链的。核酸分子可以是基因组DNA。核酸分子可以包含单链区域和具有其它结构的区域,例如发夹环、三链体和/或四链体。DNA/RNA杂交体可以在同一条链上包含DNA和RNA。优选地,DNA/RNA杂交体包含与RNA链杂交的一条DNA链。

[0021] 核酸分子可以具有任何长度。例如,核酸分子的长度可以是至少10、至少50、至少100、至少150、至少200、至少250、至少300、至少400或至少500个核苷酸或核苷酸对。靶核酸分子的长度可以是1000个或更多个核苷酸或核苷酸对、5000个或更多个核苷酸或核苷酸对或者100000个或更多个核苷酸或核苷酸对。核酸分子可以是整个基因组。核酸分子可以是包含在细胞内的所有核酸分子的整体。核酸分子可以是包含在细胞内的所有核酸分子的子选择。核酸分子可以是包含在细胞内的全部DNA。核酸分子可以是包含在细胞内的DNA的子选择,例如单个染色体。

[0022] 元件可以是一个或多个核酸分子内任何大小的核苷酸序列的一部分。元件可以是根据给定的基因组参考组装由特定坐标定义的基因座。因此,元件可以是一个或多个染色

体内的基因座。元件可以是基因组的编码或非编码序列。元件可以是异染色质(封闭染色质)或常染色质(开放染色质)内的核苷酸序列。元件可以是顺式调节元件或顺式调节模块。元件可以是启动子、增强子、沉默子、外显子、内含子。元件可以是蛋白质的结合位点,例如组蛋白、转录因子和/或反式作用因子。元件可以是侧接组蛋白的开放染色质的一部分。元件可以是CpG岛。元件可以是基因沙漠。元件可以是转录因子结合基序。元件可以是包含彼此连锁不平衡的小核苷酸多态性的区域。元件可以是单个SNP、CpG或单个核苷酸碱基。在本文所述的任何方法中,元件在一个或多个DNA分子的一级核苷酸序列中不相邻。

[0023] 在本文所述的任何方法中,“相互作用”可以指一个或多个DNA分子内元件之间的任何形式的直接或间接接触。元件可以包含在一种或多种细胞内的一个或多个DNA分子内。相互作用可以指元件之间的间接或直接相互作用,其中元件在一级DNA序列中不相邻。因此,相互作用可以提供3D基因组结构的指示。特别地,相互作用可以进一步提供DNA分子内元件的特殊组织的精确图的指示。相互作用可以是彼此接近的两个或多个元件。因此,在描述DNA分子内元件的空间组织的上下文中,彼此接近的元件也可以被认为是相互作用的,而不管两个或多个元件的接近是否会产生任何功能后果。当在本文中使用时,术语“接近”是指两个元件之间在三维空间中的距离。例如,在一级序列中接近(例如在约10、50、100、150、200或250bp或更多内)的染色体中的DNA序列元件总是彼此非常接近。在一些情况下,在染色体中的一级序列中相距较远的DNA序列元件(例如,间隔超过约200;250;300;400;500;1000;1500;2000;5000;10,000;25,000;50,000;100,000;250,000;500,000;或1,000,000bp)可以因为染色体的三级或四级结构而彼此紧邻。在一些情况下,位于不同染色体上的DNA序列元件可以因为染色体的四级结构而彼此紧邻。在一些情况下,核酸序列元件相对于一级序列是远端的,因为一个或多个元件是染色体DNA序列元件并且一个或多个其他元件是RNA(或cDNA)序列元件。因此,核酸序列元件可以是或可以在不同的核酸分子内。在这种情况下,两个或多个核酸序列元件可以因为它们形成复合物而彼此紧邻。例如,非编码RNA可以与基因组中的一个或多个DNA序列元件相关联。因此,在本文所述的任何方法中,DNA序列元件可以被认为是因为它们的接近而相互作用。

[0024] 两个或多个元件可以同时相互作用。元件可以直接地相互作用,或者可以间接地相互作用。间接相互作用可以通过与一种或多种蛋白质的直接相互作用来介导。例如,间接相互作用可以由同时与增强子和启动子结合的蛋白质复合物表示,其中该两个元件就一级序列而言彼此相距超过100,000bp。

[0025] 样品

[0026] 在本文所述的任何方法中,样品可以是任何合适的样品。样品应含有一个或多个DNA分子。样品通常是已知含有或怀疑含有一个或多个DNA分子的样品。样品可以含有一个或多个细胞。

[0027] 样品可以是生物样品。所公开的方法可以在体外对包含来自任何生物体或微生物的细胞的样品进行。所述生物或微生物通常是古细菌、原核生物或真核生物,并且通常属于五个王国之一:植物界、动物界、真菌、原核生物界和原生生物界。该方法可以在体外对从任何病毒获得或提取的样品进行。

[0028] 样品优选地是基于流体的。样品通常包括体液。体液可从人或动物获得。人或动物可能患有、疑似患有疾病或有患病风险。样品可以是尿液、淋巴液、唾液、粘液、精液或羊水,

但优选地是全血、血浆或血清。通常，样品是人来源的，但替代地，它可以来自其他哺乳动物，诸如来自商业养殖的动物，诸如马、牛、绵羊或猪，或者替代地可以是宠物，诸如猫或狗。

[0029] 替代地，通常从以下商业作物获得植物来源的样品：诸如谷类作物、豆科作物、水果或蔬菜，例如小麦、大麦、燕麦、油菜、玉米、大豆、大米、香蕉、苹果、西红柿、土豆、葡萄、烟草、豆类、扁豆、甘蔗、可可、棉花、茶或咖啡。

[0030] 样品可以是非生物样品。非生物样品优选地是流体样品。非生物样品的实例包括手术液、水诸如饮用水、海水或河水，以及用于实验室测试的试剂。例如，样品可以是包含来自两种或多种不同生物体的细胞的不均匀混合物的环境样品。

[0031] 样品可以在应用于本文所述的方法之前进行处理，例如，借助于离心处理或通过膜，所述膜过滤掉不需要的分子或细胞，例如红细胞。样品可以在获取之后立即测量。通常还可以在分析之前对样品进行储存，优选地低于-70℃。

[0032] 样品优选地包含基因组DNA。样品可以是一个或多个核。

[0033] 交联

[0034] 在本文所述的任何方法中，提供样品，其中一个或多个紧邻的DNA分子内的元件是交联的。样品可以包含一个或多个细胞。细胞可以通过任何适用于本文所述方法的交联剂进行交联。通过将交联剂应用于细胞，可以获得DNA分子的空间组织的“快照”，从而获得细胞内一个或多个DNA分子内元件之间的相互作用。可以将一种或多种交联剂应用于样品内的细胞，以共价键合彼此邻近的分子。优选地，该一种或多种交联剂将DNA与DNA、DNA与蛋白质以及蛋白质与蛋白质交联。更优选地，该方法还可以包括交联DNA分子和/或DNA相互作用蛋白。

[0035] 优选地，本文所述方法中的交联剂与蛋白质中的氨基和/或DNA中的亚氨基和氨基反应，由此能够在这些基团中的任何一个或所有基团之间形成交联。示例性交联剂包括甲醛、二琥珀酰亚胺戊二酸酯(DSG)、双[2-(N-琥珀酰亚胺-氧羰基氧基)乙基]砜(BSOCOES)、二琥珀酰亚胺二丁酸脲(DSBU)、1,5-二氟-2,4-二硝基苯(DFDNB)、己二酸二甲酯二盐酸盐(DMA)、庚二酸二甲酯(DMP)、辛二酸二甲酯(DMS)、二硫代双(丙酸琥珀酰亚胺)(DSP)、辛二酸二琥珀酰亚胺(DSS)、二琥珀酰亚胺亚砜(DSSO)、酒石酸二琥珀酰亚胺(DST)二甲基二硫代双丙酰亚胺酯(DTBP)、乙二醇双(琥珀酰亚胺琥珀酸酯)(EGS)、磺基-EGS、三-(琥珀酰亚胺)氨基三乙酸酯(TSAT)。优选地，交联是通过用交联剂处理一个或多个细胞来实现。甚至更优选地，交联是通过用甲醛和/或戊二酸二琥珀酰亚胺酯(DSG)处理细胞来实现。

