(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 9 月 7 日 (07.09.2012)

WIPO | PCT

(10) 国际公布号
WO 2012/116661 A1

(51) 国际专利分类号:
C12Q 1/68 (2006.01)

(21) 国际申请号: PCT/CN2012/071893

(22) 国际申请日: 2012 年 3 月 2 日 (02.03.2012)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201110050238.1 2011 年 3 月 3 日 (03.03.2011) 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): **刘琳 (LIU, 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)。 **何毅敏 (HE, Yimin)** [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。 **杨焕明 (YANG, Huanming)** [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照澜院商业楼 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, 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))。

(54) **Title**: DNA TAG AND USE THEREOF

(54) **发明名称** : DNA 标签及其应用

(57) **Abstract**: Provided are a DNA tag, a DNA tag adapter and a PCR tag primer for constructing a DNA tag library. Also provided are a DNA tag library and a method and kit for constructing said library. Also provided is a method of determining sequence information for one or multiple DNA samples.

(57) **摘要**:

本发明提供了用于构建 DNA 标签文库的 DNA 标签、DNA 标签接头、PCR 标签引物。本发明还提供了一种 DNA 标签文库及其制备方法以及用于构建 DNA 标签文库的试剂盒。本发明还提供了一种确定一种或多种 DNA 样品序列信息的方法。

WO 2012/116661 A1

DNA 标签及其应用

优先权信息

本申请请求 2011 年 3 月 3 日向中国国家知识产权局提交的、专利申请号为 201110050238.1 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

本发明涉及核酸测序技术领域，特别是多重核酸测序技术领域。具体地，本发明涉及 DNA 标签及其应用，更具体地，本发明提供了用于 DNA 测序的 DNA 标签、寡核苷酸、DNA 标签文库及其制备方法、确定 DNA 样品序列信息的方法、确定多种 DNA 样品序列信息的方法以及用于构建 DNA 标签文库的试剂盒。

背景技术

多重测序（Multiplex sequencing）技术，是在每个文库的接头序列和插入片段之间加入一段用于识别样品来源的 DNA 标签（也称为 Index），并利用标签特异性引物对多个文库样品混合测序，以便确定各文库以及其对应的 DNA 标签的序列，然后基于各标签序列区分不同的文库，从而确定各文库的 DNA 序列。采用多重测序技术，能够实现对多个文库样品混合测序，从而能够避免测序资源浪费。

然而，目前的多重核酸测序技术，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：

目前的多重核酸测序技术，一般是利用同样长度的 DNA 标签，混合不同文库同时进行测序，采用该技术对多个文库的混合物进行测序时，常会因为标签中碱基的偏向性而导致检测插入片段时光强参数波动，从而影响输出数据的质量，导致数据结果不可信，不能真实的反映样品的相关信息，同时也导致实验结果重复性低。

本发明旨在至少解决现有技术中存在的技术问题之一。为此，本发明提供了一组能够用于多重核酸测序的 DNA 标签（在本文中，有时也简单地称为“标签”）及其应用。

根据本发明的一个方面，本发明提供了一组分离的 DNA 标签。根据本发明的实施例，该组分离的 DNA 标签由 SEQ ID NO:（3N-2）所示的核苷酸构成，其中 N=1-6 的任意整数。在本说明书中，这些 DNA 标签分别被命名为 IndexN，其中 N=1-6 的任意整数，其序列如下表 1 所示。利用上述根据本发明实施例的 DNA 标签，通过将 DNA 标签与样品 DNA 或其等同物相连，可以精确地表征 DNA 的样品来源。由此，能够将上述 DNA 标签有效地应用到目前的多重核酸测序技术中，即利用上述 DNA 标签，可以同时构建多种样品的 DNA 标签文库（在本文中，有时也称为“标签文库”），从而可以通过将来源于不同样品的 DNA 标签文库混合之后进行测序，并且能够基于 DNA 标签对 DNA 标签文库的 DNA 序列进行

分类，从而可以获得多种样品的DNA序列信息，由此可以充分利用高通量的测序技术，例如利用Solexa测序技术，同时对多种DNA标签文库进行测序，从而提高DNA标签文库的测序效率和通量。发明人惊奇地发现，相对于目前的多重核酸测序技术，利用根据本发明实施例的DNA标签构建多种样品的DNA标签文库，并进行多重核酸测序，能够精确地对多种DNA标签文库进行区分，并且能够有效地减少数据产出偏向性问题，所得到的测序数据可靠，稳定性和可重复性非常好。

根据本发明的另一方面，本发明提供了用于将上述DNA标签引入样品DNA或其等同物中，从而能够应用于多重核酸测序技术的一组分离的寡核苷酸。根据本发明的实施例的一组分离的寡核苷酸，具有第一链和第二链，所述第一链分别由SEQ ID NO:（3N-1）所示的核苷酸构成，所述第二链分别由SEQ ID NO:（3N）所示的核苷酸构成，其中，对于相同的寡核苷酸，其第一链和第二链的N取值相同，并且N=1-6的任意整数。根据本发明的实施例，这些寡核苷酸（在本说明书中，有时也称为“DNA标签接头”、“标签接头”）分别具有如前所述的根据本发明实施例的DNA标签，因而可以通过连接反应，将相应的DNA标签引入到DNA或其等同物中。与DNA标签的命名方法类似，在本说明书中，与DNA标签IndexN相对应的寡核苷酸（DNA标签接头）被命名为IndexN adapter，其中N=1-6的任意整数，进一步，DNA标签接头的第一链（在本文中，有时也称为“有义序列”）和第二链（在本文中，有时也称为“反义序列”）分别被命名为IndexN adapter F和IndexN adapter R，其中N=1-6的任意整数，其序列如下表1所示（表中所示序列方向均是5’－3’方向）。根据本发明的实施例，可以通过将有义序列IndexN adapter F和其相应的反义序列IndexN adapter R进行等摩尔退火处理而形成相应的DNA标签接头。

表1 DNA标签（IndexN）和DNA标签接头（IndexN_adapter）序列

名称	序列(SEQ ID NO:)
Index1	GATGGA(1)
Index1 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTGATGGAT(2)
Index1 adapter R	5’-TCCATCAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(3)
Index2	CTAAGTG(4)
Index2 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTCTAAGTGT(5)
Index2 adapter R	5’-CACTTAGAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(6)
Index3	TAGGAACA(7)
Index3 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGAACAT(8)
Index3 adapter R	5’-TGTTCCTAAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC (9)
Index4	AGGCTG(10)
Index4 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGCTGT(11)
Index4 adapter R	5’-CAGCCTAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(12)

Index5	AGAGATA(13)
Index5 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTAGAGATAT(14)
Index5 adapter R	5'-TATCTCTAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(15)
Index6	GCTCTCGA(16)
Index6 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTCTCGAT(17)
Index6 adapter R	5'-TCGAGAGCAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(18)

利用上述根据本发明实施例的寡核苷酸（也可以成为 DNA 标签接头），能够有效地将 DNA 标签引入到样品的 DNA 或其等同物中，由此能够构建具有本发明的 DNA 标签的 DNA 标签文库，进而能够准确有效地进行多重核酸测序。另外，发明人惊奇地发现，当针对相同的样品，采用具有不同标签的寡核苷酸构建含有各种 DNA 标签的 DNA 标签文库时，所得到的测序数据结果的稳定性和可重复性非常好。

根据本发明的另一方面，本发明还提供了用于将上述 DNA 标签引入样品 DNA 或其等同物中的一组分离的 PCR 标签引物。根据本发明的实施例的一组分离的 PCR 标签引物，具有第一链和第二链，其中第一链分别由 SEQ ID NO: 39 所示的核苷酸构成，第二链的核苷酸序列如下所示：CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCT，其中， XXXXXX 为根据本发明实施例的一组分离的 DNA 标签的一种。该组 PCR 标签引物其分别具有如前所述的根据本发明实施例的 DNA 标签，通过采用 PCR 标签引物的 PCR 反应，可以将 PCR 标签引物引入样品的 DNA 或其等同物中，从而就将相应的 DNA 标签引入到 DNA 或其等同物中，进而能够获得含有相应的 DNA 标签的 DNA 标签文库，由此，能够将多种样品的 DNA 标签文库进行混合测序，即本发明的 PCR 标签引物能够有效地应用于多重核酸测序技术。另外，发明人惊奇地发现，当针对相同的样品，采用具有不同标签的 PCR 标签引物分别构建含有各种 DNA 标签的 DNA 标签文库时，所得到的测序数据结果的稳定性和可重复性非常好。

根据本发明的又一方面，本发明提供了一种构建 DNA 标签文库的方法。根据本发明的实施例，该方法包括以下步骤：将 DNA 样品片段化，以便获得 DNA 片段；将所得的 DNA 片段进行末端修复，以便获得经过末端修复的 DNA 片段；将经过末端修复的 DNA 片段进行 3'末端添加碱基 A，以便获得 3'末端添加碱基 A 的 DNA 片段；将 3'末端添加碱基 A 的 DNA 片段与 DNA 标签接头相连，以便获得连接产物，其中该 DNA 标签接头包含前面所述根据本发明的实施例的一组分离的 DNA 标签的一种；将所得的连接产物进行扩增，以便获得扩增产物；以及分离回收扩增产物，该扩增产物构成 DNA 标签文库。

根据本发明实施例的构建DNA标签文库的方法，能够有效地应用于多重核酸测序技术，具体地，利用根据本发明实施例的构建 DNA 标签文库的方法，能够有效地将根据本发明实施例的 DNA 标签引入到针对样品 DNA 所构建的 DNA 标签文库中，进而能够通过对 DNA 标签文库进行测序，获得样品 DNA 的序列信息以及 DNA 标签组合的序列信息，进而能够对样品 DNA 的来源进行区分。另外，发明人惊奇地发现，基于上述方法，采用不同的标签

