

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2012年6月28日 (28.06.2012)



(10) 国际公布号
WO 2012/083845 A1

- (51) 国际专利分类号:
C12Q 1/68 (2006.01)
- (21) 国际申请号: PCT/CN2011/084303
- (22) 国际申请日: 2011年12月20日 (20.12.2011)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201010600214.4 2010年12月22日 (22.12.2010) CN
- (71) 申请人 (对除美国外的所有指定国): **深圳华大基因科技有限公司 (BGI SHENZHEN CO., LIMITED)** [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼 11F-3, Guangdong 518083 (CN)。 **深圳华大基因研究院 (BGI SHENZHEN)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
- (72) 发明人; 及
(75) 发明人/申请人 (仅对美国): **胡帅星 (HU, Shuaixing)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **武靖华 (WU, Jinghua)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **樊帆 (FAN, Fan)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **陈琳 (CHEN, Lin)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **赵美茹 (ZHAO, Meiru)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **王俊 (WANG, Jun)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **汪建 (WANG, Jian)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。
- (74) 代理人: **北京清亦华知识产权代理事务所 (普通合伙) (TSINGYIHUA INTELLECTUAL PROPERTY LLC)**; 中国北京市海淀区清华园清华大学熙澜院商业楼 301 室, Beijing 100084 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA,

[见续页]

(54) Title: METHODS FOR REMOVAL OF VECTOR FRAGMENTS IN SEQUENCING LIBRARY AND USES THEREOF

(54) 发明名称: 用于除去测序文库中载体片段的方法及其用途

(57) Abstract: Methods for removal of vector fragments in sequencing library, methods for sequencing clone library of genomes and kits for sequencing genomes are provided. Wherein methods for removal of vector fragments in sequencing library include: providing labeled probes which are able to hybridize with vector fragments; hybridizing the probes with the sequencing library, so that the probes and the vector fragments form labeled double-stranded nucleic acid; and utilizing labeled molecular entity on specific binding probes to remove labeled double-stranded nucleic acid, accordingly remove vector fragments in sequencing library.

(57) 摘要:

提供了用于除去测序文库中的载体片段的方法、用于对基因组克隆文库进行测序的方法以及用于基因组测序的试剂盒。其中, 用于除去测序文库中的载体片段的方法包括以下步骤: 提供经过标记的探针, 该探针能够与载体片段进行杂交; 将探针与测序文库进行杂交, 以便探针与载体片段形成带标记的双链核酸; 以及利用特异性结合探针上的标记的分子实体, 去除带标记的双链核酸, 从而除去测序文库中的载体片段。

WO 2012/083845 A1



RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。
- 包括说明书序列表部分(细则 5.2(a))。

用于除去测序文库中载体片段的方法及其用途

优先权信息

本申请请求 2010 年 12 月 22 日向中国国家知识产权局提交的、专利申请号为 201010600214.4 的专利申请的优先权和权益，并且通过参照将其全文并入此处。

技术领域

本发明涉及分子生物学领域，特别是基因组测序领域，具体地，本发明涉及用于除去测序文库中载体片段的方法及其用途。更具体地，本发明提供了一种用于除去测序文库中的载体片段的方法、一种用于对基因组克隆文库进行测序的方法以及一种用于基因组测序的试剂盒。

背景技术

De novo 测序法也称从头测序法，其不需要任何基因序列信息即可对某个物种的基因组进行测序，在测序后，用生物信息学的分析方法对测序获得的序列进行拼接、组装，能够获得该物种的基因组序列图谱，从而能够应用于该物种的其它基因组学、遗传学研究。由此，对生物进行 *De novo* 测序意义重大。

现阶段常利用 Illumina 公司的 DNA 测序平台对生物进行 *De novo* 测序，具体地，主要通过制备该生物的测序文库，然后对其进行测序实现。其中，根据测序片段的大小，基于 Illumina 公司的 DNA 测序平台的测序文库制备方法分为小片段 DNA 的测序文库制备和大片段 DNA 的测序文库制备。而为了解决非单倍体基因组序列杂合度过高、*De novo* 测序序列拼接组装难度大的问题，目前通常采用以 Fosmid 克隆文库为模板构建 DNA 小片段测序文库（在本文中有时也称为构建 Fosmid 小片段测序文库）的方法。其中，Fosmid 克隆文库是通过将某种生物的总 DNA 与 Fosmid 载体一起以重组的形式转移到宿主细胞中，然后通过细胞增殖形成多个克隆的整体而得到的。对于 Fosmid 载体，可插入的基因组 DNA 片段的合适的长度为大约 40 kb。构建 Fosmid 小片段测序文库，是指将一定数量的 Fosmid 克隆混合到一起作为建库模板进行测序文库的制备。