[0036] 本领域技术人员可以适当地选择交联步骤的条件。例如，选择合适的缓冲液和/或温度进而使用任何给定的交联剂来实现所需的交联度是在本领域技术人员的常规技能范围内。

[0037] 裂解和分裂

[0038] 在本文所述的任何方法中，可以裂解交联的细胞。来自细胞的DNA分子可以在细胞裂解的同时或之后分裂。在本文所述的任何方法中，可以裂解交联的细胞，同时细胞内的DNA分子被分裂。

[0039] 细胞可以通过任何合适的方案裂解。细胞裂解可以通过物理破坏或基于溶液的方案执行。物理中断方案可以是“机械性”方案。物理破坏和机械方案可以将裂解细胞分裂，同时细胞内包含的DNA分子被分裂。适用于所公开方法的示例性物理破坏方案是使用waring

blender polytron。适用于所公开方法的另一个示例性物理破坏方案是使用Dounce均质器。适用于所公开方法的另一个示例性物理破坏方案是使用Potter-Elevehjem均质器。适用于所公开方法的另一个示例性物理破坏方案是使用超声处理。适用于所公开方法的另一个示例性物理破坏方案是使用冻融循环。适用于所公开方法的另一个示例性物理破坏方案是使用研杵和研钵。适用于所公开方法的物理破坏的优选示例是珠磨。

[0040] 珠磨包括将珠子与样品组合并物理搅拌该组合，从而导致样品中细胞的分裂。用于珠磨的珠子可以是适用于本文所述方法的任何合适的材料。例如，珠子可以是陶瓷、金属或玻璃。优选地，珠子是玻璃。珠子可以是适用于机械物理破坏方案中的样品的任何尺寸。珠子的直径可以小于5mm。珠子的直径可以小于3mm。珠子的直径可以小于1mm。优选地，珠子的直径在0.1mm至1mm之间。甚至更优选地，珠子的直径为0.5mm。除了所用珠子的大小和数量之外，还可以根据样品的体积通过任何合适的方法搅拌组合的样品和珠子。通过珠磨搅动细胞可以同时溶解细胞和分裂细胞内包含的DNA分子。

[0041] 一种示例性的搅拌方法是涡旋。可以使用任何标准实验室台式涡旋。优选地，涡旋具有强度设置。优选地，搅拌步骤在标准实验室台式涡旋上可选择的最高强度下执行。本文描述的方法中的任何珠磨步骤可以通过一个单一的搅拌周期来执行。搅拌周期可以少于30分钟。搅拌周期可以少于20分钟。搅拌周期可以是15分钟。

[0042] 本文描述的方法中的任何珠磨步骤可以通过搅拌和冷却循环来执行。例如，珠磨步骤可以包括由冷却温育步骤分离的5个搅拌循环。搅拌循环可以持续任何合适的时间段。分离冷却温育步骤可以持续任何合适的时间段。

[0043] 进行珠磨的样品优选地在珠磨方案期间保持冷却。珠磨可以在低于室温下执行。珠磨可以在低于18℃下执行。珠磨可以在4℃下执行。进行珠磨的样品可以在搅拌步骤之间在低于室温的温度下温育。进行珠磨的样品可以在搅拌步骤之间在低于18℃的温度下温育。进行珠磨的样品可以在搅拌步骤之间在4℃下温育。优选地，珠磨包括三个3分钟的涡旋步骤，其中涡旋在标准实验室台式涡旋上可选择的最高强度下执行，每次涡旋间隔4℃下或冰上的2分钟温育。

[0044] 包含在进行裂解的细胞内的DNA分子随后可以被分裂。尽管在本文所述的任何方法中可以裂解交联的细胞，但同时将细胞内的DNA分子分裂。同时进行的细胞机械分裂和交联DNA机械分裂的步骤意味着：由于方案中的步骤数量较少，因此引发错误的能力降低。单步式细胞裂解和DNA分裂使得这些方法得到简化和精简。

[0045] 例如，可以以任何方式组合所描述的珠磨参数，以实现细胞裂解和/或所需的DNA分裂程度。

[0046] DNA分子可以通过任何合适的方法分裂。DNA分裂包括将DNA分子分解成更小的片段。DNA分子可以来自已裂解的细胞或细胞核。优选地，细胞被裂解并且包含在细胞内的DNA分子在单个步骤中同时被分裂。

[0047] DNA可以被机械地分裂。优选地，该方法包括：裂解其中紧邻的一个或多个DNA分子内的元件是交联的细胞，并且同时机械地分裂细胞内的DNA分子。优选地，细胞被机械地裂解并且DNA分子被机械地分裂。甚至更优选地，细胞被机械地裂解并且包含在细胞内的DNA分子在单个步骤中同时被机械地分裂。任何上述裂解细胞的机械方法都可以分裂DNA。

[0048] 本公开的方法可以包括通过珠磨进行的机械分裂。珠磨可以应用于例如完整细

胞、完整细胞核、裂解细胞、裂解细胞核和/或分离DNA。优选地，本文所述方法的交联细胞被裂解，同时细胞内的DNA分子被分裂。更长的珠磨持续时间或更大的珠磨强度会导致DNA分子被分裂成更小的碎片。同时进行的细胞机械分裂和交联DNA机械分裂的步骤意味着：由于方案中的步骤数量较少，因此引发错误的能力降低。

[0049] 单步式细胞机械裂解和机械DNA分裂使得这些方法得到简化和精简。此外，所公开的方法提供了已经以不依赖序列的方式获得的DNA片段。这意味着，该方法的涉及机械分裂(例如珠磨)的分裂步骤不会受到过多或过少的限制酶基序或DNA化学修饰的影响。因此，当检测到元件之间的相互作用时，本方法的机械分裂步骤中偏差的缺失能够提高基因组覆盖率。根据本方法的DNA机械分裂能够在不牺牲分辨率的情况下最大程度地绘制测序数据。通过机械分裂步骤(例如珠磨)的不同方面，可以对片段大小进行微调。

[0050] DNA分子可以被分裂成适合于所选测序平台以应用于本文所述的方法的任何大小。机械分裂步骤可以产生至少约100bp的DNA分子，例如至少约250bp、至少约500bp、至少约1kbp、至少约2kbp、至少约5kbp、至少约5kbp、至少约10kbp或至少约15kbp。例如，片段可以具有约100bp至约15kbp的长度，例如约250bp、约500bp、约1kbp或约2kbp直至约5kbp、约10kbp或约15kbp。

[0051] 邻位连接

[0052] 在本文所述的任何方法中，分裂DNA分子可以被邻位连接。本上下文中的邻位连接具有形成多联体序列的作用，由此代表原始一个或多个DNA分子内的元件的DNA片段与在三维空间中邻近但在一级序列中不相邻的其他片段共价连接。因此，多联体序列表明在一个或多个DNA分子中哪些DNA元件彼此之间相互作用。

[0053] 在分裂步骤之后，分裂DNA分子样品优选地被稀释。如果没有稀释，则可能导致与随机地邻近溶液中的其他交联分裂DNA分子的交联分裂DNA分子的虚假邻位连接事件。作为对样品溶液进行稀释的替代方法，可以基于DNA片段的大小在琼脂糖凝胶上分离DNA片段，从而降低与随机地邻近其他交联分裂DNA分子的交联分裂DNA分子的虚假邻位连接事件的可能性。

[0054] 本文描述的任何方法中的邻位连接步骤可以用本领域已知的任何合适的DNA连接酶执行。示例性的连接酶和试剂盒可以包括T4 DNA连接酶、Tfi DNA连接酶、DNA连接酶I、DNA连接酶II、DNA连接酶III、DNA连接酶IV、小尺寸DNA连接酶、NEB的Blunt/TA预混液或NEB的快速连接™试剂盒中的一种或多种。优选地，连接酶能够连接双链DNA片段的平末端。