构建的含有根据本发明实施例的 DNA 标签的 DNA 标签文库应用于多重核酸测序技术时，相对于常规的多重核酸测序技术，能够有效地减少数据产出偏向性问题，所得到的测序数据结果可靠、稳定，可重复性非常好。

进一步，本发明提供了一种 DNA 标签文库。根据本发明的实施例，该 DNA 标签文库是通过根据本发明实施例的构建 DNA 标签文库的方法获得的。发明人发现，本发明的 DNA 标签文库能够有效地应用于高通量测序平台，从而能够有效地获得准确、可靠的测序数据，并且所得的测序数据稳定性和可重复性好。

根据本发明的另一方面，本发明提供了一种确定 DNA 样品序列信息的方法。根据本发明的实施例，其包括下列步骤：根据本发明实施例的制备 DNA 标签文库的方法构建所述 DNA 样品的 DNA 标签文库；以及对所述 DNA 标签文库进行测序，以便确定所述 DNA 样品的序列信息。基于该方法，能够有效地获得 DNA 标签文库中 DNA 样品的序列信息以及 DNA 标签的序列信息，从而能够对 DNA 样品的来源进行区分。另外，发明人惊奇地发现，利用根据本发明实施例的方法确定 DNA 样品序列信息，能够有效地减少数据产出偏向性的问题，并且能够精确地对多种 DNA 标签文库进行区分。

根据本发明的再一方面，本发明还提供了一种确定多种 DNA 样品序列信息的方法。根据本发明的实施例，其包括以下步骤：针对所述多种样品的每一种，分别独立地根据本发明实施例的构建 DNA 标签文库的方法，构建所述 DNA 样品的 DNA 标签文库，其中，不同的 DNA 样品采用相互不同并且已知序列的 DNA 标签；将所述多种样品的 DNA 标签文库进行组合，以便获得 DNA 标签文库混合物；对所述 DNA 标签文库混合物进行测序，以便获得所述 DNA 样品的序列信息以及所述标签的序列信息；以及基于所述标签的序列信息对所述 DNA 样品的序列信息进行分类，以便确定所述多种样品的 DNA 序列信息。由此，根据本发明实施例的该方法，可以充分利用高通量测序技术，例如利用 Solexa 测序技术，同时对多种样品的 DNA 标签文库进行测序，从而能够提高 DNA 标签文库测序的效率和通量，同时可以提高确定多种 DNA 样品序列信息的效率，并且相对于常规的多重核酸测序技术，能够有效减少数据产出偏向性的问题，测序数据更加准确、可靠。

根据本发明的又一方面，本发明提供了一种用于构建 DNA 标签文库的试剂盒。根据本发明的实施例，该试剂盒包括：6 种分离的寡核苷酸，所述分离的寡核苷酸具有第一链和第二链，所述第一链分别由 SEQ ID NO:（3N-1）所示的核苷酸构成，所述第二链分别由 SEQ ID NO:（3N）所示的核苷酸构成，其中，对于相同的寡核苷酸，其第一链和第二链的 N 取值相同，并且 N=1-6 的任意整数，其中，所述 6 种分离的寡核苷酸分别设置在不同的容器中。由此，利用该试剂盒，能够方便地将根据本发明实施例的 DNA 标签引入到构建的 DNA 标签文库中。从而，根据本发明实施例的试剂盒能够有效地应用于多重核酸测序技术。

根据本发明的又一方面，本发明提出了一组不同长度，优选梯度长度的标签用于测序标签文库的构建和/或测序的用途，其中所述标签是一段寡聚核苷酸序列，优选是 2－100bp 的核苷酸序列。

根据本发明的一个实施例，其中将所述标签包含在用于扩增目的序列的 PCR 引物中，

从而构成各自相对应的标签 PCR 引物，通过 PCR 方法引入待测序序列中，所述 PCR 标签引物用作 PCR 的 5'引物，或 3'引物，或者同时用作 PCR 的 5'引物和 3'引物。

根据本发明的一个实施例，其中在所述标签 PCR 引物中，所述标签嵌入用于扩增目的序列的 PCR 引物中，或者通过或不通过连接子与用于扩增目的序列的 PCR 引物的 5'端或 3'端相连，从而构成各自相对应的标签 PCR 引物。

根据本发明的一个实施例，其中将所述标签包含在标签文库的接头中，从而构成各自相对应的标签接头，所述标签接头用作标签文库的 5'接头，3'接头，或者同时用作标签文库的 5'接头和 3'接头。

根据本发明的一个实施例，其中所述标签插入接头中，或通过或不通过连接子连接在接头的末端，优选地不通过连接子连接在接头的末端，从而构成自相对应的标签接头。

根据本发明的一个实施例，其中所述标签构成标签 PCR 引物和标签接头，同时用于测序标签文库的构建和/或测序。

根据本发明的又一方面，本发明提出了一组不同长度，优选梯度长度的标签，其用于测序标签文库的构建和/或测序。

根据本发明的又一方面，本发明提出了一组梯度标签，所述一组梯度标签包括如下或由如下组成：表 1 所示的 6 个梯度标签或者与其相差 1 个碱基的梯度标签中的至少 2 个，或至少 3 个，或至少 4 个，或至少 5 个，或全部 6 个，

所述一组梯度标签优选地至少包括表 1 所示的 6 个梯度标签中的 Index1 和 Index2，或 Index3 和 Index4，或 Index5 和 Index6，或者他们任何两个或多个的组合。

根据本发明的实施例，其中所述相差 1 个碱基包括对表 1 所示的 6 个梯度标签的序列中 1 个碱基的取代、添加或缺失。

根据本发明的又一方面，本发明提出了前面所述的梯度标签用于测序标签文库，特别是 Illumina/Solexa 测序标签文库的构建和/或测序的用途，其中所述梯度标签包含在用于测序标签文库，特别是 Illumina/Solexa 测序标签文库的接头 2 的 5'末端中，从而构成各自相对应的梯度标签接头 2，其用作测序标签文库，特别是 Illumina/Solexa 测序标签文库的 3'接头。

根据本发明的实施例，所述梯度标签包含在接头 2 的 5'末端中，包括所述梯度标签通过或不通过连接子与接头 2 的 5'末端相连，或者插入接头 2 的 5'末端中，优选的是不通过连接子与接头 2 的 5'末端相连。

根据本发明的又一方面，本发明提出了使用前述方法构建的测序标签文库，特别是 Illumina/Solexa 测序标签文库。

根据本发明的又一方面，本发明提出了包含前面所述的梯度标签的一组梯度标签接头 2，其在 5'末端包含权利要求 1 所述的梯度标签，并且优选地用作测序标签文库，特别是 Illumina/Solexa 测序标签文库的 3'接头，所述一组梯度标签接头 2 包括如下或由如下组成：表 1 所示的 6 个梯度标签接头 2 或者与其中包含的梯度标签序列相差 1 个碱基的接头中的至少 2 个，或至少 3 个，或至少 4 个，或至少 5 个，或全部 6 个，

所述一组梯度标签接头 2 优选地至少包括表 1 所示的 6 个梯度标签接头 2 中的 Index1 adapter2 F/R 和 Index2 adapter2 F/R，或 Index3 adapter2 F/R 和 Index4 adapter2 F/R，或 Index5 adapter2 F/R 和 Index6 adapter2 F/R，或者他们任何两个或多个的组合。

根据本发明的实施例，本发明提出了梯度标签接头 2，其中所述相差 1 个碱基包括对梯度标签序列中 1 个碱基的取代、添加或缺失。

根据本发明的实施例，本发明提出了梯度标签接头 2 用于测序标签文库，特别是 Illumina/Solexa 测序标签文库的构建和/或测序的用途，所述梯度标签接头 2 用作测序标签文库，特别是 Illumina/Solexa 测序标签文库的 3'接头。

根据本发明的实施例，本发明提出了使用所述的梯度标签接头 2 构建的测序标签文库，特别是 Illumina/Solexa 测序标签文库，其中所述梯度标签接头 2 用作测序标签文库，特别是 Illumina/Solexa 测序标签文库的 3'接头。

根据本发明的又一方面，本发明提出了一种构建测序标签文库，特别是 Illumina/Solexa 测序标签文库的方法，所述方法的特征在于使用一组具有不同长度，优选梯度长度的标签的接头用作测序标签文库，特别是 Illumina/Solexa 测序标签文库的 3'接头。

根据本发明的实施例，构建测序标签文库的方法包括:

1）提供 n 个总基因组 DNA 样品，所述基因组 DNA 样品来自任何真核生物样品，包括但不限于人的基因组 DNA 样品;

2）打断 DNA: 通过机械法打断 DNA，产生带有粘性末端的 DNA 片段，所述机械法包括但不限于使用 Bioruptor、Hydroshear 和 Covaris;

3）末端修复: 通过连接反应将 DNA 片段的粘性末端补平;

4）末端加 A: 通过连接反应在 DNA 片段的平末端加上一个腺嘌呤碱基 A;

5）添加 5'接头和 3'接头;

6）通过 PCR 对目的片段进行扩增，最后通过回收目的片段文库;

7）混合: 当 n>1 时，将各样品的 PCR 扩增产物混合在一起。

根据本发明的实施例，其中所述方法的特征在于使用不同的选自表 1 的梯度标签接头 2 或者与其中包含的梯度标签序列相差 1 个碱基的接头用作测序标签文库，特别是 Illumina/Solexa 测序标签文库的 3'接头。