然而，目前构建 Fosmid 小片段测序文库的方法仍有待改进。

发明内容

本发明是基于发明人的下列发现而完成的：

当利用 Solexa DNA 测序仪对由 Fosmid 克隆文库建立的测序文库进行测序时，由于在建库过程中对 Fosmid 克隆进行了打断处理，Fosmid 载体也被打断成了许多 1kb 以下的小片段 DNA，因而测序文库中包含有大量的 Fosmid 载体片段，因此，当对整个测序文库进行测序时，会产生大量的不需要的载体序列数据，并且这些不需要的序列数据还会影响后续的数据分析，造成测序资源浪费、测序成本高、效率低。

本发明旨在至少解决现有技术中存在的技术问题之一。由此，本发明提供了用于除去测序文库中的载体片段的方法及其用途，以便在利用 Solexa DNA 测序仪对测序文库进行测序前，将 Fosmid 载体片段将去除，以避免对 Fosmid 载体进行测序，减少不必要的的数据读取及分析，从而能够显著降低测序成本，提高测序效率。

根据本发明的一个方面，本发明提供了一种用于除去测序文库中的载体片段的方法。根据本发明的实施例，该方法包括以下步骤：

首先，提供经过标记的探针，该探针能够与载体片段进行杂交。根据本发明的实施例，载体是 Fosmid 载体。其中，在本文中所使用的术语“杂交”，是指相互间具有互补序列的两个单链核酸分子在一定条件下（适宜的温度及离子强度等）按碱基互补配对原则退火形成双链核酸的过程，也称为核酸杂交。核酸杂交可以在 DNA-DNA 之间，也可在 DNA-RNA 或 RNA-RNA 之间进行，只要杂交双方存在互补序列，可以进行碱基配对即可。通常，杂交的双方是待测核酸分子和已知核酸分子，在杂交体系中已知序列的核酸分子称作探针（probe）。此外，核酸杂交包括固-液相杂交和液相杂交，其中液相杂交是在溶液中进行的杂交反应，其是指待测核酸分子与已知核酸分子（探针）在溶液中退火形成杂交复合物。在本文中，所使用的术语“载体片段”可以是整个 Fosmid 载体，也可以是将整个 Fosmid 载体打断后形成的所有载体的片段，由此，在本文中的表达方式“探针能够与载体片段进行杂交”表示探针的序列可以与整个 Fosmid 载体序列互补，从而能够与整个 Fosmid 载体进行杂交，也可以是探针的序列与上述所有载体的片段的序列互补，从而能够与所有载体的片段进行杂交。

其次，将探针与测序文库进行杂交，以便探针与载体片段形成带标记的双链核酸。当利用经过标记的探针进行杂交时，在杂交反应结束后，能够利用探针上的标记分离和检测杂交后的双链，其中，标记是本领域内是已知的，可以包括但不限于放射性同位

素或核素、生物素、吖啶翁酯 (acridinium ester) 或 polyA 等。根据探针上的标记的类型, 可以利用本领域内已知的相应的核酸分离和检测方法, 包括但不限于羟基磷灰石 (HAP) 法或亲和吸附法(例如参见 Henegariu O 等人, (1999). Custom fluorescent-nucleotide synthesis as an alternative method for nucleic acid labeling. Nature Biotechnology. 18:345-348; Ezaki T 等人, 1989. Fluorometric Deoxyribonucleic Acid-Deoxyribonucleic Acid Hybridization in Microdilution Wells as an Alternative to Membrane Filter Hybridization in which Radioisotopes Are Used To Determine Genetic Relatedness among Bacterial Strains. Int. J. of Systemic Bacteriology 29 (3): 224-229; Herrington C 等人, 1998. PCR 3: PCR in situ hybridization: a practical approach, Volume 3. Oxford: Oxford University Press, 通过参照将其全文并入本文), 将杂交产物进行分离和检测。在本文中所使用的表达方式“探针与载体片段形成带标记的双链核酸”, 是指探针可以与整个 Fosmid 载体杂交形成带标记的双链核酸, 也可以是探针与上述所有载体的片段杂交带标记的双链核酸。根据本发明的实施例, 将探针与测序文库进行液相杂交反应。根据本发明的一些实施例, 可以将探针与测序文库以大约 1:1 至大约 2:1 的质量比进行杂交反应。根据本发明的具体示例, 在进行杂交反应前, 可以进一步包括向测序文库中添加接头封闭剂。根据本发明的实施例, 接头封闭剂和测序文库的比例可以为大约 0.3 pM/ng-0.8 pM/ng。根据本发明的一个具体示例, 接头封闭剂和测序文库的比例为 0.5 pM/ng。根据本发明的实施例, 杂交反应的时间可以为 1, 4, 16, 24 或更多个小时。

然后利用特异性结合探针上的标记的分子实体, 去除带标记的双链核酸, 从而除去测序文库中的载体片段。根据本发明的实施例, 前述经过标记的探针可以是经过生物素标记的探针, 并且分子实体是亲和素。根据本发明的一些实施例, 分子实体是链霉亲和素, 根据本发明的具体示例, 分子实体形成于磁珠上。