[0055] 在本文所述的任何方法中，执行邻位连接是为了提供5C或Hi-C文库。

[0056] 在邻位连接步骤之前但在DNA分子分裂步骤之后，仍然交联的DNA分裂DNA分子可能会经历“末端修复”。DNA片段末端修复可以促进邻位连接。可以使用任何合适的DNA末端修复方案。用于修复分裂DNA末端的示例性产品是NEBNext末端修复酶。优选地，分裂DNA分子具有“平末端”，从而提供具有平末端的DNA片段用于邻位连接。在本方法中可以使用能够使双链DNA分子平末端化的任何酶或试剂盒。

[0057] 在本文所述的任何方法中，在邻位连接步骤之后，可以通过任何合适的方法逆转交联。适用于本文所述方法的交联逆转方法可以取决于在本文所述方法的交联步骤中使用的一种或多种交联剂。例如，其中甲醛是在本文所述的方法中使用的交联剂，交联可以通过以下中的任一种来逆转：与高盐(例如NaCl)一起温育并在65℃下长时间温育，或者与结合

了RNaseA和蛋白酶K的Tris HCl缓冲液在65℃下长时间温育。

[0058] 连接DNA片段可以例如具有至少约250bp、至少约500bp、至少约1kbp、至少约2kbp、至少约5kbp、至少约10kbp或至少约15kbp的长度。例如，片段可以具有约250bp至约100kbp的长度，例如约2kbp、约5kbp或约10kbp直至约15kbp、约50kbp或约100kbp。

[0059] DNA纯化

[0060] 在本文所述的任何方法中，可以在邻位连接和交联逆转步骤之后纯化DNA分子。在本文所述的任何方法中，DNA可以通过本领域已知的适合于纯化DNA的任何方法来执行。优选地，应用于本方法的纯化方法提供对于测序而言足够纯的DNA。用于纯化DNA的示例性方法包括有机提取方法，例如苯酚-氯仿和乙醇沉淀、Chelex提取纯化和固相纯化，以及本领域任何已知的DNA纯化试剂盒。优选地，在本文所述的方法中使用的纯化步骤使用固相可逆固定(SPRI)珠子。

[0061] 大小选择

[0062] 本文所述的任何方法可以包括在任何合适阶段选择所需大小的DNA的步骤。可以在分裂交联DNA分子之后和/或在邻位连接一个或多个分裂DNA分子之后和/或逆转连接DNA分子中的交联之后选择所需大小的片段。大小选择可以通过任何合适的方法执行。在本文所述的任何方法中，包括选择所需大小的DNA片段的步骤的任何包括都将优选地在任何测序步骤之前立即发生。示例性的DNA大小选择方法包括在琼脂糖凝胶上分离DNA片段，然后切除包含所需大小的DNA片段的凝胶并纯化，SPRI珠子或BluePippin(Sage Science)。

[0063] 所需的DNA片段大小可以是变化的，这具体取决于将使用什么测序平台对连接DNA分子进行测序。200bp至500bp之间的片段大小是基于Illumina的测序方法的首选。当将来自Oxford Nanopore Technologies的测序方法应用于本文所述的方法时，所需的DNA片段大小通常超过500bp，优选地超过1kb，甚至更优选地超过3kb。

[0064] 例如，Oxford Nanopore Technologies的测序平台可以用于对长度超过3kb的DNA分子进行测序。优选地，形成多联体DNA片段的连接片段的长度足以唯一地映射到参考基因组组装。甚至更优选地，形成多联体DNA片段的连接片段足够长且具有足够高的读取质量，以便唯一地映射到参考基因组组装。

[0065] 富集

[0066] 本文公开的方法还可以包括富集一个或多个关注的DNA分子的步骤。关注的DNA分子可以在方法中任何被认为是适合的阶段进行富集。在某些情况下，连接DNA分子可以在测序之前进行富集。在其他情况下，连接DNA分子可以在测序之前立即进行富集。在其他情况下，连接DNA分子可以在纯化之后进行富集。在其他情况下，连接DNA分子可以在选择所需大小的DNA片段后进行富集。在其他情况下，连接DNA分子可以在纯化和大小选择之后进行富集。

[0067] 关注的DNA分子可以通过任何合适的方法进行富集。例如，关注的DNA分子可以是其相互作用的配偶体受到关注的特定元件。

[0068] 一种示例性的富集方法是通过将互补碱基序列的一个或多个标记的寡核苷酸与DNA内的一个或多个特定关注区域杂交，其中标记是亲和标签，并且其中关注的DNA分子是分离的且因此被富集，通过用亲和标签的结合配偶体靶向亲和标签，并丢弃与该结合配偶体无关的任何DNA。示例性的亲和标签和结合配偶体是生物素和链霉亲和素。

[0069] 其他示例性的富集方法是通过反向聚合酶链反应(PCR)。在本方法的上下文中的通过反向PCR的富集可以包括将连接DNA分子环化,并且其中与环化DNA分子的特定靶区域互补的碱基序列的一对引物序列(因此,一个或多个DNA分子的元件)引发在反向上的PCR延伸,并且其中靶区域及其侧翼(相互作用)序列被扩增。

[0070] 其他示例性的富集方法是通过半特异性PCR。在本方法的上下文中的通过半特异性PCR的富集可以包括用末端制备酶混合物对连接DNA分子进行处理,以产生dA加尾的可连接末端,其中该可连接末端随后可以连接到通用PCR接头,并且其中针对DNA分子的靶元件的序列特异性引物可以与包含在PCR接头内的单一通用PCR引物组合,并且扩增靶及其侧翼(相互作用)序列。可以用对应的引物设计来研究靶的任一侧。

[0071] 接头

[0072] 本文所述的任何方法还可以包括在对DNA进行测序之前将接头添加到DNA片段的末端的步骤。优选地,接头是测序接头。测序接头可以是PCR测序接头。任何合适的测序接头都可以应用于本文所述的方法,这具体取决于所使用的测序平台,任何合适的测序平台均可以用于本文所述的方法中。更优选地,在本文所述的方法中使用与Oxford Nanopore Technologies的测序平台兼容的接头。

[0073] Oxford Nanopore测序接头可以包含至少一个单链多核苷酸或非多核苷酸区域。例如,用于纳米孔测序的Y-接头在本领域中是已知的。Y衔接子通常包含(a)双链区和(b)单链区或在另一端不互补的区域。如果Y接头包含单链区,那么它可以被描述为具有突出端。Y接头中非互补区的存在赋予了接头Y形状,因为两条链通常不彼此杂交,不同于双链部分。Y接头可以包含一个或多个锚。

[0074] Y接头优选地包含优选地螺旋到孔中的前导序列。前导序列通常包括聚合物。聚合物优选带负电荷。聚合物优选是多核苷酸,例如DNA或RNA、修饰的多核苷酸(例如脱碱基DNA)、PNA、LNA、聚乙二醇(PEG)或多肽。前导优选包含多核苷酸,并且更优选包含单链多核苷酸。单链前导序列最优选包含DNA的单链,例如poly dT区段。前导序列优选地包含一个或多个间隔子。

[0075] 前导序列可以是任何长度,但其长度通常是10到150个核苷酸,例如20到150个核苷酸。前导的长度通常取决于在方法中使用的膜嵌入纳米孔。

[0076] 前导序列优先旋入跨膜孔中,并且从而促进多核苷酸通过孔的移动。

[0077] Y接头可以包含捕获序列、亲和标签或孔系链,其在与接头连接的双链区展开时显露。捕获序列或标签用于在DNA分子随着DNA分子的第一链通过孔而展开时防止DNA分子的第二链扩散来远离纳米孔,其中孔与系链结合或用寡核苷酸标记,该寡核苷酸包含与Y接头中的捕获序列互补的序列,Y-接头上的标签的亲和配偶体。可以使用本领域中已知的任何方法将接头连接到DNA分子。可以使用连接酶(如T4 DNA连接酶、大肠杆菌DNA连接酶、Taq DNA连接酶、Tma DNA连接酶以及9°N DNA连接酶)来连接接头中的一个或两个。或者,可以使用下文论述的方法向DNA分子中加入接头。