根据本发明的实施例，构建测序标签文库的方法包括:

1）提供 n 个总基因组 DNA 样品，n 为整数且 $1 \leq n \leq 6$，优选地 $2 \leq n \leq 6$，所述基因组 DNA 样品来自任何真核生物样品，包括但不限于人的基因组 DNA 样品;

2）打断 DNA: 通过机械法打断 DNA，产生带有粘性末端的 DNA 片段，所述机械法包括但不限于使用 Bioruptor、Hydroshear 和 Covaris;

3）末端修复: 通过连接反应将 DNA 片段的粘性末端补平;

4）末端加 A: 通过连接反应在 DNA 片段的平末端加上一个腺嘌呤碱基 A;

5）添加接头 1 和梯度标签接头 2: 通过连接反应将接头 1 和梯度标签接头 2 与带有 A-末端的 DNA 片段进行连接;

6）通过 PCR 对目的片段进行扩增，最后通过回收目的片段文库;

7）混合：当 n>1 时，将各样品的 PCR 扩增产物混合在一起。

根据本发明的实施例，其中所述接头 1 包括如下接头：5'-TACACTCTTTCCCTACACG ACGCTCTTCCGATCTATCACT(SEQ ID NO:37)和 5'- /GTGATAGATCGGAAGAGCACAC GTCTGAACTCCAGTCAC(SEQ ID NO:38)。

根据本发明的实施例，其中所述梯度标签接头 2 包括表 1 所示的 6 个梯度标签接头 2 或与其中包含的梯度标签序列相差 1 个碱基的接头中的至少 2 个，或至少 3 个，或至少 4 个，或至少 5 个，或全部 6 个，

所述一组梯度标签接头 2 优选地至少包括表 1 所示的 6 个梯度标签接头 2 中的 Index1 adapter2 F/R 和 Index2 adapter2 F/R，或 Index3 adapter2 F/R 和 Index4 adapter2 F/R，或 Index5 adapter2 F/R 和 Index6 adapter2 F/R，或者他们任何两个或多个的组合。

根据本发明的实施例，其中所述相差 1 个碱基包括梯度标签序列中 1 个碱基的取代、添加或缺失。

根据本发明的实施例，其中步骤 6）中的 PCR 使用如下 PCR 引物:

PCR Primer 1:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCT TCCGATCT(SEQ ID NO:39)，和

PCR Primer 2:

5'- CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT(SEQ ID NO:40)。

根据本发明的实施例，其中步骤 6）中回收目的片段文库是通过琼脂糖凝胶电泳以及切胶回收进行。

根据本发明的实施例，其进一步包括如下步骤;

8）测序：将各样品的 PCR 扩增产物利用测序技术，特别是 Illumina/Solexa 测序技术进行测序。

根据本发明的实施例，其中利用测序技术进行测序中使用的测序引物包括当构建 DNA Pair-end 文库时，使用测序引物为 Sequencing Primer1：5' – ACACTCTTTCCCTACACGACGCTCTTC CGATCT。

根据本发明的实施例，本发明还提出了根据前述方法构建的测序标签文库，特别是 Illumina/Solexa 测序标签文库。

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。

附图说明

本发明的上述和/或附加的方面和优点从结合下面附图对实施例的描述中将变得明显和容易理解，其中：

图 1：显示了根据本发明实施例的构建 DNA 标签文库的方法的流程示意图;

图 2: 显示了根据本发明实施例的利用 Illumina/Solexa 测序平台对 DNA 标签文库进行测序时，各测序反应的序列读取顺序及所读出序列的组成的示意图;

图 3: 显示了根据本发明实施例的本发明的 DNA 标签文库与对照例的 DNA 标签文库的前 10 个 Illumina/Solexa 测序循环的质量值的比较结果，其中，

图 3A: 显示了实施例 1 的本发明的 DNA 标签文库的前 10 个 Illumina/Solexa 测序循环的质量值，

图 3B: 显示了对照例的 DNA 标签文库的前 10 个 Illumina/Solexa 测序循环的质量值;

图 4: 显示了根据本发明实施例的本发明的 DNA 标签文库与对照例的 DNA 标签文库的各 Illumina/Solexa 测序循环的光强的比较结果，其中，

图 4A: 显示了实施例 1 的本发明的 DNA 标签文库的各 Illumina/Solexa 测序循环的光强信号平均值，

图 4B: 显示了对照例的 DNA 标签文库的各 Illumina/Solexa 测序循环的光强信号平均值;

图 5: 显示了根据本发明实施例的本发明的 DNA 标签文库与对照例的 DNA 标签文库的 Illumina/Solexa 测序循环的碱基分布的比较结果，其中，

图 5A: 显示了实施例 1 的本发明的 DNA 标签文库的 Illumina/Solexa 测序循环的碱基百分比分布图，

图 5B: 显示了对照例的 DNA 标签文库的 Illumina/Solexa 测序循环的碱基百分比分布图; 以及

图 6: 显示了根据本发明实施例的本发明的 DNA 标签文库与对照例的 DNA 标签文库的 Illumina/Solexa 测序循环的错误率的比较结果，其中，

图 6A: 显示了实施例 1 的本发明的 DNA 标签文库的 Illumina/Solexa 测序循环的错误率图，

图 6B: 显示了对照例的 DNA 标签文库的 Illumina/Solexa 测序循环的错误率图。

发明详细描述

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，仅用于解释本发明，而不能理解为对本发明的限制。

需要说明的是，在本发明的描述中，除非另有说明，“多个”的含义是两个或两个以上。

DNA 标签

根据本发明的一个方面，本发明提供了一组分离的 DNA 标签。根据本发明的实施例，这些分离的 DNA 标签分别由 SEQ ID NO:（3N-2）所示的核苷酸序列构成，其中 N=1-6 的任意整数。在本说明书中，这些 DNA 标签分别被命名为 IndexN，其中 N=1-6 的任意整数，

其序列如前面表 1 所示，在此不再赘述。

在本发明中所使用的术语“DNA”可以是任何包含脱氧核糖核苷酸的聚合物，包括但不限于经过修饰的或者未经修饰的 DNA。利用根据本发明实施例的 DNA 标签，通过将 DNA 标签与样品的 DNA 或其等同物相连，得到具有标签的 DNA 标签文库，通过对 DNA 标签文库进行测序，可以获得样品 DNA 的序列以及标签的序列，进而基于标签的序列可以精确地表征 DNA 的样品来源。由此，利用上述 DNA 标签，可以同时构建多种样品的 DNA 标签文库，从而可以通过将来源于不同样品的 DNA 标签文库进行混合，同时进行测序，基于 DNA 标签对样品的 DNA 序列进行分类，获得多种样品的 DNA 的序列信息。从而可以充分利用高通量的测序技术，例如利用 Solexa 测序技术，同时对多种样品的 DNA 进行测序，从而提高了通过高通量测序技术的效率和通量，降低了确定 DNA 样品序列信息的成本。这里所使用的表述方式“DNA 标签与样品的 DNA 或其等同物相连”应做广义理解，其包括 DNA 标签可以与样品的 DNA 直接相连，以构建 DNA 标签文库，也可以与和样品的 DNA 具有相同序列的核酸（例如可以是相应的 RNA 序列或 cDNA 序列，其与 DNA 具有相同的序列）相连。

本申请的发明人发现：在本发明中，为了设计有效的 DNA 标签，首先需要考虑标签序列之间的序列差异程度和碱基识别率。其次，在标签混合量少于 6 个样品的情况下，必须考虑到混合后的标签上的每个碱基位点的 GT 含量。因为 Solexa 测序过程中，碱基 G 和 T 的激发荧光一样，碱基 A 和 C 的激发光是一样的，因此必须考虑碱基“GT”含量与碱基“AC”含量的“平衡”，最适碱基“GT”含量为 50%，能保证标签识别率最高和错误率最低。最后，还要考虑数据产出的可重复性和准确性，即为了实现能够有效构建 DNA 标签文库并进行测序，所构建的一组 DNA 标签需要能够保证结果可靠，可重复性高，也就是针对同样的 DNA 样品，可以保证利用该组 DNA 标签中的不同标签构建的 DNA 标签文库，能够获得一致的测序结果，因而可以确保实验结果可靠且重复性高。另外，还需要同时避免标签序列出现 3 或 3 个以上连续的碱基的出现，因为 3 个或 3 个以上连续的碱基会增加序列在合成过程中或测序过程中的错误率，同时也要尽量避免 DNA 标签接头和 PCR 标签引物自身形成发夹结构。

为此，本申请的发明人进行了大量的筛选工作，并且选定了根据本发明实施例的一组分离的 DNA 标签，其分别由 SEQ ID NO:（3N-2）所示的核苷酸序列构成，其中 N=1-6 的任意整数。其序列如前面表 1 所示，不再赘述。根据本发明的实施例，该组分离的 DNA 标签为梯度标签，具体地，Index1 为 6bp，Index2 为 7bp，Index1 为 8bp，则 Index1-3 具有 1bp 递增的梯度，而同样，Index4 为 6bp，Index5 为 7bp，Index6 为 8bp，则 Index4-6 具有 1bp 递增的梯度，由此，能够有效减少标签中碱基的偏向性，进而将这些 DNA 标签应用到多重核酸测序技术中时，能够有效地减少测序过程中的光强参数的波动，从而能够显著提高输出数据的质量，进一步，能够显著提高测序数据的可靠性、稳定性和可重复性。此外，各 DNA 标签之间的差异在 5 个碱基以上，当 DNA 标签的碱基中的任意 1 个碱基出现测序错误或合成错误，都不影响到标签的最终识别。这些标签可以应用于任何 DNA 标签文库的构