发明人惊奇地发现, 利用根据本发明实施例的用于除去测序文库中的载体片段的方法, 能够有效地除去 Fosmid 小片段测序文库中的载体片段, 从而基于高通量测序平台例如 Illumina Solexa 测序平台, 对处理过的测序文库进行测序后, 能够有效地避免浪费测序资源和产生不需要的测序数据, 最终降低测序成本, 提高测序效率。

具体地, 根据本发明的实施例, 本发明的用于除去测序文库中的载体片段的方法可以包括下列步骤:

- 1) 制备经标记的探针, 该探针能够与载体或其片段杂交;

2) 将探针与测序文库进行杂交反应, 从而使探针与载体或其片段形成带标记的双链核酸;

3) 利用特异性结合探针上的标记的分子实体, 去除带标记的双链核酸, 从而除去测序文库中的载体片段。

根据本发明的实施例, 载体用于构建基因组克隆文库, 并且优选是 Fosmid 载体。根据本发明的具体示例, 探针用生物素进行标记, 并且特异性结合标记的分子实体是亲和素。根据本发明的实施例, 优选分子实体为链霉亲和素, 根据本发明的一个实施例, 链霉亲和素可以形成于磁珠。根据本发明的实施例, 将探针与测序文库进行液相杂交反应。

根据本发明的实施例, 可以通过下列步骤制备生物素标记的探针:

1) 使用生物素化的 dNTP 对载体进行 PCR 扩增, 并纯化回收 PCR 产物;

2) 将 PCR 产物片段化并回收纯化, 从而获得生物素标记的探针。

根据本发明的实施例, 可以利用 Covaris 仪将 PCR 产物片段化。根据本发明的具体示例, 可以将 PCR 产物片段化为大约 200bp-500bp 的片段。根据本发明的一个实施例, 可以将 PCR 产物片段化为大约 300bp 的片段。

根据本发明的实施例, 可以通过在 95°C 下使探针与测序文库变性, 然后在 65°C 下退火来进行杂交反应。根据本发明的具体示例, 可以将探针与测序文库以大约 1:1 至大约 2:1 的质量比进行杂交反应。

根据本发明的实施例, 任选地, 在进行探针与测序文库的杂交之前, 可以向测序文库中加入接头封闭剂 (block)。根据本发明的具体示例, 接头封闭剂是能够与测序文库构建过程中所使用的接头杂交的核酸, 优选, 接头封闭剂的核苷酸序列与接头的核苷酸序列完全互补, 由此, 在进行探针与测序文库的杂交时, 接头封闭剂与接头杂交, 从而避免发生接头与探针的杂交以及接头间配对连接。根据本发明的实施例, 加入的接头封闭剂和文库的比例可以为大约 0.3pM/ng-0.8pM/ng, 例如 0.5pM/ng。当加入的接头封闭剂的量过大时, 文库中的载体片段的去除效率会降低, 然而当加入的接头封闭剂的量过少时, 大量的非载体的文库 DNA 片段会被探针捕获而被去除。根据本发明的实施例, 杂交可以进行 1, 4, 16, 24 或更多个小时。

根据本发明的另一方面, 本发明提供了一种用于对基因组克隆文库进行测序的方法, 其特征在于, 包括以下步骤:

首先, 利用基因组克隆文库构建测序文库, 其中测序文库中包含载体片段。根据本发明的实施例, 基因组克隆文库是 Fosmid 克隆文库, 且载体是 Fosmid 载体。根据本发明的

一些实施例，可以通过下列步骤构建测序文库：将基因组克隆文库中的 DNA 进行片段化，以便获得 DNA 片段；将接头连接至 DNA 片段的两端，以便获得连接产物；将连接产物进行 PCR 扩增，以便获得扩增产物；以及回收纯化扩增产物，该扩增产物构成测序文库。

接着，利用根据本发明实施例的用于除去测序文库中的载体片段的方法除去所述测序文库中的载体片段。根据本发明的实施例，可以在构建测序文库后，或在构建测序文库的过程中，除去测序文库中的载体片段；根据本发明的具体示例，在将基因组克隆文库中的 DNA 进行片段化后，除去测序文库中的载体片段。

最后，对去除载体片段后的测序文库进行测序。根据本发明的实施例，测序的方法不受特别限制。根据本发明的一个实施例，可以利用 Solexa 测序仪进行测序。

发明人惊奇地发现，利用根据本发明实施例的用于对基因组克隆文库进行测序的方法，能够有效地对基因组克隆文库即 Fosmid 小片段测序文库进行测序，由于该方法能够有效地除去 Fosmid 小片段测序文库中的载体片段，因此能够避免测序资源浪费和产生不需要的测序数据，从而能够降低测序成本，提高测序效率，并且所得测序结果准确，数据分析容易。由此，该方法能够有效地应用于现阶段大规模开展的集中化商业性多物种基因组测序，从而能够节省成本，提高效率，进一步，能够将根据本发明实施例的用于对基因组克隆文库进行测序的方法应用到分析、疾病诊断以及个性化(个体化)医疗等新领域。