[0078] 在一个实施方案中,该方法包括对样品中的一个或多个分子加以修饰,使得它们在一端包含Y接头并且在另一端包含发夹环。可以使用任何修饰方式。

[0079] 用于纳米孔测序的发夹环接头是本领域中已知的。发夹环可以设置在DNA分子的一端,该方法优选地还包括在DNA分子的一端为DNA分子提供发夹环。DNA分子的两条链可以

在一端与发夹环连接。

[0080] 测序

[0081] 本文所述的方法还可以包括对连接DNA分子进行测序的步骤。对连接DNA分子进行测序的步骤可以是为了确定其全部或部分序列。可以采用任何合适的测序技术来确定连接DNA分子的序列。在本公开的方法中，可以使用高通量(所谓的“第二代”、“第三代”和“下一代”技术)对连接DNA分子进行测序。

[0082] 在第二代技术中，并行地对大量DNA分子进行测序。通常，数以万计的分子以高密度锚定在给定位置，并在依赖于DNA合成的过程中确定序列。反应通常由连续的试剂输送和洗涤步骤(例如，以允许掺入可逆标记终止子碱基)以及扫描步骤(以确定碱基掺入的顺序)组成。这种类型的基于阵列的系统可从例如Illumina,Inc.(加利福尼亚州圣地亚哥)商购获得。

[0083] 第三代技术通常被定义为无需在检测步骤之间停止测序过程。例如，氢离子在掺入过程中发生的碱基特异性释放可以在微孔系统(例如，可从Life Technologies获得的Ion Torrent系统)环境中检测到。类似地，在焦磷酸测序中，对焦磷酸盐(PPi)的碱基特异性释放进行检测和分析。在纳米孔测序技术中，使DNA分子通过纳米孔或将其定位在纳米孔附近，并且在DNA分子相对于纳米孔移动之后，确定各个碱基的同一性。例如，这种类型的系统可从Oxford Nanopore Technologies商购获得。在另一种技术中，DNA聚合酶被限制在“零模式波导”中，并用伽马标记的磷核苷酸的荧光检测来确定所掺入的碱基的同一性(例如，参见Pacific Biosciences)。

[0084] 本文所述的方法可以包括分析测序数据以检测细胞内一个或多个DNA分子内元件之间的相互作用。分析测序数据可以包括识别来自一个或多个DNA分子内不同元件的串联序列，从而检测一个或多个DNA元件中的相互作用元件。

[0085] 下面的非限制性实施例说明了本发明，并且不是限制性的。

[0086] 实施例1：

[0087] 本实施例描述了适用于本公开的示例性实验室工作流程，即一种用于确定细胞内元件之间的相互作用的方法。具体地，方法用于研究在一级序列中不相邻的基因组元件之间的相互作用。除了使组装得到简化之外，所公开的方法在应用于宏基因组样品时还提供了一种将质粒与其宿主基因组相关联的方式。用于确定一个或多个DNA分子内元件之间的相互作用的替代方案采用限制性酶切在执行邻位连接之前对交联DNA进行分裂。然而，限制性酶切相当耗时，而且，限制酶的选择受样品中基因组的核苷酸组成的影响，而后者并不总是事先知晓的——尤其是在进行宏基因组研究时。本公开提供了一种通过使用机械分裂(例如，珠磨)来避免限制性酶切以同时裂解细胞和分裂DNA的方法。珠磨也可以用于从裂解细胞中分裂DNA。

[0088] 材料与方法

[0089] 图1中提供了当前描述的示例性方法的代表性示意图。更详细地，收集10^9完整微生物细胞，并通过离心(15000g下持续5分钟)沉淀。然后用PBS洗涤细胞一次，并通过相同的离心程序再次沉淀。将细胞重新悬浮在预混合缓冲液中：1.2mL PBS+34uL 37％甲醛(1％最终甲醛浓度)，并在室温下温育30分钟以交联DNA和蛋白质。加入170ul 1M甘氨酸溶液(125mM最终浓度)并使混合物再温育30分钟，以使得甘氨酸淬灭交联反应。然后，细胞通过

离心(15000g下持续5分钟)沉淀,并弃去上清液。之后可以将沉淀的细胞储存在-20℃下备用。

[0090] 然后将细胞沉淀物(～10μl)重新悬浮在200μl珠磨溶液(200μl 1x TBS、2μl 100x Halt Protease Inhibitor(Thermofisher)、2μl Triton X-100)中。然后,将100μl 0.5毫米直径的玻璃珠子(Qiagen)添加到悬浮液中,再将悬浮液以最高速度涡旋(VWR®Vortexer Mini 120v)3x5分钟,每次涡旋间隔冰上的2分钟温育。珠磨步骤为接下来的邻位连接步骤产生游离DNA末端。相比之下,大多数HiC制备物出于相同目的使用限制性酶切,而后者会受到基因组覆盖偏差、较低分辨率和更复杂的实验室步骤的影响。然后将悬浮液短暂离心以分离玻璃珠子。此时的细胞裂解物被发现位于上清液中。然后将裂解物转移到新的盆中并进一步离心(15000g下持续5分钟),之后再弃去上清液。然后将沉淀物重新悬浮在500μl 1x TBS中并进一步离心(17000g下持续5分钟)。弃去上清液,将沉淀物重新悬浮在200μl H_2O中。然后将重新悬浮的分裂DNA稀释10倍,再使用Qubit(Thermofisher)测量其浓度。之后,对1至5μg的重新悬浮的分裂DNA进行DNA末端修复("平末端化"),由之前的珠磨步骤产生的DNA片段。用于平末端化的反应混合物如表1中所述。将反应物在20℃下温育30分钟。

[0091] 表1

[0092]

1至5μg的分裂DNA	xμL
NEBNext末端修复缓冲液10x	10μL
酶混合物	5μL
H_2O	85-x uL
总计	100μL

[0093] 在温育30分钟后,将反应混合物在台式离心机上以最大速度离心5分钟,随后弃去上清液。将沉淀物重新悬浮在200μl的水中。然后,将重新悬浮的末端修复DNA稀释10倍,再使用Qubit测量其浓度。

[0094] 之后,对末端修复的DNA进行邻位连接。T4连接酶连接反应被设置为1-2ng/μl的DNA浓度。用于T4连接酶连接反应的反应混合物如表2所述。反应在室温下温育4小时,偶尔混合。

[0095] 表2

[0096]

10x T4连接酶缓冲液	100μL
DNA	1-2μg
水	到1000μL
T4 DNA连接酶(NEB,2M单位/mL)	10μL

[0097] 在4小时的温育时间之后,将反应混合物在17000g下离心5分钟。然后在不干扰沉淀物的前提下去除750μl的上清液。然后加入40μl的5M NaCl,涡旋混合物以使沉淀物重新悬浮。混合物在65℃下温育过夜以灭活T4连接酶和去交联。多联体DNA分子应该在该方法的邻位连接步骤之后形成。因此,出于QC目的,通过琼脂糖凝胶电泳应该可以看到多联体产物。

[0098] 如果需要,现在可以合并多个250μl连接DNA溶液。然后将5μl的Triton-x100和50μl的10％Tween-20添加到每250μl连接DNA溶液中。然后加水至总体积为1ml。为了净化DNA,将45μl蛋白酶K溶液(Qiagen)和2μl的100mg/ml RNaseA溶液加入到连接DNA溶液中,并在37

℃下温育30分钟。然后，将350μl的Qiagen缓冲液B2添加到反应中，同时在50℃下再温育30分钟。

[0099] 然后进行苯酚:氯仿提取和乙醇沉淀，以纯化DNA，即经过净化的DNA。然后通过nanodrop和琼脂糖凝胶电泳评估DNA的质量，并使用Qubit对其进行量化。