建。目前尚未有关于这些标签应用于 DNA 样品测序的文库构建并通过 Solexa 技术测序的报道。

根据本发明的一些实施例，所采用的一组 DNA 标签由如下组成：表 1 所示 DNA 标签或与之相差 1 个碱基的 DNA 标签中的至少 2 个，或至少 3 个，或至少 4 个，或至少 5 个，或全部 6 个。具体地，根据本发明的实施例，所述一组 DNA 标签优选地至少包括表 1 所示的 6 个 DNA 标签的 Index1 和 Index2，或 Index3 和 Index4，或 Index5 和 Index6，或者它们任何两个或多个的组合。在本发明的一些具体示例中，其中所述相差 1 个碱基包括对表 1 所示的 6 个 DNA 标签的序列中 1 个碱基的取代、添加或缺失。

根据本发明的实施例，本发明还提供了将根据本发明实施例的 DNA 标签在多重核酸测序技术中的用途，具体地，即 DNA 标签在 DNA 标签文库构建及测序的用途。其中，可以将本发明的 DNA 标签包含在 DNA 标签文库的接头中，从而构成各自相对应的 DNA 标签接头，该 DNA 标签接头用作 DNA 标签文库的 5'接头，3'接头，或者同时用作 DNA 标签文库的 5'接头和 3'接头，由此，能够通过连接反应将 DNA 标签引入到 DNA 标签文库中。根据该用途的实施例，DNA 标签插入 DNA 标签接头的末端中，或通过或不通过连接子连接在 DNA 接头的末端，优选地不通过连接子连接在 DNA 标签接头的末端中，从而构成各自相对应的 DNA 标签接头。此外，可以将本发明的 DNA 标签包含在用于扩增目的序列的 PCR 引物中，从而构成各自相对应的 PCR 标签引物，该 PCR 标签引物用作 PCR 反应的 5'引物，或 3'引物，或者同时用作 PCR 反应的 5'引物和 3'引物，由此，能够通过 PCR 方法将 DNA 标签引入到 DNA 标签文库中。根据该用途的实施例，DNA 标签嵌入 PCR 标签引物中，或者通过或不通过连接子与用于扩增目的序列的 PCR 引物的 5'末端或 3'末端相连，从而构成各自相对应的 PCR 标签引物。

根据本发明的一个具体示例，本发明提供了前面所述的 DNA 标签用于 DNA 标签文库，特别是 Illumina/Solexa 测序 DNA 标签文库的构建和/或测序的用途。在本发明提供的所述用途中，DNA 标签包含在 DNA 标签文库，特别是 Illumina/Solexa 测序 DNA 标签文库的 3'接头的 5'末端中，从而构成各自相对应的 DNA 标签接头，其用作 DNA 标签文库，特别是 Illumina/Solexa 测序 DNA 标签文库的 3'接头。

根据本发明的一个具体示例，本发明提供了前面所述的 DNA 标签用于 DNA 标签文库，特别是 Illumina/Solexa 测序 DNA 标签文库的构建和/或测序的用途。其中 DNA 标签包含在 DNA 标签文库，特别是 Illumina/Solexa 测序 DNA 标签文库的 3'接头的 5'末端中，其中 DNA 标签通过或不通过连接子与该 3'接头的 5'末端相连，或者插入该 3'接头的 5'末端中。优选的是不通过连接子与 3'接头的 5'末端相连。其中所述连接子是 1－10 个碱基的序列，优选地 1－5 个碱基的序列，更优选 1－3 个碱基的序列。

寡核苷酸、PCR 标签引物以及构建 DNA 标签文库的方法

根据本发明的又一方面，本发明提供了一组分离的寡核苷酸，其可以用于将前面所述的 DNA 标签引入到样品的 DNA 中，进而构建 DNA 标签文库。根据本发明的实施例，该分离的寡核苷酸具有第一链和第二链，其中第一链分别由 SEQ ID NO:（3N-1）所示的核苷

酸构成，第二链分别由 SEQ ID NO:（3N）所示的核苷酸构成，其中，对于相同的寡核苷酸，其第一链和第二链的 N 取值相同，并且 N=1-6 的任意整数。在本文中所使用的表达方式“对于相同的寡核苷酸，其第一链和第二链的 N 取值相同”是指，采用序列表中的相应核苷酸分别作为第一链和第二链时，构成第一链的核苷酸与构成第二链的核苷酸能够形成稳定的二聚体，具体地，例如当 N=3 时，采用 SEQ ID NO: 8 作为第一链，SEQ ID NO: 9 作为第二链。本领域技术人员能够理解，可以通过分别将构成相应寡核苷酸的第一链与第二链进行退火处理，而形成相应的寡核苷酸。在本说明书中，与 DNA 标签相对应的寡核苷酸（DNA 标签接头）分别被命名为 IndexN adapter，其中 N=1-6 的任意整数，其第一链和第二链分别被命名为 IndexN adapter F 和 IndexN adapter R，其中 N=1-6 的任意整数，其序列如前面表 1 所示，在此不再赘述。根据本发明的实施例，上述寡核苷酸分别具有如前所述的根据本发明实施例的 DNA 标签，因而可以通过连接反应，将相应的 DNA 标签引入到样品的 DNA 或其等同物中。具体地，这些寡核苷酸的序列如前面表 1 所示，在此不再赘述。

具体地，根据本发明的实施例，该组 DNA 标签接头，其在 5’末端含有本发明的 DNA 标签，并且优选地用作 DNA 标签文库，特别是 Illumina/Solexa 测序 DNA 标签文库的 3’接头。根据本发明的一些实施例，该组 DNA 标签接头包括或由如下组成：表 1 所示的 6 个 DNA 标签接头或者与其包含的 DNA 标签序列相差 1 个碱基的 DNA 标签接头中的至少 2 个，或至少 3 个，或至少 4 个，或至少 5 个，或全部 6 个。根据本发明的具体示例，这些 DNA 标签接头优选地至少包括表 1 所示的 6 个 DNA 标签接头 2 中的 Index1 adapter F/R 和 Index2 adapter F/R，或 Index3 adapter F/R 和 Index4 adapter F/R，或 Index5 adapter F/R 和 Index6 adapter F/R，或者他们任何两个或多个的组合。在本发明的一些具体示例中，前面所述的 DNA 标签接头中所述相差 1 个碱基包括对表 1 所示的 6 个 DNA 标签的序列中 1 个碱基的取代、添加或缺失。

根据本发明的一个具体示例，本发明提供了 DNA 标签接头用于 DNA 标签文库，特别是 Illumina/Solexa 测序 DNA 标签文库构建和/或测序的用途，所述 DNA 标签接头用作 DNA 标签文库，特别是 Illumina/Solexa 测序 DNA 标签文库的 3’接头。

根据本发明的另一方面，本发明提供了一组分离的 PCR 标签引物，其可以用于将前面所述的 DNA 标签引入到样品的 DNA 中，进而构建 DNA 标签文库。根据本发明的实施例，该组分离的 PCR 标签引物具有第一链和第二链，其中第一链分别由 SEQ ID NO: 39 所示的核苷酸构成，第二链的核苷酸序列如下所示：CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT，其中，XXXXXX 为根据本发明实施例的一组分离的 DNA 标签的一种。根据本发明的实施例，本发明的 PCR 标签引物分别具有如前所述的根据本发明实施例的 DNA 标签，通过采用 PCR 标签引物的 PCR 反应，能够将 PCR 标签引物引入样品的 DNA 或其等同物中，进而将相应的 DNA 标签引入到 DNA 或其等同物中，进而能够将 DNA 标签引入到 DNA 标签文库中，从而通过各样品的 DNA 标签文库进行混合测序能够实现对多种样品的测序，并能够有效地减少数据产出偏向性的问题，获得的测序数据可靠、稳定，可重复性好。根据本发明的一些实施例，本发明的 PC

R 标签引物，可以用作 PCR 反应的 5’引物，或 3’引物，或者同时用作 PCR 反应的 5’引物和 3’引物，由此，能够通过 PCR 方法将 DNA 标签引入到 DNA 标签文库中。

进一步，根据本发明的实施例，还提供了使用上述 DNA 标签接头和 PCR 标签引物构建的 DNA 标签文库。

根据本发明的另一方面，本发明还提供了一种利用上述 DNA 标签接头构建 DNA 标签文库的方法。具体地，根据本发明的实施例，参考图 1，该方法可以包括:

首先，将 DNA 样品片段化，以便获得 DNA 片段。根据本发明的实施例，将 DNA 样品片段化的方法不受特别限制。根据本发明的一些实施例，片段化是通过选自雾化法、酶切法和超声打断法的至少一种进行的。根据本发明的一些具体示例，利用超声打断法将 DNA 样品片段化。根据本发明的一个具体示例，利用 Covaris 打断仪将 DNA 样品片段化。根据本发明的实施例，获得的 DNA 片段的长度可以为大约 500bp，由此，能够进一步提高构建 DNA 标签文库以及后续测序的效率。根据本发明的实施例，DNA 样品的来源并不受特别限制。根据本发明的一些实施例，DNA 样品可以来源于任何常见的生物样品，例如植物、动物或微生物。根据本发明的另一些实施例，DNA 样品可以来源于选自拟南芥、水稻、人、小鼠和大肠杆菌的至少一种。从而，利用根据本发明实施例的方法，能够有效地构建多种常见模式生物的 DNA 标签文库。根据本发明的具体示例，DNA 样品可以为人 DNA 样品，更具体地，可以为人基因组 DNA 样品。