具体地，根据本发明的实施例，本发明的用于对基因组克隆文库进行测序的方法可以包括下列步骤：

- 1) 利用基因组克隆文库构建测序文库，并除去测序文库中的载体片段；
- 2) 对去除载体片段后的测序文库进行测序。

根据本发明的实施例，基因组克隆文库是 Fosmid 克隆文库，并且载体是 Fosmid 载体。根据本发明的具体示例，可以利用根据本发明的用于除去测序文库中的载体片段的方法来去除测序文库中的载体片段。

根据本发明的一些实施例，测序文库的构建可以通过以下步骤进行：

- 1) 将基因组克隆文库中的 DNA 片段化；
- 2) 将接头连接至经片段化的 DNA 的两端，以便获得添加了接头的 DNA；
- 3) 将添加了接头的 DNA 进行 PCR 扩增，并回收纯化扩增产物，该扩增产物构成测序文库。

根据本发明的实施例，可以在构建测序文库后，除去测序文库中的载体片段。根

据本发明的一些实施例，可以在构建测序文库的过程中，除去测序文库中的载体片段。根据本发明的一个具体示例，可以在片段化基因组克隆文库中的 DNA 后，除去载体片段。

根据本发明的实施例，可以使用 Solexa 测序仪对测序文库进行测序。

根据本发明的又一方面，本发明提供了一种用于基因组测序的试剂盒，其特征在于，包括：载体，其用于构建基因组克隆文库；经标记的探针，其能够与载体或其片段进行杂交；以及分子实体，其能够特异性结合探针上的标记。根据本发明的实施例，载体是 Fosmid 载体。根据本发明的一些实施例，探针是经过生物素标记的探针，并且分子实体是亲和素。根据本发明的具体示例，分子实体是链霉亲和素。根据本发明的一个实施例，分子实体为形成于磁珠上的链霉亲和素。本领域的技术人员可以理解，本发明的试剂盒中还可以包括基因组克隆文库构建、测序等所需的其它试剂，根据本发明的实施例，该试剂盒中可以进一步包括接头和接头封闭剂。关于接头封闭剂的特征及作用，前面已作详细描述，在此不再赘述。

发明人发现，根据本发明实施例的试剂盒能够有效地用于基因组测序，具体地，利用试剂盒中的载体能够有效地构建基因组克隆文库，进而利用试剂盒中的经标记的探针以及分子实体能够有效地将文库中的载体片段去除，从而基于高通量测序平台，能够有效地对去除载体片段的文库进行基因组测序。由此，将本发明的试剂盒应用于基因组测序，能够有效地避免测序资源浪费和产生不需要的测序数据，从而能够降低测序成本，提高测序效率，并且所得测序结果准确，数据分析容易。

具体地，根据本发明的实施例，本发明的用于基因组测序的试剂盒可以包括：用于构建基因组克隆文库的载体，能够与载体或其片段杂交的经标记的探针以及能够特异性结合探针的标记的分子实体。

根据本发明的实施例，载体是 Fosmid 载体。根据本发明的一些实施例，探针用生物素进行标记，并且特异性结合标记的分子实体是亲和素，根据本发明的具体示例，优选分子实体为链霉亲和素。根据本发明的一个实施例，链霉亲和素可以形成于磁珠上。根据本发明的实施例，试剂盒还可以包括其他试剂，包括但不限于接头和接头封闭剂。

需要说明的是，根据本发明实施例的用于除去测序文库中的载体片段的方法及其用途是本申请的发明人经过艰苦的创造性劳动和优化工作而完成的。

本发明的附加方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本发明的实践了解到。

具体实施方式

下面将结合实施例对本发明的方案进行解释。本领域技术人员将会理解，下面的实施例仅用于说明本发明，而不应视为限定本发明的范围。实施例中未注明具体技术或条件的，按照本领域内的文献所描述的技术或条件（例如参考 J.萨姆布鲁克等著，黄培堂等译的《分子克隆实验指南》，第三版，科学出版社）或者按照产品说明书进行。所用试剂或仪器未注明生产厂商者，均为可以通过市购获得的常规产品，例如可以采购自 Illumina 公司。

实施例 1

1、制备探针

以 Fosmid 载体（其序列如 SEQ ID NO: 1 所示）为模板，通过 PCR 反应制备探针。其中，PCR 的反应体系如下：

LA Taq	0.5 μ L
10 \times LA 缓冲液	5 μ L
dNTP 混合溶液（1 μ M）*	20 μ L
用于制备探针的正向引物（10 μ M）	5 μ L
用于制备探针的反向引物（10 μ M）	5 μ L
模板（50 μ M）	1 μ L
水	13.5 μ L
总体积	50 μ L