[0100] 为了扩增经过纯化的DNA，进行PCR模板制备。经过纯化的DNA用FFPE(NEB)和Ultra-II end-prep模块(NEB)处理。反应混合物如表3中所述。

[0101] 表3

[0102]

H_2O中的DNA	xμL(100ng-500ng)
NEBNext FFPE DNA修复缓冲液	3.5μL
NEBNext FFPE DNA修复混合物	2μL
UltraII End-prep反应缓冲液	3.5μL
UltraII End-prep酶混合物	3μL
总计	60μL

[0103] 将反应混合物混合、离心并在热循环仪中在20℃下温育15分钟，然后在65℃下温育5分钟。使用0.4x SPRI珠(例如，每60μl反应24μl SPRI)对反应进行净化，以选择长度超过1kb的DNA。然后，将与Oxford Nanopore Technologies的测序平台兼容的PCR测序接头连接到DNA。然后使用0.4x SPRI珠子净化DNA。

[0104] 为了获得用于纳米孔的最佳DNA量，DNA样品不应具有大量的高分子量扩增子。因此，建议先导PCR实验确定最佳的PCR循环数，首先执行8x到12x的循环，然后通过琼脂糖凝胶电泳使这些循环可视化。可以进行多次25μl PCR反应，如表4所示。

[0105] 表4

[0106]

	体积
DNA模板	5-10ng
ONT PCR引物混合物(PRM)	0.5μL
H_2O	到12.5μL
NEB长扩增热启动2x mastermix	12.5μL
总计	25μL

[0107] 对于基于纳米孔的测序应用，对于大于2到3kb的PCR产物，然后通过凝胶提取或Bluepippin(SAGE Science)选择最佳PCR循环DNA产物的大小。然后，根据本领域已知的纳米孔文库制备和测序方案执行基于纳米孔的测序。

[0108] 生物信息学工作流程

[0109] 图2提供了适用于目前描述的示例性方法的生物信息学分析工作流程的代表性示意图。在任何描述的方法中，生物信息学分析工作流程可以用于从使用应用于宏基因组样品的方法而得出的测序数据获得宏基因组接触图。更详细地，纳米孔测序数据是通过早先描述的方法生成的，以提供纳米孔测序读数。通过使用BWA-SW(Li H.和Durbin R.(2010) Fast and accurate long-read alignment with Burrows-Wheeler.Transform.Bioinformatics,Epub.[PMID:20080505])，读数首先与染色体序列和染色体外序列(例如质粒)的参考序列集合进行比对。每个比对的读数都被过滤以保留遍历大部分读数的最小比对集合。然后将参考基因组分成大小相等的分箱，纳米孔测序读数的每个比对片段被分配一个

分箱。最后，根据所有纳米孔测序读数计算出分箱到分箱接触的总数，并在接触图中将其可视化。通过确定哪些染色体与元件有最多的接触，可以将染色体外元件分配给它们的宿主。

[0110] 结果

[0111] 图3示出的结果表明，上述方法和工作流程产生的结果证明了益生菌样品中染色体内和染色体外接触的识别。来自益生菌食品补充剂样品的基因组DNA(其包含15种已知的细菌菌株(图3A))应用于上述方法和工作流程，并生成纳米孔测序数据。样品内细菌染色体和质粒的接触图是根据上面的生物信息学工作流程而制备的(图3B和图3C)。平均核苷酸同一性图(图3D)揭示了物种之间低水平的虚假相互作用，最有可能是由于纳米孔测序读取映射的模糊性。图3E和图3F总结了每个细菌染色体的接触。质粒与预期的宿主基因组相关联并识别了染色体内的相互作用，这对于分箱化以及接触图的组装而言很有价值。

[0112] 实施例2

[0113] 本实施例描述了适用于本公开的另一示例性实验室工作流程，即一种用于确定细胞内元件之间的相互作用的方法。具体地，方法用于研究在一级序列中不相邻的基因组元件之间的相互作用。除了使组装得到简化之外，所公开的方法在应用于宏基因组样品时还提供了一种将质粒与其宿主基因组相关联的方式。用于确定一个或多个DNA分子内元件之间的相互作用的替代方案采用限制性酶切在执行邻位连接之前对交联DNA进行分裂。然而，限制性酶切相当耗时，而且，限制酶的选择受样品中基因组的核苷酸组成的影响，而后者并不总是事先知晓的——尤其是在进行宏基因组研究时。本公开提供了一种通过使用机械分裂(例如，珠磨)来避免限制性酶切以同时裂解细胞和分裂DNA的方法。珠磨也可以用于从裂解细胞中分裂DNA。

[0114] 材料与方法

[0115] 样品收集和交联

[0116] 通过本领域已知的方法收集和分离大2-3×10^9个完整细菌细胞。将细胞重悬于预混合的交联缓冲液(1.2mL PBS，34μL 37％甲醛(1％甲醛最终浓度))并在室温下温育30分钟，偶尔混合。170微米加入L的1M甘氨酸(125mM甘氨酸终浓度)以淬灭交联反应，并将样品在室温下再孵育20分钟。然后将样品在17000g下离心5分钟，弃去上清液，用1X TBS洗涤交联的细胞沉淀物(固定细胞)。将样品在17000g下再离心5分钟，弃去上清液。然后可以将沉淀物在-80℃下储存备用。

[0117] 细胞裂解和末端修复

[0118] 如果先前是在-80℃下储存，则固定细胞的沉淀物应在冰上解冻。细胞(大约10μL)随后在200μL珠磨溶液中重悬，如表5所示。珠磨溶液的总体积可以根据需要按比例放大/缩小。

[0119] 表5

[0120]

试剂	体积
1X TBS	200μL
100X停止蛋白酶抑制剂	4μL
Triton X-100	2μL
总计	206μL

[0121] 然后将100μL的0.5mm直径的玻璃珠子(Qiagen)添加到重新悬浮的细胞中，然后将

样品以最高速度涡旋3x5分钟，每五分钟在冰上相隔2分钟。将样品短暂离心以消除气泡，并通过上下吹打样品溶液来重新混合样品，以产生均质的细胞裂解物。然后，将裂解物转移到新管中并在17000g下离心5分钟。弃去上清液，将沉淀物重新悬浮在500μL 1X TBS中。样品再次在17000g下离心5分钟，随后弃去上清液。将沉淀物重新悬浮在50μL H_2O中，充分混合，DNA浓度优选地通过Qubit(ThermoFisher Scientific)测量。末端修复反应如表6所示地配制。

[0122] 表6

[0123]

试剂	体积
高达5μg的染色质DNA	xμL
NEBNext末端修复缓冲液(10X)	10μL
NEBNext末端修复酶混合物	5μL
H_2O	85-xμL
总计	100μL

[0124] 末端修复反应在20℃下温育30分钟，然后在17000g下离心五分钟。弃去上清液，沉淀物用1X TBS洗涤。然后将样品在17000g下短暂离心，随后弃去上清液。然后，可以将沉淀物在-20℃下储存备用。

[0125] 邻位连接

[0126] 将经过末端修复的DNA的沉淀物重新悬浮在200μL H_2O中并通过Qubit进行定量。然后如下表7所示建立T4连接。

[0127] 表7

[0128]

试剂	体积
H_2O	1618-xμL
H_2O中的0.9ug DNA	xμL
10X T4连接酶缓冲液(Thermofisher)	180μL
充分混合	
T4 DNA连接酶(Thermofisher,30weissU/uL)	6μL
总计	1800μL

[0129] 邻位连接反应中DNA的最终浓度为0.5ng/μL。将反应在22℃下温育四小时，偶尔混合。然后将样品在17000g下离心5分钟。去除1375μL的上清液，剩下大约475μL的上清液。将25μL的5M NaCl添加至最终体积为500μL。如果需要，现在可以在同一样品管中合并多个500μL反应。然后，将样品温育过夜以进行解交联。

[0130] 净化

[0131] 每500μL去交联反应，按顺序加入表8中列出的试剂。

[0132] 表8

[0133]