其次，将 DNA 片段进行末端修复，以便获得经过末端修复的 DNA 片段。根据本发明的实施例，将 DNA 片段进行末端修复的方法不受特别限制。根据本发明的一个具体示例，末端修复是利用 Klenow 片段、T4 DNA 聚合酶和 T4 多核苷酸激酶进行的，其中该 Klenow 片段具有 5’→3’聚合酶活性和 3’→5’外切酶活性，但缺少 5’→3’外切酶活性。

接着，将经过末端修复的 DNA 片段进行 3’末端添加碱基 A，以便获得 3’末端添加碱基 A 的 DNA 片段。根据本发明的实施例，将经过末端修复的 DNA 片段进行 3’末端添加碱基 A 的方法不受特别限制。根据本发明的具体示例，可以利用 Klenow (3’-5’ exo-)将经过末端修复的 DNA 片段进行 3’端添加碱基 A。根据本发明的实施例，将经过末端修复的 DNA 片段的两条寡核苷酸链的 3’末端均添加碱基 A。

接下来，将 3’末端添加碱基 A 的 DNA 片段与 DNA 标签接头相连，以便获得连接产物，其中 DNA 标签接头包含根据本发明实施例的一组分离的 DNA 标签的一种。根据本发明的实施例，将 3’末端添加碱基 A 的 DNA 片段与 DNA 标签接头相连的方法不受特别限制。根据本发明的一些具体示例，将 3’末端添加碱基 A 的 DNA 片段与 DNA 标签接头相连是利用 T4 DNA 连接酶进行的。根据本发明的实施例，在 3’末端添加碱基 A 的 DNA 片段的 3’末端和 5’末端的至少一处连接 DNA 标签接头。根据本发明一些具体示例，在 3’末端添加碱基 A 的 DNA 片段的 3’末端连接 DNA 标签接头。根据本发明的实施例，DNA 标签接头可以为选自前面所述的根据本发明实施例的分离的寡核苷酸的一种。

然后，将得到的连接产物进行扩增，以便获得扩增产物。根据本发明的实施例，将连接产物进行扩增之前，进一步包括对连接产物进行片段选择的步骤。根据本发明的一些实施例，可以利用2%琼脂糖凝胶电泳进行片段选择。根据本发明的一个具体示例，经过片段选择的连接产物的长度可以为约620bp。根据本发明的实施例，将连接产物进行扩增，是通过PCR反应进行的，该PCR反应采用分别具有如SEQ ID NO: 39和SEQ ID NO: 40所示核苷酸序列的引物。

最后，分离回收获得的扩增产物，该扩增产物构成DNA标签文库。根据本发明的实施例，分离回收扩增产物的方法也不受特别限制，本领域技术人员可以根据扩增产物的特点选择适当的方法和设备进行分离，例如可以通过电泳并且回收特定长度的PCR扩增产物的方法进行回收。根据本发明的实施例，可以通过选自琼脂糖凝胶电泳、磁珠纯化和纯化柱纯化的至少一种分离扩增产物。根据本发明的具体示例，分离回收的扩增产物的长度可以为大约620bp。

利用根据本发明实施例的构建DNA标签文库的方法，能够有效地将根据本发明实施例的DNA标签引入到针对DNA样品所构建的DNA标签文库中。从而可以通过对DNA标签文库进行测序，获得DNA样品的序列信息以及DNA标签的序列信息，从而能够对DNA样品的来源进行区分。本发明提供的构建DNA标签文库的方法能够充分的利用高通量测序平台，满足高通量测序的需求，节省测序资源，从而降低测序成本，并且相对于常规的多重核酸测序技术，能够有效地减少数据产出偏向性的问题，获得的测序数据可靠、稳定，可重复性好。

具体地，根据本发明的实施例，本发明的构建DNA标签文库的方法可以包括:

1）提供n个基因组DNA样品，其中基因组DNA样品可以来自任何真核生物样品，包括但不限于人的基因组DNA样品;

2）打断DNA，可以通过机械法打断DNA，产生带有粘性末端的DNA片段，所述机械法包括但不限于使用Bioruptor、Hydroshear和Covaris;

3）末端修复，通过连接反应将DNA片段的粘性末端补平，以便获得经过末端修复的DNA片段;

4）末端加A，通过连接反应在经过末端修复的DNA片段的平末端加上一个腺嘌呤碱基A，以便获得3’末端添加碱基A的DNA片段;

5）添加5’接头和3’接头，将接头1和DNA标签接头分别作为DNA标签文库的5’接头和3’接头，并将3’末端添加碱基A的DNA片段与接头1和DNA标签接头相连，以便获得连接产物，其中DNA标签接头包含根据本发明实施例的一组分离的DNA标签的一种;

6)通过PCR对连接产物进行扩增，并通过琼脂糖凝胶电泳以及切胶分离回收扩增产物，该扩增产物构成DNA标签文库;

7）混合：当n>1时，将各样品的DNA标签文库混合在一起，以便获得DNA标签文库混合物。

进一步，根据本发明的一个具体示例，本发明的构建DNA标签文库的方法还可以包括:

1)提供 n 个基因组 DNA 样品，n 为整数且 1≤n≤6，优选地 2≤n≤6，其中基因组 DNA 样品可以来自任何真核生物样品，包括但不限于人的基因组 DNA 样品；

2）打断 DNA，可以通过机械法打断 DNA，产生带有粘性末端的 DNA 片段，所述机械法包括但不限于使用 Bioruptor、Hydroshear 和 Covaris；

3）末端修复，通过连接反应将 DNA 片段的粘性末端补平，以便获得经过末端修复的 DNA 片段；

4）末端加 A，通过连接反应在经过末端修复的 DNA 片段的平末端加上一个腺嘌呤碱基 A，以便获得 3’末端添加碱基 A 的 DNA 片段；

5）连接 5’接头和 DNA 标签接头：将 DNA 标签接头作为 DNA 标签文库的 3’接头，通过连接反应将 5’接头和 DNA 标签接头与 3’末端添加碱基 A 的 DNA 片段进行连接，以便获得连接产物，其中 DNA 标签接头包含根据本发明实施例的一组分离的 DNA 标签的一种；

6)通过 PCR 对连接产物进行扩增，并通过琼脂糖凝胶电泳以及切胶分离回收扩增产物，该扩增产物构成 DNA 标签文库；

7）混合：当 n>1 时，将各样品的 DNA 标签文库混合在一起，以便获得 DNA 标签文库混合物。

根据本发明的一个实施例，上述本发明的构建 DNA 标签文库的方法中使用的 Illumina/Solexa 测序文库的 5’接头为接头 1，其序列如下所示：

5’-TACACTCTTTCCCTACACGACGCTCTTCCGATCTATCACT(SEQ ID NO:37)；

5’-GTGATAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(SEQ ID NO:38)。

根据本发明的一个实施例，上述本发明的构建 DNA 标签文库的方法中使用的 DNA 标签接头包括或由如下组成：表 1 所示的 6 个 DNA 标签接头或者与其包含的 DNA 标签序列相差 1 个碱基的 DNA 标签接头中的至少 2 个，或至少 3 个，或至少 4 个，或至少 5 个，或全部 6 个。根据本发明的具体示例，这些 DNA 标签接头优选地至少包括表 1 所示的 6 个 DNA 标签接头 2 中的 Index1 adapter F/R 和 Index2 adapter F/R，或 Index3 adapter F/R 和 Index4 adapter F/R，或 Index5 adapter F/R 和 Index6 adapter F/R，或者他们任何两个或多个的组合。在本发明的一些具体示例中，前面所述的 DNA 标签接头中所述相差 1 个碱基包括对表 1 所示的 6 个 DNA 标签的序列中 1 个碱基的取代、添加或缺失。

根据本发明的一个实施例，上述本发明的构建 DNA 标签文库的方法中的 PCR 反应采用如下 PCR 引物：

PCR Primer 1:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT(SEQ ID NO:39)，和

PCR Primer 2:

CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT(SEQ ID NO:40)。

根据本发明的一个实施例，本发明的构建 DNA 标签文库的方法可以进一步包括：

8）测序：利用测序技术，特别是 Illumina/Solexa 测序技术，将各样品的 DNA 标签文

库混合物进行测序。

根据本发明的一个具体示例，可以将本发明的一组不同长度的DNA标签分别包含在用于扩增目的序列的PCR引物中，从而构成各自相对应的PCR标签引物，由此，本发明的构建DNA标签文库的方法可以包括：首先将总DNA样品利用机械法或酶切法打断成一定长度的DNA片段，并在将DNA片段与接头连接后，再通过PCR标签引物对含有DNA片段（即前面所述的目的序列）的连接产物进行扩增，最后通过琼脂糖电泳及切胶回收扩增产物，从而获得扩增产物构成的DNA标签文库。

根据本发明的一个具体示例，可以将本发明的一组不同长度的DNA标签分别嵌入现有文库的接头（例如末端）中，从而构成各自相对应的DNA标签接头，由此，本发明的构建DNA标签文库的方法可以包括：首先将总DNA样品利用机械法或酶切法打断成一定长度的DNA片段，并在将DNA片段进行末端修复并形成随机的粘性末端后，将其与DNA标签接头进行连接，以便获得连接产物，然后通过特定的PCR引物对含有DNA片段（即前面所述的目的序列）的连接产物进行扩增，最后通过琼脂糖电泳及切胶回收扩增产物，从而获得扩增产物构成的DNA标签文库。