注：*表示所使用的 dNTP 混合溶液是以 15:85 的比例配置的生物素-dNTP 与普通 dNTP 的混合溶液，并且其终浓度是 1 μ M。

其中，用于制备探针的正向引物为：5'-CCTGGGGTGCCTAATGAGTG-3' (SEQ ID NO: 2)。用于制备探针的反向引物：5'-CGTCGTTTTACAACGTCGTGA-3' (SEQ ID NO: 3)。

PCR 反应条件为：95 $^{\circ}$ C，2 分钟；12 个循环的 95 $^{\circ}$ C，30 秒、65 $^{\circ}$ C，30 秒、72 $^{\circ}$ C，8 分钟；72 $^{\circ}$ C，10 分钟；4 $^{\circ}$ C 保存。

PCR 完成后，利用 DNA clean and ConcentratorTM-25 试剂盒对 PCR 产物进行纯化，然后利用 Covaris S2 仪将 PCR 产物进行打断，以便获得 DNA 片段，然后通过凝胶电泳检测确定 DNA 片段的大小。检测结果显示，DNA 片段的大小主要为大约 300bp。然后利用

MinElute PCR 纯化试剂盒纯化回收 DNA 片段，并将其溶于 20 μ l 洗脱缓冲液中，从而获得生物素标记的探针。然后，利用 Qubit (HS) 定量探针的浓度，备用。

2、探针与文库杂交

首先，利用 CopyControl™ HTP Fosmid 文库构建试剂盒(Epicentre, USA)，按照生产商的详细说明，制备小麦条锈菌(*Puccinia striiformis f.sp tritici*)基因组 DNA 的 Fosmid 克隆文库。

接着，利用 Illumina 公司的混合样品制备寡核苷酸试剂盒 (Multiplexing sample preparation oligonucleotide kit, PE-400-1002) 构建小麦条锈菌 DNA 的测序文库，备用。

然后，将上述探针和小麦条锈菌 DNA 的测序文库 (在本实施例中有时将其简称为“文库”) 按照下表中的配比制备两份杂交体系，并分别命名为样品 1 和样品 2:

文库	120 ng
2 × SC 杂交缓冲液	7.5 μ l
SC 杂交组件 A	3 μ l
探针	120 ng
接头封闭剂 1	0.06 nM
接头封闭剂 2	0.06 nM
水	补足至 15 μ l

注: SC 杂交缓冲液和 SC 杂交组件 A 来源于序列捕获杂交试剂盒。

其中，在各杂交体系中，文库的用量为 120ng，探针的用量为 120ng (通过 Qubit 进行定量)，并且杂交体系中添加有接头封闭剂。该接头封闭剂来自标签测序引物试剂盒 (Illumina 公司) 和混合 Rd2 测序引物 (Multiplexing Rd2 Sequencing Primer 试剂盒, Illumina 公司)，其中接头封闭剂的序列如下所示:

接头封闭剂 1: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACG ACGCTCTTCCGATCT-3' (SEQ ID NO: 4);

接头封闭剂 2: 5'-ACAAGCAGAAGACGGCATAACGAGATAAGCAATGGTGA CTGGA GTTCAGACGTGTGCTCTTCCGATCT-3' (SEQ ID NO: 5)。

在各杂交体系中，两种接头封闭剂的用量都是 0.06 nM，从而接头封闭剂与文库的量的比值为 0.5 pM/ng。

然后，将上述制备好的两份杂交体系进行杂交，其中杂交条件为：在 95 $^{\circ}$ C 下变性 10

分钟，然后在 65℃ 下按照下表中的指定时间进行杂交，从而获得杂交产物，备用。

样品	探针: 文库	是否加接头封闭剂	杂交时间 (小时)
1	1:1	是	24
2	1:1	是	4

3、载体片段的捕获和分离

根据制造商的说明书，利用链霉亲和素磁珠 (M280 磁珠) 捕获和分离上述杂交产物中能够与探针杂交的文库 DNA (即，载体片段)，然后将来源于样品 1 和 2 的去除了载体片段的文库 DNA 分别进行收集，并标记为 24-1-D 和 4-12-D，并测定二者的浓度，结果如下所示：

样品名称	24-1-D	4-12-D
浓度(nM/L)	3.92	6.08
回收率	0.228967	0.355133

此外，分别收集来自样品 1 和 2 的与探针杂交的文库 DNA (即，载体片段)，并标记为 24-1-V 和 4-12-V。

根据制造商的说明书，利用 QIAquick PCR 纯化试剂盒分别将上述获得的文库 DNA，即 24-1-D、4-12-D、24-1-V 以及 4-12-V 进行纯化，备用。

4、文库 DNA 的扩增

利用 PFX 酶将上述获得的文库 DNA 进行 PCR 扩增，其中 PCR 反应体系如下：

DNA	38μl
PFX 酶	1μl
PFX 缓冲液(10×)	5μl
dNTP(10mM)	2μl
MgSO ₄	2μl
用于文库 DNA 扩增的正向引物(10 pm/μl)	1μl
用于文库 DNA 扩增的反向引物(10 pm/μl)	1μl
总体积	50μl