试剂	体积
10% Tween-20	25 μL
10% Triton X-100	25 μL
溶菌酶（250 mg/mL）	5 μL
RNaseA (Qiagen)	2.5 μL

[0134]

混合并在室温下温育 10 分钟	
Qiagen 20 mg/mL 蛋白酶 K	12.5 μL

[0135] 将反应物在37℃下温育30分钟。将170μL的Qiagen缓冲液B2添加到上述每个反应中，然后将反应在50℃下温育30分钟。然后通过苯酚:氯仿:异戊醇和后续的异丙醇沉淀来纯化DNA。

[0136] 测序

[0137] 经过纯化的DNA可以1)直接制备来用于使用Oxford Nanopore Technologies的标准文库制备工作流程/试剂盒的测序(这使得天然DNA修饰(表观基因组学)能够保留在所生成的测序数据中);或者2)通过使用Oxford Nanopore Technologies的PCR测序工作流程/试剂盒由PCR扩增。

[0138] 对于基于纳米孔的测序应用，对于大于2到3kb的PCR产物，然后通过凝胶提取或Bluepippin(SAGE Science)选择最佳PCR循环DNA产物的大小。然后，根据本领域已知的纳米孔文库制备和测序方案执行基于纳米孔的测序。

[0139] 生物信息学工作流程

[0140] 图2提供了适用于目前描述的示例性方法的生物信息学分析工作流程的代表性示意图。在任何描述的方法中，生物信息学分析工作流程可以用于从使用应用于宏基因组样品的方法而得出的测序数据获得宏基因组接触图。更详细地，纳米孔测序数据是通过早先描述的方法生成的，以提供纳米孔测序读数。通过使用BWA-SW(Li H.和Durbin R.(2010) Fast and accurate long-read alignment with Burrows-Wheeler.Transform.Bioinformatics,Epub.[PMID:20080505])，读数首先与染色体序列和染色体外序列(例如质粒)的参考序列集合进行比对。每个比对的读数都被过滤以保留遍历大部分读数的最小比对集合。然后将参考基因组分成大小相等的箱，纳米孔测序读数的每个比对片段被分配一个箱。最后，根据所有纳米孔测序读数计算出箱到箱接触的总数，并在接触图中将其可视化。通过确定哪些染色体与元件有最多的接触，可以将染色体外元件分配给它们的宿主。

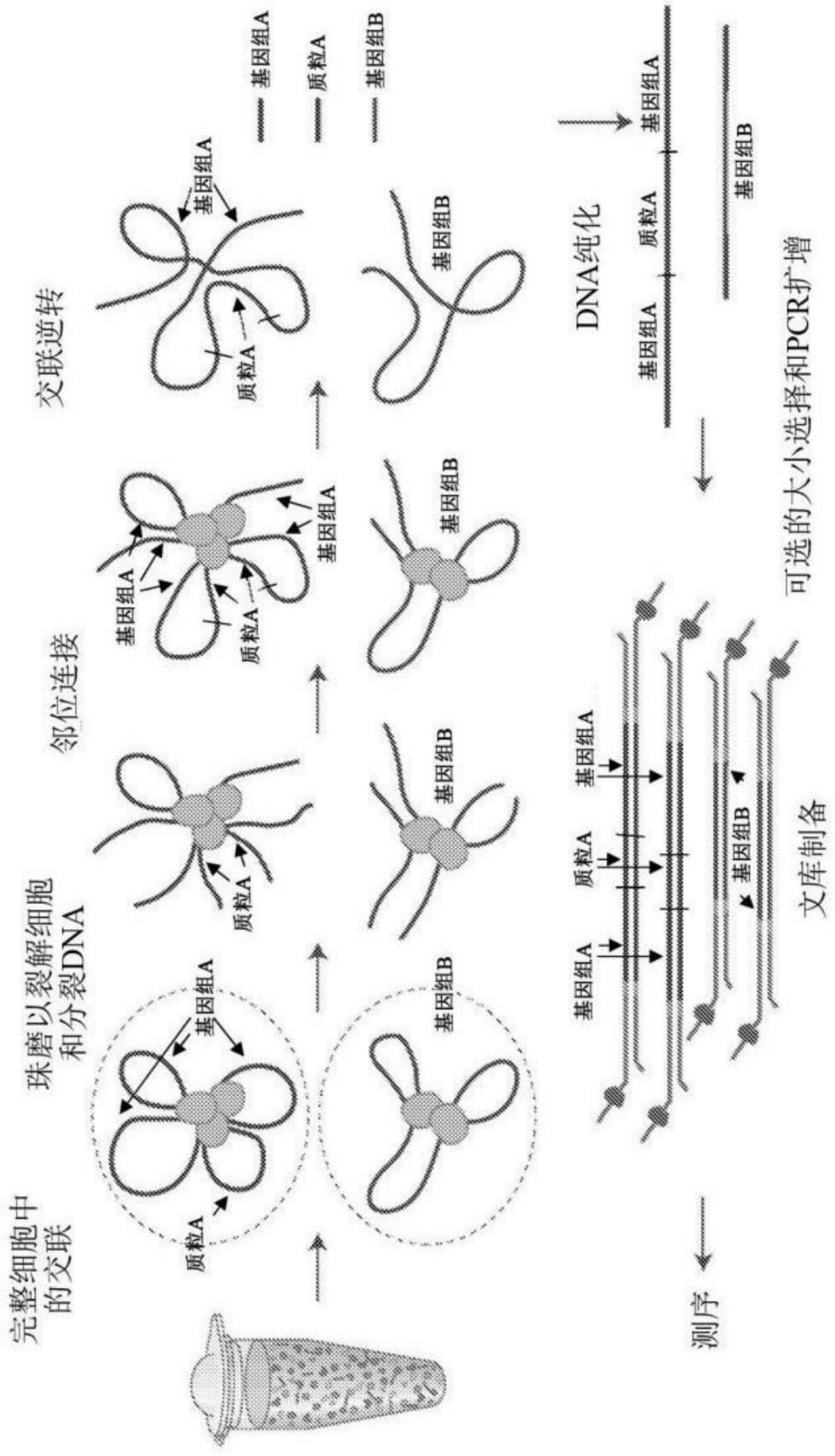
完整细胞中的交联
珠磨以裂解细胞和分裂DNA
邻位连接
交联逆转
DNA纯化
可选的大小选择和PCR扩增
文库制备
测序
质粒A
基因组A
基因组B