根据本发明的一个具体示例，可以将本发明的一组不同长度的DNA标签分别包含在用于扩增目的序列的PCR引物中，从而构成各自相对应的PCR标签引物，同时将本发明的一组不同长度的DNA标签分别嵌入现有文库的接头（例如末端）中，从而构成各自相对应的DNA标签接头，由此，本发明的构建DNA标签文库的方法可以包括：首先将总DNA样品利用机械法或酶切法打断成一定长度的DNA片段，然后将DNA片段与DNA标签接头进行连接，以便获得连接产物，再通过PCR标签引物对含有DNA片段（即前面所述的目的序列）的连接产物进行扩增，最后通过琼脂糖电泳及切胶回收扩增产物，从而获得扩增产物构成的DNA标签文库。

根据本发明的实施例，本发明的提供的构建DNA标签文库，特别是Illumina/Solexa测序DNA标签文库的方法，使用选自表1所示的本发明的DNA标签接头用作DNA标签文库，特别是Illumina/Solexa测序DNA标签文库的3'接头。

根据本发明的再一方面，本发明还提供了一种用于构建DNA标签文库的试剂盒。根据本发明的实施例，该试剂盒包括：6种分离的寡核苷酸，这些分离的寡核苷酸具有第一链和第二链，其第一链分别由SEQ ID NO:（3N-1）所示的核苷酸构成，第二链分别由SEQ ID NO:（3N）所示的核苷酸构成，其中，对于相同的寡核苷酸，其第一链和第二链的N取值相同，并且N=1-6的任意整数，其中，6种分离的DNA标签接头分别设置在不同的容器中。由此，利用该试剂盒，能够方便地将根据本发明实施例的DNA标签引入到构建的DNA标签文库中。当然，本领域技术人员能够理解，试剂盒中还可以包含其他用于构建DNA标签文库的常规组件，在此不再赘述。

DNA标签文库及测序方法

根据本发明的又一方面，本发明还提供了一种DNA标签文库，其是根据本发明的构建DNA标签文库的方法所构建的。该具有根据本发明实施例的DNA标签的DNA标签文库可

以有效地应用于高通量测序技术例如 Solexa 技术，从而可以通过获得标签序列，来对所获得的核酸序列信息例如 DNA 序列信息来精确地进行样品来源分类，并且所得的测序数据准确、可靠，稳定性和可重复性好。此外，根据本发明的一些具体示例，本发明的 DNA 标签文库采用 DNA 标签接头作为 DNA 标签文库，特别是 Illumina/Solexa 测序标签文库的 3'接头。

根据本发明的又一方面，本发明还提供了一种确定 DNA 样品序列信息的方法。根据本发明的实施例，其包括：根据本发明实施例的构建 DNA 标签文库的方法，构建 DNA 标签文库；接着，对所构建的 DNA 标签文库进行测序，以便确定 DNA 样品的序列信息。基于该方法，能够有效地获得 DNA 标签文库中 DNA 样品的序列信息以及 DNA 标签的序列信息，从而能够对 DNA 样品的来源进行区分。另外，发明人惊奇地发现，利用根据本发明实施例的方法确定 DNA 样品序列信息，能够有效地减少数据产出偏向性的问题，并且能够精确地对多种 DNA 标签文库进行区分，获得的测序数据准确、可靠，稳定性和可重复性好。根据本发明的实施例，可以采用任何已知的方法对所构建的 DNA 标签文库进行测序，其类型并不受特别限制。根据本发明的一些示例，可以利用 Solexa 测序技术对 DNA 标签文库进行测序。根据本发明的实施例，可以根据具体情况选择合适的测序引物进行测序。

进一步，可以将上述确定 DNA 样品序列信息的方法应用于多种样品。例如，根据本发明的实施例，本发明提供了一种确定多种 DNA 样品序列信息的方法。根据本发明的实施例，其包括以下步骤：针对多种样品的每一种，分别独立地根据本发明的实施例的构建 DNA 标签文库的方法，构建该 DNA 样品的 DNA 标签文库，其中，不同的 DNA 样品采用相互不同并且已知序列的 DNA 标签；将多种样品的 DNA 标签文库进行组合，以便获得 DNA 标签文库混合物；对 DNA 标签文库混合物进行测序，以便获得 DNA 样品的序列信息以及所述标签组合的序列信息；以及基于标签组合的序列信息对 DNA 样品的序列信息进行分类，以便确定多种样品的 DNA 序列信息。这里，所使用的术语“多种”为至少 2 种。根据本发明的实施例，对 DNA 标签文库混合物进行测序的方法不受特别限制。根据本发明的一些具体示例，可以利用 Solexa 测序技术对所得的 DNA 标签文库混合物进行测序，从而获得 DNA 样品的序列信息以及标签的序列信息。由此，根据本发明实施例的该方法，可以充分利用高通量的测序技术，例如利用 Solexa 测序技术，同时对多种样品的 DNA 文库进行测序，从而能够提高 DNA 文库测序的效率和通量，并且能够有效地减少数据产出偏向性的问题，所得的测序数据准确、可靠，稳定性和可重复性好。

需要说明的是，根据本发明实施例的 DNA 标签及其应用，是本申请的发明人经过艰苦的创造性劳动和优化工作才完成的。

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。

对照例:

利用 Multiplexing Sample Preparation Oligonucleotide 试剂盒（PE-400-1001），参照试剂盒的标准流程，采用下表 2 中的 6 种常规标签（PE IndexN，其中 N=1-6 的任意整数）及其对应的常规标签接头（PE IndexN adapter，包括正义序列 PE IndexN adapter F 和反义序列 PE IndexN adapter R，其中 N=1-6 的任意整数）构建 DNA 标签文库:

表 2 常规标签（PE IndexN）和常规标签接头（PE IndexN_adapter）序列（表中所示序列方向均是 5’－3’方向）

名称	序列(SEQ ID NO:)
PE Index1	TGACCA(19)
PE Index1 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACCAT(20)
PE Index1 adapter R	5'-TGGTCAAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(21)
PE Index2	CAGATC(22)
PE Index2 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGATCT(23)
PE Index2 adapter R	5'-GATCTGAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(24)
PE Index3	GCCAAT(25)
PE Index3 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCAATT(26)
PE Index3 adapter R	5'-ATTGGCAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(27)
PE Index4	CTTGTA(28)
PE Index4 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGTAT(29)
PE Index4 adapter R	5'-TACAAGAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(30)
PE Index5	ACAGTG(31)
PE Index5 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTACAGTGT(32)
PE Index5 adapter R	5'-CACTGTAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(33)
PE Index6	ATCACG(34)
PE Index6 adapter F	TACACTCTTTCCCTACACGACGCTCTTCCGATCTATCACGT(35)
PE Index6 adapter R	5'-CGTGATAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC(36)

1、DNA 样品检测

首先，取 1~2μg 人类外周血基因组 DNA 样品，进行样品检测。具体地，使用 NanoDrop 1000 检测 DNA 样品的浓度、OD260/280 比值和 OD260/230 比值，并利用琼脂糖凝胶电泳检测样品的完整性，检测合格后，备用。

其中，DNA 样品质量合格的标准为:

样品浓度：样品的浓度最低不应低于 100ng/μl;

样品纯度：OD260/280 比值应在 1.8～2.0 之间，没有蛋白、多糖和 RNA 污染；

样品完整性：电泳结果显示 DNA 样品应没有降解；

然后，取适量检测合格的 DNA 样品，按以下步骤构建 DNA 标签文库，其中，为保证文库制备的质量，要求样品总量不低于 45μg:

2、片段化

将上述检测合格的 DNA 样品片段化，以便获得 DNA 片段。其中，将 DNA 样品片段化的方法有两种，分别是雾化法（Nebulization）和 Covaris 打断法，均可将样品 DNA 打碎至 100~800 bp 范围的片段且主带在 500bp 左右（若样品为已打断的 DNA，则可以跳过此步）。本对照例采用雾化法将DNA样品片段化，获得的DNA片段的长度为 100-800bp，主带 500bp。

3、末端修复

1）按照下表中的配比，在 1.5 ml 的离心管中配制末端修复反应体系：

DNA 片段	30 μl
超纯水(Ultra Pure Water)	45 μl
10x 多聚核苷酸激酶缓冲液	10 μl
dNTP 混合物（每种均为 10 mM）	4 μl
T4 DNA 聚合酶	5 μl
Klenow 片段	1 μl
T4 多聚核苷酸激酶	5 μl
总体积	100 μl

2）将上述配制的末端修复反应体系在 Thermomixer 中，于 20℃下进行温浴 30 min，以便获得经过末端修复的 DNA 片段。

3）其后使用 QIAquick PCR 纯化试剂盒（Qiagen），将经过末端修复的 DNA 片段进行柱纯化，并溶于 34μl 的洗脱缓冲液（Elution Buffer）中，备用。

4、3’末端添加碱基 A

1）按照下表中的配比，在 1.5 ml 的离心管中配制 3’末端添加碱基 A 的反应体系：

经过末端修复的 DNA 片段	32 μl
10x blue 缓冲液	5 μl
dATP	10 μl
Klenow (3’-5’ exo-)	3 μl
总体积	50 μl

2) 将上述配制的 3’末端添加碱基 A 的反应体系，在 Thermomixer 中，于 37℃下进行温浴 30 min，以便获得 3’末端添加碱基 A 的 DNA 片段。

3) 其后使用 QIAquick PCR 纯化试剂盒（Qiagen），将 3’末端添加碱基 A 的 DNA 片段进行柱纯化，并溶于 12μl 的洗脱缓冲液中，备用。