其中，用于文库 DNA 扩增的正向引物为：5'- CAAGCAGAAGACGGC ATACGA -3' (SEQ ID NO: 6)，用于文库 DNA 扩增的反向引物：5'- AATGATACGGCGACCACCGAGATC -3'

(SEQ ID NO: 7)。

PCR 反应条件为：94℃，2 分钟；12 个循环的 94℃，15 秒、58℃，30 秒、72℃，30 秒；72℃，5 分钟；4℃ 保存。由此，获得扩增产物。

然后，将扩增产物纯化后溶于 30μl 超纯水中，并利用 Nanodrop 测定获得的扩增产物的浓度，结果如下：

样品名称	24-1-D	4-12-D	24-1-V	4-12-V
浓度 (ng/μl)	369.9	381.9	290.5	320.3

5、载体去除情况的分析

在本实施例中，通过 Q-PCR 来分析利用本发明的用于除去测序文库中的载体片段方法去除文库中载体片段的效果。在分析时，选择未经本发明的方法处理的样品作为对照。为确定去除载体片段的效果，将未经处理的样品和经处理的样品稀释到同一浓度，并进行 Q-PCR 以检测这两种样品中载体片段的含量。然后通过下列公式，利用未经处理的样品和经处理的样品的 Ct 值来计算富集度：富集度 = $E^{Ct(\text{经处理的样品}) - Ct(\text{对照样品})}$ ，其中，E 表示扩增效率，例如，当扩增效率为 100% 时，E = 2，表示扩增效率为每个扩增循环使得扩增子的拷贝数增加 2 倍。富集度表示对照样品中载体片段的量相对于经处理的样品中载体片段的量的倍数，因此，由富集度的值即可得出载体去除的效果，例如当富集度为 N 时，其表示经过本发明的方法处理的样品去除了 $1 - 1/N$ 的载体片段。

然后，以去除了载体片段的文库 DNA(即上述 24-1-D 和 4-12-D)以及未去除载体片段的文库 DNA(即小麦条锈菌 DNA 的测序文库)为模板，进行 Q-PCR 反应以检测各文库中载体片段的含量。其中，Q-PCR 反应体系如下：

SYBR Premix	10 μl
用于 Q-PCR 的正向引物(10μm)	0.4 μl
用于 Q-PCR 的反向引物(10μm)	0.4 μl
Rox Reference Dye II	0.4 μl
模板 DNA (1ng/μl)	1 μl
双蒸水	7.8 μl
总体积	20 μl

注：SYBR Premix 和 Rox Reference Dye II 来源于 SYBR® Premix Ex Taq™。

其中,用于 Q-PCR 的正向引物为: 5'-TTgTTCCACgCCTGCTGAgTTGT-3' (SEQ ID NO: 8), 用于 Q-PCR 的反向引物为: 5'-ATCCCGAATTTgCTCCTCCATCCAC-3' (SEQ ID NO: 9)。

Q-PCR 反应条件为: 95℃, 30 秒; 40 个循环的 95℃, 15 秒、60℃, 1 分钟。

Q-PCR 的结果如下:

样品	杂交时间	去除了载体片段的文库 DNA 的平均 Ct 值	未去除载体片段的文库 DNA 的平均 Ct 值	ΔCt	富集度
24-1-D	24 小时	14.34798	11.60962	2.73836	6.22614425
4-12-D	4 小时	14.09166	11.65803	2.43363	5.079686748

从上述结果可以看出, 载体片段的富集度都大于 5, 这表明经本发明的方法处理的样品中载体片段的含量小于对照样品的 1/5, 从而表明本发明的方法去除了初始文库中至少 80% 的载体片段。

实施例 2

1、制备探针

利用实施例 1 中所采用的制备探针方法的制备探针, 备用。

2、探针与文库杂交

首先, 利用来自一名中国成年男性的血液基因组 DNA, 构建 Fosmid 克隆文库, 接着根据制造商的说明书, 利用混合样品制备寡核苷酸试剂盒 (Multiplexing sample preparation oligonucleotide kit, Illumina 公司, PE-400-1002) 构建测序文库, 称为 YH Fosmid 测序文库。然后, 将上述制备的探针和 YH Fosmid 测序文库 (在本实施例中有时将其简称为“文库”) 按照下表中的配比制备杂交体系并进行杂交:

文库	50 ng
2 × SC 杂交缓冲液	7.5 μ l
SC 杂交组件 A	3 μ l
探针	100 ng
接头封闭剂 1	0.025 nM
接头封闭剂 2	0.025 nM

水	补足至 15 μ l
---	----------------

其中，在上述杂交体系中，文库与探针的质量比为 1:2，并且接头封闭剂与实施例 1 的接头封闭剂相同。杂交条件为：在 95℃ 下进行变性 10 分钟，然后在 65℃ 下进行杂交 24 小时，从而获得杂交产物，备用。