图1

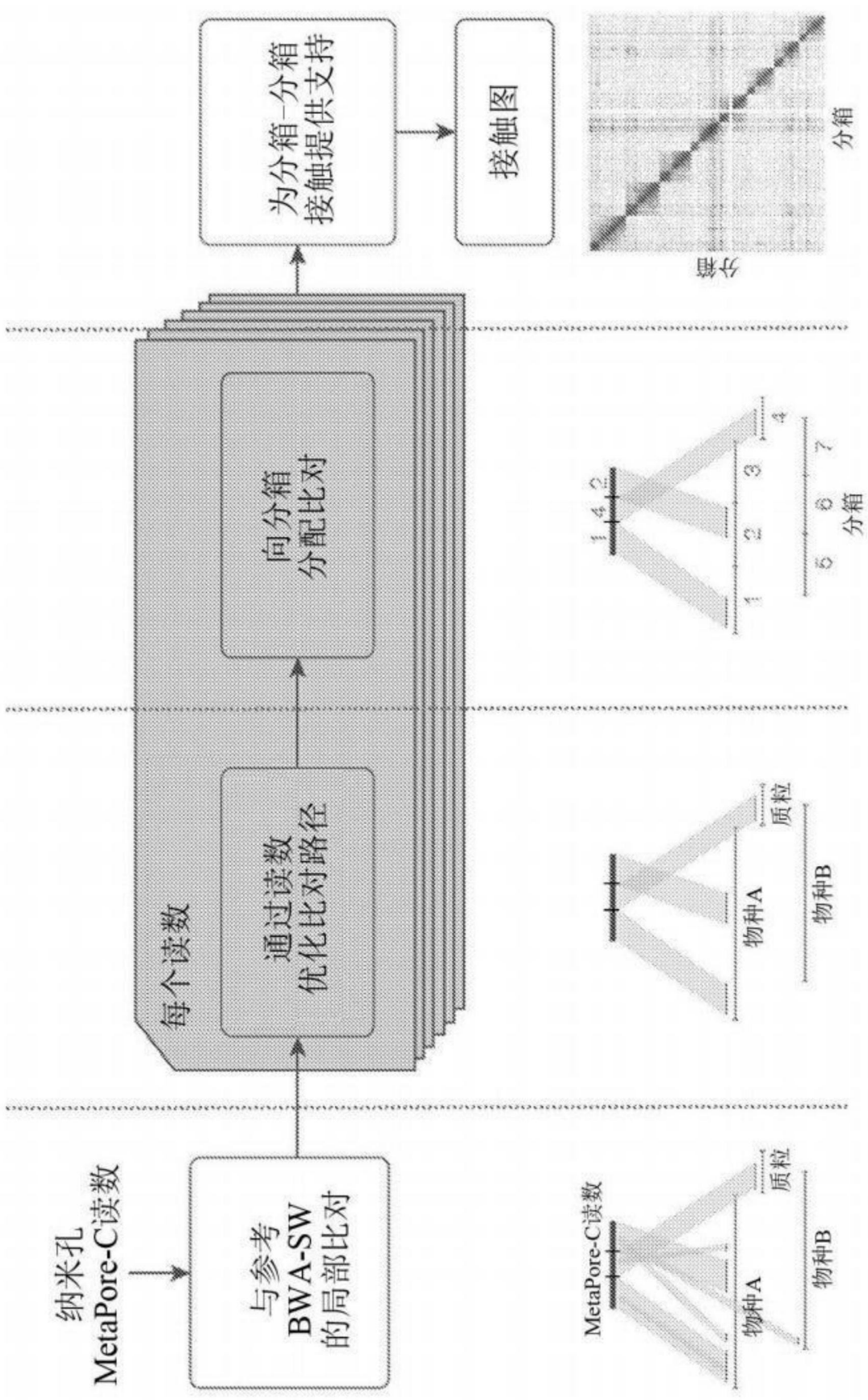
纳米孔
MetaPore-C读数
与参考
BWA-SW
的局部比对
每个读数
通过读数
优化比对路径
向分箱
分配比对
为分箱-分箱
接触提供支持
接触图
分箱
分箱
MetaPore-C读数
物种A
质粒
物种B
物种A
质粒
物种B

图2

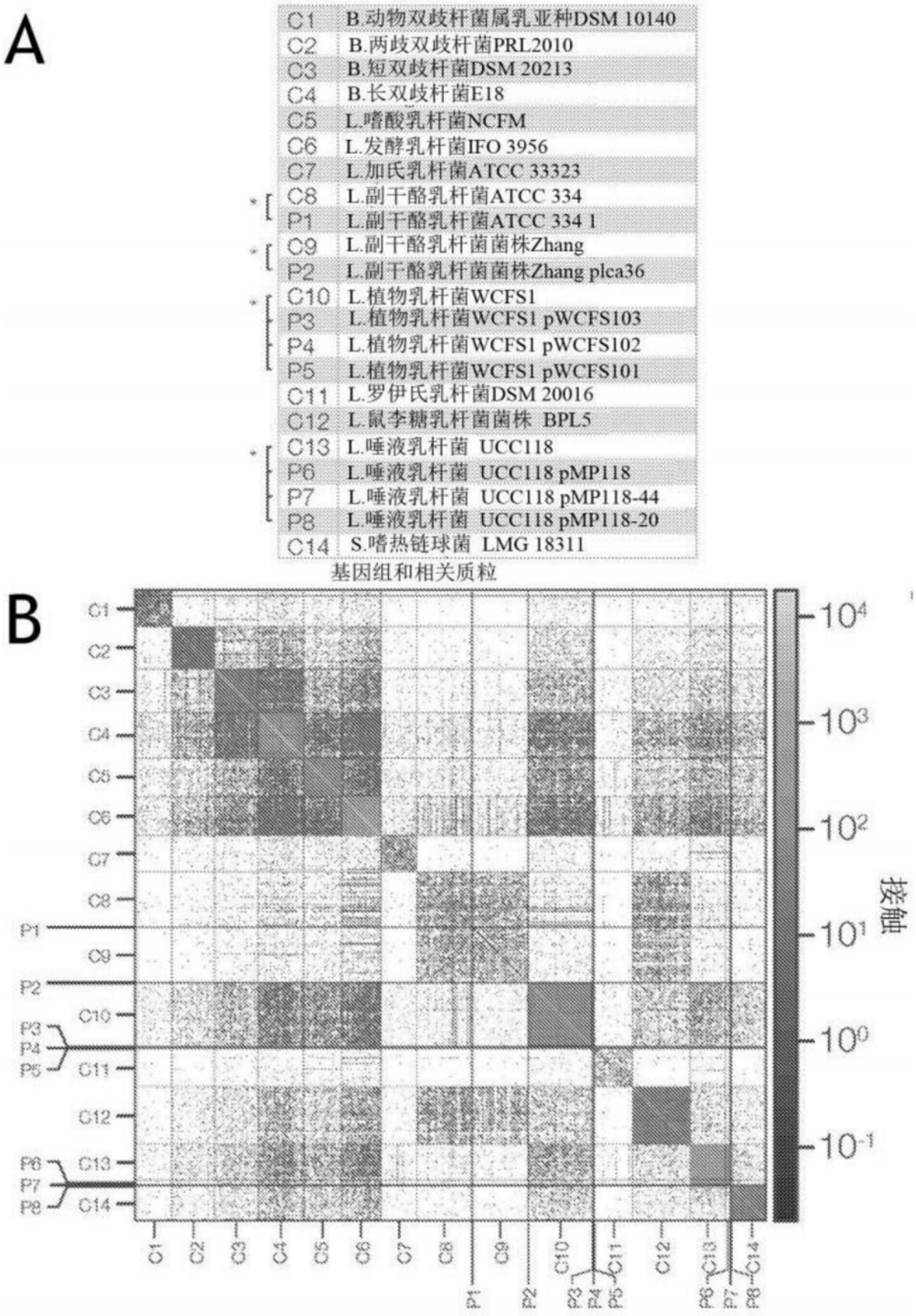
A
C1 B.动物双歧杆菌属乳亚种DSM 10140
C2 B.两歧双歧杆菌PRL2010
C3 B.短双歧杆菌DSM 20213
C4 B.长双歧杆菌E18
C5 L.嗜酸乳杆菌NCFM
C6 L.发酵乳杆菌IFO 3956
C7 L.加氏乳杆菌ATCC 33323
C8 L.副干酪乳杆菌ATCC 334
P1 L.副干酪乳杆菌ATCC 334 1
C9 L.副干酪乳杆菌菌株Zhang
P2 L.副干酪乳杆菌菌株Zhang plca36
C10 L.植物乳杆菌WCFS1
P3 L.植物乳杆菌WCFS1 pWCFS103
P4 L.植物乳杆菌WCFS1 pWCFS102
P5 L.植物乳杆菌WCFS1 pWCFS101
C11 L.罗伊氏乳杆菌DSM 20016
C12 L.鼠李糖乳杆菌菌株 BPL5
C13 L.唾液乳杆菌 UCC118
P6 L.唾液乳杆菌 UCC118 pMP118
P7 L.唾液乳杆菌 UCC118 pMP118-44
P8 L.唾液乳杆菌 UCC118 pMP118-20
C14 S.嗜热链球菌 LMG 18311
基因组和相关质粒
B
接触
10^4
10^3
10^2
10^1
10^0
10^{-1}