5、连接接头（Adapter）

1）按照下表中的配比，在 1.5 ml 的离心管中配制连接接头的反应体系：

3’末端添加碱基 A 的 DNA 片段	10 μl
2x Rapid 连接缓冲液	25 μl
常规标签接头混合物（PE IndexN Adapter Oligo Mix）	10 μl
T4 DNA Ligase	5 μl
总体积	50 μl

常规标签接头混合物，是通过将上表 2 所示的相对应的有义序列 PE IndexN adapter F 和反义序列 PE IndexN adapter R 经退火处理而形成的常规标签接头，进行组合获得的。

2）将上述配制的连接接头的反应体系，在 Thermomixer 中，于 20℃下进行温浴 15 min，以便获得连接产物。

3）其后使用 QIAquick PCR 纯化试剂盒（Qiagen），将连接产物进行柱纯化，并溶于 30μl 的洗脱缓冲液中，备用。

6、片段选择

1）将上述获得的连接产物于 2%琼脂糖凝胶上，100V 下进行电泳 120 min;

2）切取 620bp（切胶的片段长度可通过 n+120bp 计算，其中 n=插入片段大小）位置的胶块;

3）其后使用 QIAquick 凝胶提取试剂盒（Qiagen）回收切取的胶块，以便获得目的片段，然后将其溶于 40μl 的洗脱缓冲液中，备用。

7、 PCR 反应

1）按照下表中的配比，在 0.2 ml 的 PCR 管中配制 PCR 反应体系：

目的片段	10 μl
Phusion DNA 聚合酶	25 μl
PCR primer 1	1 μl
PCR primer 2	1 μl
超纯水（UltraPure™ Water）	13 μl
总体积	50 μl

其中，该 PCR 反应采用如下的 PCR 引物：

PCR Primer 1:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT(SEQ ID NO:39)，和

PCR Primer 2:

CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT(SEQ ID NO:40)。

2）然后，将上述配制的 PCR 反应体系，在热循环仪中运行下列程序：

PE 文库

98℃ 30s

98℃ 10s
65℃ 30s
72℃ 30s 或 50s
} 10 个循环

72℃ 5min

4℃ ∞

由此，获得扩增产物，备用。

8、分离回收扩增产物

将扩增产物在 2%琼脂糖凝胶上，以 100V 进行电泳 120 min，然后切取 620bp（切胶的片段长度可通过 n+120bp 计算，其中 n=插入片段大小）位置的胶块，其后使用 QIAquick 凝胶提取试剂盒（Qiagen）回收切取的胶块，以便获得 DNA 标签文库，然后将其溶于 40μl 的洗脱缓冲液中，备用。

实施例 1:

利用 Multiplexing Sample Preparation Oligonucleotide 试剂盒（PE-400-1001），参照试剂盒的标准流程，采用表 1 中的根据本发明实施例的 6 种 DNA 标签及其对应的 DNA 标签接头，构建 DNA 标签文库，具体步骤同对照例，其中，不同之处在于：

步骤 5，使用 DNA 标签接头混合物（Index Adapter Oligo mix）代替常规标签接头混合物（PE Index Adapter Oligo Mix），其中，DNA 标签接头混合物，是通过将表 1 所示的有义序列 IndexN adapter F 和反义序列 IndexN adapter R 经退火处理而形成的 DNA 标签接头，进行组合获得的。

步骤 7，PCR 反应的程序为：

98℃ 30s

98℃ 10s
65℃ 30s
72℃ 30s 或 50s
} 12 个循环

72℃ 5min

4℃ ∞

由此，获得 DNA 标签文库，备用。

实施例 2:

利用 HiSeq2000 测序仪，严格按照仪器推荐的流程，对对照例及实施例 1 所构建的 DNA 标签文库进行测序，以便获得测序数据。其中，所采用的测序引物为：

Sequencing Primer1:

5' – ACACTCTTTCCCTACACG ACGCTCTTCCGATCT（SEQ ID NO: 41）

图 2 则显示了利用 Solexa 测序平台对 DNA 标签文库进行测序时，各 Reads 的序列读取顺序及其序列组成的示意图。如图 2 所示，其中 Read1 表示测序反应 1 所测出来的序列，Read 1 Seq Primer 表示所采用的测序引物。

然后，利用数据处理软件 HiSeq Control Software 对获得的对照例及实施例 1 所构建的 DNA 标签文库的测序数据，进行处理、分析和比较，结果见图 3-图 6。其中数据处理软件包括但不限于 HiSeq Control Software（HCS）、Pipeline、CASAVA、SOAP 和 ELAND。

图 3 显示了实施例 1 的本发明的 DNA 标签文库与对照例的 DNA 标签文库的前 10 个测序循环（cycle）的质量值（Q20）的比较结果，其中，A 显示了实施例的本发明的 DNA 标签文库的前 10 个测序循环的质量值，B 显示了对照例的 DNA 标签文库的前 10 个测序循环的质量值。如图 3 所示，横坐标表示循环数，纵坐标表示质量情况。其中，质量值（Q-Value）可以反映测序质量，且其介于 0-40 之间，在此范围内，数值越高表示质量越好。而 Q20 是指质量值大于 20 的碱基在所有碱基中所占的比例，可以反映测序出来的序列质量好坏，数值越接近 1，说明测序质量越好。由图 3 可知，将本发明的 6 条呈梯度长度变化的 DNA 标签嵌入接头中，构建本发明的 DNA 标签文库，并使用 Illumina/Solexa 技术对 DNA 标签文库进行测序时，该本发明的 DNA 标签文库的前 10 个测序循环的质量值 Q20 一直都维持在 0.9；而将 6bp 固定长度的常规标签嵌入接头中，构建对照例的 DNA 标签文库，并使用 Illumina/Solexa 技术对 DNA 标签文库进行测序，在测第一个插入片段碱基即第 7 个循环时，由于碱基分布的变化，在该循环处出现了明显的质量下降现象，且该对照例的 DNA 标签文库的前 10 个测序循环的质量值 Q20 为 0.8。则图 3 表明，两个 DNA 标签文库的前 10 个测序循环（cycle）的质量值存在明显差异，且本发明的 DNA 标签文库的测序质量较好。

图 4-图 6 则是实施例 1 的本发明的 DNA 标签文库与对照例的标签文库的光强、碱基分布和错误率随循环数的变化的比较结果。图 4 显示了实施例 1 的 DNA 标签文库与对照例的标签文库的各测序循环的光强的比较结果，其中，A 显示了实施例 1 的本发明的 DNA 标签文库的各测序循环的光强信号平均值，B 显示了对照例的 DNA 标签文库的各轮测序循环的光强信号平均值。如图 4 所示，横坐标（Cycle）表示循环数，纵坐标表示光强信号平均值（Signal Mean）。图 5 显示了根据本发明实施例的本发明的 DNA 标签文库与对照例的 DNA 标签文库的 Solexa 测序数据沿 Reads 位置的碱基分布的比较结果，其中，A 显示了实施例 1 的本发明的 DNA 标签文库的 Solexa 测序数据沿 Reads 位置的碱基百分比分布图，B 显示了对照例的 DNA 标签文库的 Solexa 测序数据沿 Reads 位置的碱基百分比分布图。如图 5 所示，横坐标表示沿 Reads 的位置，纵坐标表示在该 Reads 的位置的各碱基百分比，该图显示了每次测序中测到的各种碱基比例。图 6 显示了根据本发明实施例的本发明的 DNA 标签文库与对照例的 DNA 标签文库的 Solexa 测序数据沿 Reads 位置的错误率的比较结果，其中，A 显示了实施例 1 的本发明的 DNA 标签文库的 Solexa 测序数据沿 Reads 位置的错误率图，B 显示了对照例的 DNA 标签文库的

Solexa 测序数据沿 Reads 位置的错误率图。如图 6 所示，横坐标表示沿 Reads 的位置，纵坐标表示错误率（即在该 Reads 的位置，测序错误发生的比例），实线表示错误率（即在该 Reads 的位置，测序错误发生的比例），虚线表示无法分析的碱基比例，该图显示了不同文库在错误率上的区别。

由图 4-图 6 可知，从光强、碱基分布和错误率随循环数的变化这三种参数来看，实施例 1 的本发明的 DNA 标签文库与对照例的标签文库并无明显差异。表明，使用本发明的呈梯度长度变化的 DNA 标签构建的 DNA 标签文库和使用固定长度的常规标签构建的 DNA 标签文库，在整体上并无明显差异，使用本发明的 DNA 标签并不会影响 DNA 标签文库的整体测序结果，且测序从梯度标签过渡到插入片段（Insert Fragment）时，能显著提高在过度处的碱基的质量值。由此，说明本发明的 DNA 标签，优于常规的多重核酸测序技术中的常规标签。

而对于 HiSeq2000 测序仪产量而言，假设碱基簇密度为 300 万/tile，PF 为 87%（其中，PF 即 Pass Filter，是指 Clean Data 占 Raw Data 的比例），则相对于对照例的标签文库，本发明的 DNA 标签文库，在一次 HiSeq 测序仪运行时可以增加 83.5M 的数据，并且能增加数据的可用性。

以上对本发明的 DNA 标签文库与对照例的标签文库的各参数的比较结果表明，本发明的 DNA 标签及 DNA 标签文库，具有显著的创造性。

工业实用性

本发明的用于构建 DNA 标签文库的 DNA 标签、DNA 标签接头、PCR 标签引物、DNA 标签文库及其制备方法、确定 DNA 样品序列信息的方法、确定多种 DNA 样品序列信息的方法以及用于构建 DNA 标签文库的试剂盒，能够有效地应用于样品 DNA 的 DNA 测序文库的构建以及测序，并且能够有效减少数据产出偏向性的问题，测序结果准确、可靠，稳定性和可重复性好。