3、载体片段的捕获和分离

利用实施例 1 中采用的捕获和分离载体片段的方法，对所得的杂交产物中能够与探针杂交的文库 DNA，即载体片段，进行捕获和分离，然后收集去除了载体片段的文库 DNA，并标记为 YH-DNA-1，并测定其浓度，结果如下所示：

样品名称	YH-DNA-1
浓度 (nM/L)	2.28
回收率	39.95%

收集与探针杂交的文库 DNA(即，载体片段)，并标记为 YH-Vector-1。

根据制造商的说明书，利用 QIAquick PCR 纯化试剂盒纯化获得的文库 DNA，备用。

4、利用实施例 1 中采用的将文库 DNA 进行 PCR 扩增的方法，对上述获得的文库 DNA 进行 PCR 扩增，以便获得扩增产物，然后将扩增产物纯化后溶于 30 μ l 超纯水中，并利用 Nanodrop 测定获得的扩增产物的浓度，结果如下：

样品	YH-DNA-1	YH-Vector-1
浓度 (ng/ul)	238.1	246.2

5、载体去除情况的分析

利用实施例 1 中采用的对利用本发明的用于除去测序文库中的载体片段方法去除文库中载体片段的效果进行分析的方法，通过 Q-PCR 来分析本实施例中去除文库中载体片段的效果，Q-PCR 的结果如下：

样品	探针和文库的质量比	去除了载体片段的文库 DNA 的平均 Ct 值	未去除载体片段的文库 DNA 的平均 Ct 值	Δ Ct	富集度
YH-DNA-1	2:1	14.28705311	11.633825	2.6536	5.869373

从上述结果可以看出，载体片段的富集度大于 5，这表明经本发明的方法处理的样品

中载体片段的含量小于对照样品的 1/5。因此，本发明的方法去除了初始文库中至少 80% 的载体片段。

工业实用性

本发明的用于除去测序文库中的载体片段的方法、用于对基因组克隆文库进行测序的方法以及用于基因组测序的试剂盒，能够有效地除去 Fosmid 小片段测序文库中的载体片段，进而能够应用于基因组克隆文库的测序，从而能够降低测序成本，提高测序及数据分析的效率。

尽管本发明的具体实施方式已经得到详细的描述，本领域技术人员将会理解。根据已经公开的所有教导，可以对那些细节进行各种修改和替换，这些改变均在本发明的保护范围之内。本发明的全部范围由所附权利要求及其任何等同物给出。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示意性实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

权利要求书

- 1、一种用于除去测序文库中的载体片段的方法，其包括以下步骤：
提供经过标记的探针，所述探针能够与所述载体片段进行杂交；
将所述探针与所述测序文库进行杂交，以便所述探针与所述载体片段形成带标记的双链核酸；以及
利用特异性结合探针上的标记的分子实体，去除所述带标记的双链核酸，从而除去所述测序文库中的载体片段。
- 2、根据权利要求1所述的方法，其特征在于，所述载体是 Fosmid 载体。
- 3、根据权利要求1所述的方法，其特征在于，所述经过标记的探针是经过生物素标记的探针，并且所述分子实体是亲和素。
- 4、根据权利要求3所述的方法，其特征在于，所述分子实体是链霉亲和素。
- 5、根据权利要求4所述的方法，其特征在于，所述分子实体形成于磁珠上。
- 6、根据权利要求1所述的方法，其特征在于，将所述探针与所述测序文库进行液相杂交反应。
- 7、根据权利要求1所述的方法，其特征在于，将所述探针与所述测序文库以大约 1:1 至大约 2:1 的质量比进行杂交反应。
- 8、根据权利要求1所述的方法，其特征在于，在进行杂交反应前，进一步包括向测序文库中添加接头封闭剂。
- 9、根据权利要求8所述的方法，其特征在于，所述接头封闭剂和所述测序文库的比例为大约 0.3 pM/ng-0.8 pM/ng。
- 10、根据权利要求9所述的方法，其特征在于，所述接头封闭剂和所述测序文库的比例为 0.5 pM/ng。
- 11、根据权利要求1所述的方法，其特征在于，所述杂交反应的时间为 1，4，16，24 或更多个小时。
- 12、一种用于对基因组克隆文库进行测序的方法，其特征在于，包括以下步骤：
利用基因组克隆文库构建测序文库，其中所述测序文库中包含载体片段；
利用根据权利要求1-11任一项所述的方法除去所述测序文库中的载体片段；以及
对去除载体片段后的测序文库进行测序。
- 13、根据权利要求12所述的方法，其特征在于，所述基因组克隆文库是 Fosmid 克隆