图3

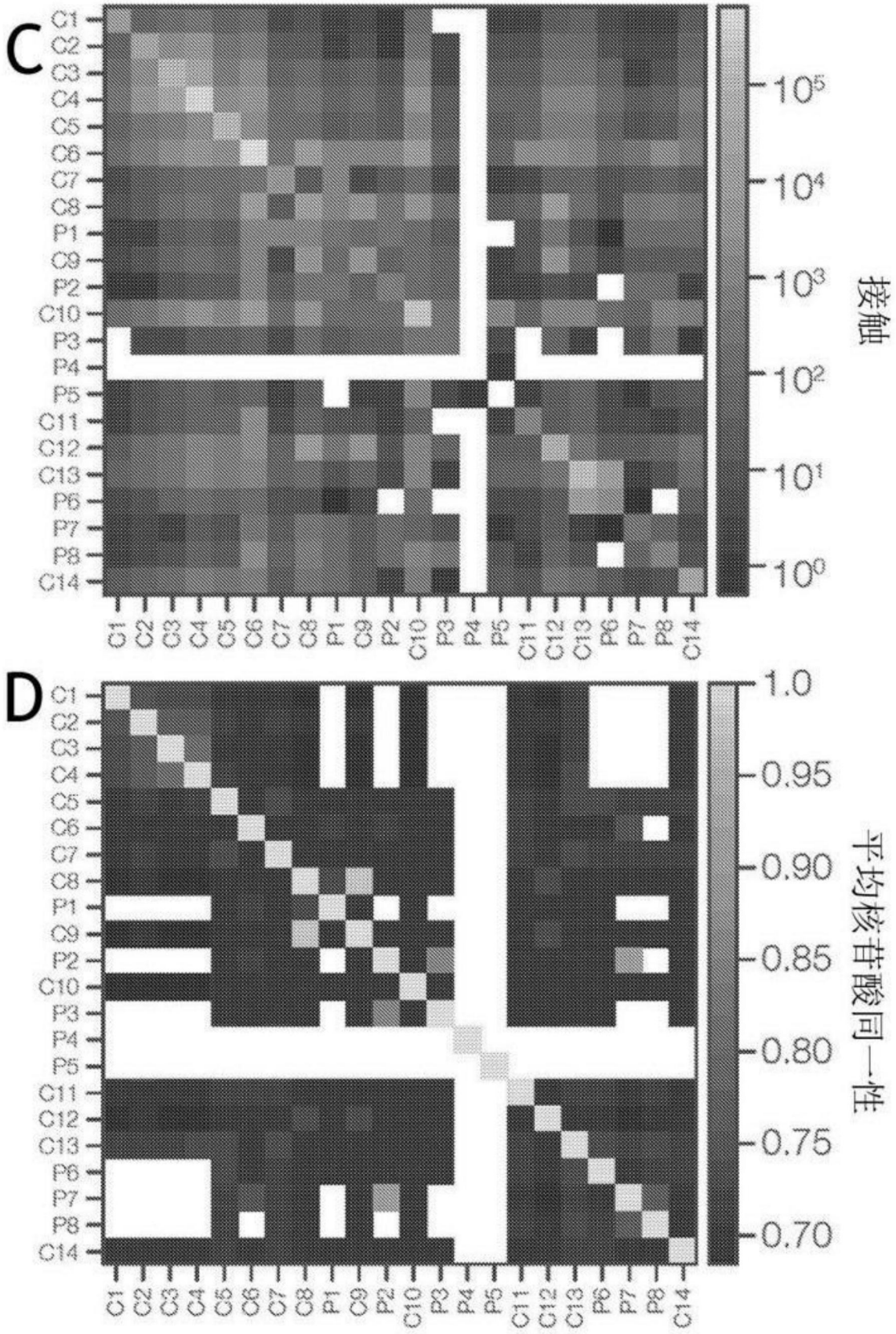
C
C1
C2
C3
C4
C5
C6
C7
C8
P1
C9
P2
C10
P3
P4
P5
C11
C12
C13
P6
P7
P8
C14
10^5
10^4
10^3
10^2
10^1
10^0
接触
D
1.0
0.95
0.90
0.85
0.80
0.75
0.70
平均核苷酸同一性

图3续

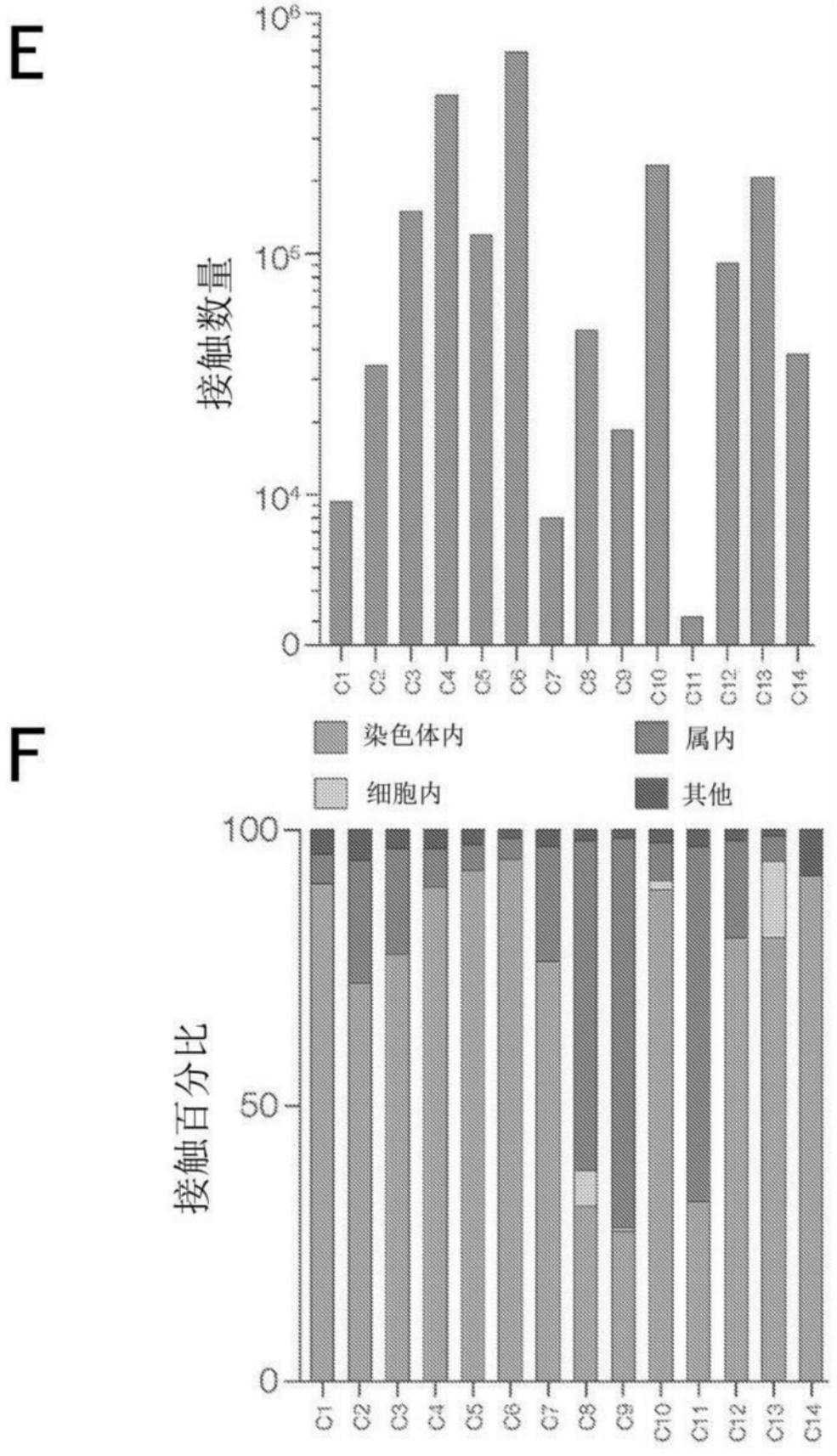
E
接触数量
10^6
10^5
10^4
0
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
F
染色体内
属内
细胞内
其他
接触百分比
100
50
0

图3续