文库，且所述载体是 Fosmid 载体。

14、根据权利要求 12 所述的方法，其特征在于，所述构建测序文库是通过下列步骤进行的：

将所述基因组克隆文库中的 DNA 进行片段化，以便获得 DNA 片段；

将接头连接至所述 DNA 片段的两端，以便获得连接产物；

将所述连接产物进行 PCR 扩增，以便获得扩增产物；以及

回收纯化所述扩增产物，所述扩增产物构成所述测序文库。

15、根据权利要求 12 所述的方法，其特征在于，在构建测序文库后，或在构建测序文库的过程中，除去所述测序文库中的载体片段；

16、根据权利要求 12 所述的方法，其特征在于，在将所述基因组克隆文库中的 DNA 进行片段化后，除去所述测序文库中的载体片段。

17、根据权利要求 12 所述的方法，其特征在于，利用 Solexa 测序仪进行测序。

18、一种用于基因组测序的试剂盒，其特征在于，包括：

载体，所述载体用于构建基因组克隆文库；

经标记的探针，所述探针能够与所述载体片段进行杂交；以及

分子实体，所述分子实体能够特异性结合所述探针上的标记。

19、根据权利要求 18 所述的试剂盒，其特征在于，所述载体是 Fosmid 载体。

20、根据权利要求 18 所述的试剂盒，其特征在于，所述探针是经过生物素标记的探针，并且所述分子实体是亲和素。

21、根据权利要求 20 所述的试剂盒，其特征在于，所述分子实体是链霉亲和素。

22、根据权利要求 21 所述的试剂盒，其特征在于，所述分子实体为形成于磁珠上的链霉亲和素。

23、根据权利要求 18 所述的试剂盒，其特征在于，进一步包括接头和接头封闭剂。

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/084303

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, SIPOABS, DWPI, CNTXT, WOTXT, EPTXT, USTXT, CNKI, GOOGLE SCHOLAR, Elsevier Science, Pub Med: DE NOVE sequencing method, de novo sequencing method, FOSMID cloning library, FOSMID small segment sequencing library, blocking reagent, FOSMID, BLOCK, probe, hybrid+, carrier, vector, vehicle

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2009325239 A1 (LOK, S.), 31 December 2009 (31.12.2009), the whole document	1-23
A	KIM et al., Cloning, Sequencing, and Characterization of the Pradimicin Biosynthetic Gene Cluster of Actinomadura hibisca P157-2, J. MICROBIOL. BIOTECHNOL, 2007, vol. 17, no. 5, pages 830-839	1-23
A	DONG, Xiaoyi et al., Screening, Identifying and Function Analysis of Polyketide Synthase I Cluster from the Environmental Strain X-2 Which Produce Macrolactins, MICROBIOLOGY, 2008, vol. 35, no. 9, pages 1367-1372	1-23

Further documents are listed in the continuation of Box C. See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&” document member of the same patent family</p>
---	---

Date of the actual completion of the international search 14 March 2012 (14.03.2012)	Date of mailing of the international search report 29 March 2012 (29.03.2012)
---	---

<p>Name and mailing address of the ISA: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451</p>	<p>Authorized officer ZHAO, Yanhao Telephone No.: (86-10) 62411043</p>
--	---

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2011/084303

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
US 2009325239 A1	31.12.2009	EP 2310524 A1	20.04.2011
		WO 2010003316 A1	14.01.2010
		CN 102165073 A	24.08.2011

A. 主题的分类		
C12Q 1/68 (2006.01) i		
按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类		
B. 检索领域		
检索的最低限度文献(标明分类系统和分类号)		
IPC:C12Q		
包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))		
CNABS,SIPOABS,DWPI,CNTXT,WOTXT,EPTXT,USTXT,CNKI,GOOGLE SCHOLAR, Elsevier Science, PubMed: DE NOVE 测序法, 从头测序法, FOSMID 克隆文库, FOSMID 小片段测序文库, 福斯质粒, 封闭剂, 探针, 杂交, FOSMID, BLOCK, probe, hybrid+, carrier, vector, vehicle		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	US2009325239 A1 (LOK S) 31.12 月 2009 (31.12.2009),全文	1-23
A	Kim 等, Cloning, Sequencing, and Characterization of the Pradimicin Biosynthetic Gene Cluster of Actinomadura hibisca P157-2, J. Microbiol. Biotechnol., 2007, 第 17 卷, 第 5 期, 第 830-839 页	1-23
A	董晓毅等, 产 Macrolactins 的海洋细菌 X-2 中 I 型 PKS 基因簇的筛选鉴定与功能分析, 微生物学通报, 2008, 第 35 卷, 第 9 期, 第 1367-1372 页	1-23
<input type="checkbox"/> 其余文件在 C 栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件		“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件
国际检索实际完成的日期 14.3 月 2012(14.03.2012)		国际检索报告邮寄日期 29.3 月 2012 (29.03.2012)
ISA/CN 的名称和邮寄地址: 中华人民共和国国家知识产权局 中国北京市海淀区蓟门桥西土城路 6 号 100088 传真号: (86-10)62019451		受权官员 赵彦豪 电话号码: (86-10) 62411043

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2011/084303

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
US2009325239 A1	31.12.2009	EP2310524 A1	20.04.2011
		WO2010003316 A1	14.01.2010
		CN102165073 A	24.08.2011