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、一组分离的 DNA 标签，其由 SEQ ID NO:（3N-2）所示的核苷酸构成，其中 N=1-6 的任意整数。

2、一组分离的寡核苷酸，所述分离的寡核苷酸具有第一链和第二链，其中所述第一链分别由 SEQ ID NO:（3N-1）所示的核苷酸构成，所述第二链分别由 SEQ ID NO:（3N）所示的核苷酸构成，其中，对于相同的寡核苷酸，其第一链和第二链的 N 取值相同，并且 N=1-6 的任意整数。

3、一组分离的 PCR 标签引物，所述分离的 PCR 标签引物具有具有第一链和第二链，其中所述第一链分别由 SEQ ID NO: 39 所示的核苷酸构成，所述第二链的核苷酸序列如下所示:

CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT，

其中，所述 XXXXXX 为权利要求 1 所述的一组分离的 DNA 标签的一种。

4、一种构建 DNA 标签文库的方法，其特征在于，包括以下步骤:

将 DNA 样品片段化，以便获得 DNA 片段;

将所述 DNA 片段进行末端修复，以便获得经过末端修复的 DNA 片段;

将所述经过末端修复的 DNA 片段进行 3’末端添加碱基 A，以便获得 3’末端添加碱基 A 的 DNA 片段;

将所述 3’末端添加碱基 A 的 DNA 片段与 DNA 标签接头相连，以便获得连接产物，其中所述 DNA 标签接头包含选自权利要求 1 所述的一组分离的 DNA 标签的一种;

将所述连接产物进行扩增，以便获得扩增产物；以及

分离回收所述扩增产物，所述扩增产物构成所述 DNA 标签文库。

5、根据权利要求 4 所述的方法，其特征在于，所述片段化是通过选自雾化法、酶切法和超声打断法的至少一种进行的。

6、根据权利要求 5 所述的方法，其特征在于，利用超声打断法将所述 DNA 样品片段化。

7、根据权利要求 6 所述的方法，其特征在于，利用 Covaris 打断仪将所述 DNA 样品片段化。

8、根据权利要求 4 所述的方法，其特征在于，所述 DNA 片段的长度为大约 500bp。

9、根据权利要求 4 所述的方法，其特征在于，所述 DNA 样品为人 DNA 样品。

10、根据权利要求 9 所述的方法，其特征在于，所述 DNA 样品为人基因组 DNA 样品。

11、根据权利要求 4 所述的方法，其特征在于，所述末端修复是利用 Klenow 片段、T4 DNA 聚合酶和 T4 多核苷酸激酶进行的，其中所述 Klenow 片段具有 5’→3’聚合酶活性和 3’→5’外切酶活性，但缺少 5’→3’外切酶活性，

12、根据权利要求 3 所述的方法，其特征在于，利用 Klenow (3'-5' exo-)将所述经过末端修复的 DNA 片段进行 3'端添加碱基 A。

13、根据权利要求 12 所述的方法，其特征在于，将所述经过末端修复的 DNA 片段的两条寡核苷酸链的 3'末端均添加碱基 A。

14、根据权利要求 4 所述的方法，其特征在于，将所述 3'末端添加碱基 A 的 DNA 片段与 DNA 标签接头相连是利用 T4 DNA 连接酶进行的。

15、根据权利要求 4 所述的方法，其特征在于，在所述 3'末端添加碱基 A 的 DNA 片段的 3'末端和 5'末端的至少一处连接 DNA 标签接头。

16、根据权利要求 15 所述的方法，其特征在于，在所述 3'末端添加碱基 A 的 DNA 片段的 3'末端连接 DNA 标签接头。

17、根据权利要求 4 所述的方法，其特征在于，所述 DNA 标签接头为选自权利要求 2 所述的分离的寡核苷酸的一种。

18、根据权利要求 4 所述的方法，其特征在于，将所述连接产物进行扩增之前，进一步包括对所述连接产物进行片段选择的步骤。

19、根据权利要求 18 所述的方法，其特征在于，利用 2%琼脂糖凝胶电泳进行所述片段选择。

20、根据权利要求 19 所述的方法，其特征在于，所述连接产物的长度为约 620bp。

21、根据权利要求 4 所述的方法，其特征在于，将所述连接产物进行扩增，是通过 PCR 反应进行的，所述 PCR 反应采用分别具有如 SEQ ID NO: 39 和 SEQ ID NO: 40 所示核苷酸序列的引物。

22、根据权利要求 4 所述的方法，其特征在于，通过选自琼脂糖凝胶电泳、磁珠纯化和纯化柱纯化的至少一种分离所述扩增产物。

23、根据权利要求 4 所述的方法，其特征在于，所述扩增产物的长度为大约 620bp。

24、一种 DNA 标签文库，其是通过根据权利要求 3-23 任一项所述的方法获得的。

25、一种确定 DNA 样品序列信息的方法，其包括下列步骤:

根据权利要求 3-23 任一项所述的方法构建所述 DNA 样品的 DNA 标签文库；以及

对所述 DNA 标签文库进行测序，以便确定所述 DNA 样品的序列信息。

26、根据权利要求 25 所述的方法，其特征在于，利用 Solexa 测序技术对所述 DNA 标签文库进行测序。

27、一种确定多种 DNA 样品序列信息的方法，其包括下列步骤:

针对所述多种 DNA 样品的每一种，分别独立地根据权利要求 3-23 任一项所述的方法，构建所述 DNA 样品的 DNA 标签文库，其中，不同的 DNA 样品采用相互不同并且已知序列的 DNA 标签，其中所述多种为 2-6 种;

将所述多种 DNA 样品的 DNA 标签文库进行组合，以便获得 DNA 标签文库混合物;

对所述 DNA 标签文库混合物进行测序，以便获得所述 DNA 样品的序列信息以及所述标签的序列信息；以及

基于所述标签的序列信息对所述 DNA 样品的序列信息进行分类，以便确定所述多种样品的 DNA 序列信息。

28、根据权利要求 27 所述的方法，其特征在于，利用 Solexa 测序技术对所述 DNA 标签文库混合物进行测序。

29、一种用于构建 DNA 标签文库的试剂盒，其包括:

6 种分离的寡核苷酸，所述分离的寡核苷酸具有第一链和第二链，其中所述第一链分别由 SEQ ID NO:（3N-1）所示的核苷酸构成，所述第二链分别由 SEQ ID NO:（3N）所示的核苷酸构成，其中，对于相同的寡核苷酸，其第一链和第二链的 N 取值相同，并且 N=1-6 的任意整数，

其中，所述 6 种分离的寡核苷酸分别设置在不同的容器中。

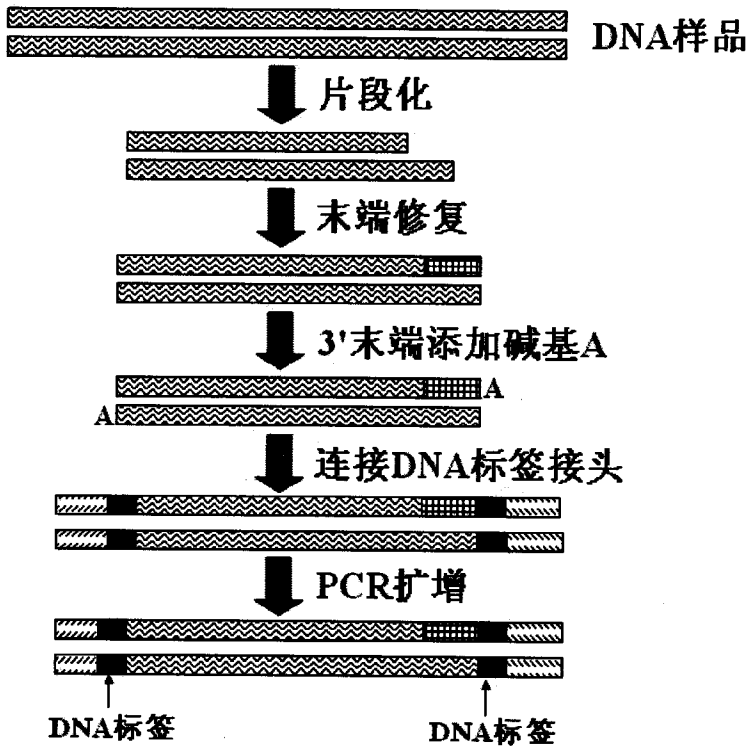


图 1

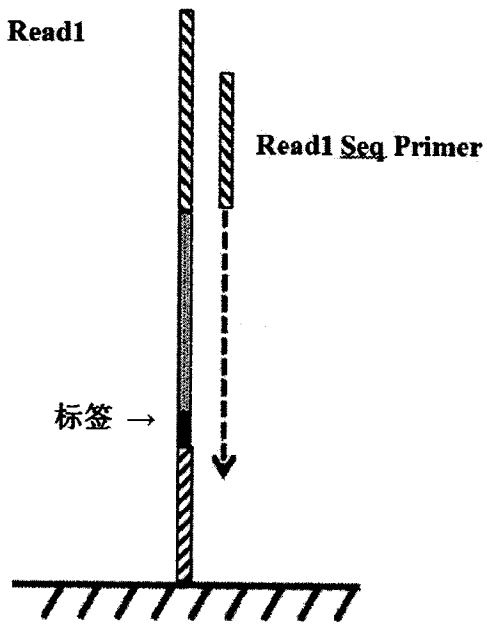


图 2

替换页(细则第26条)

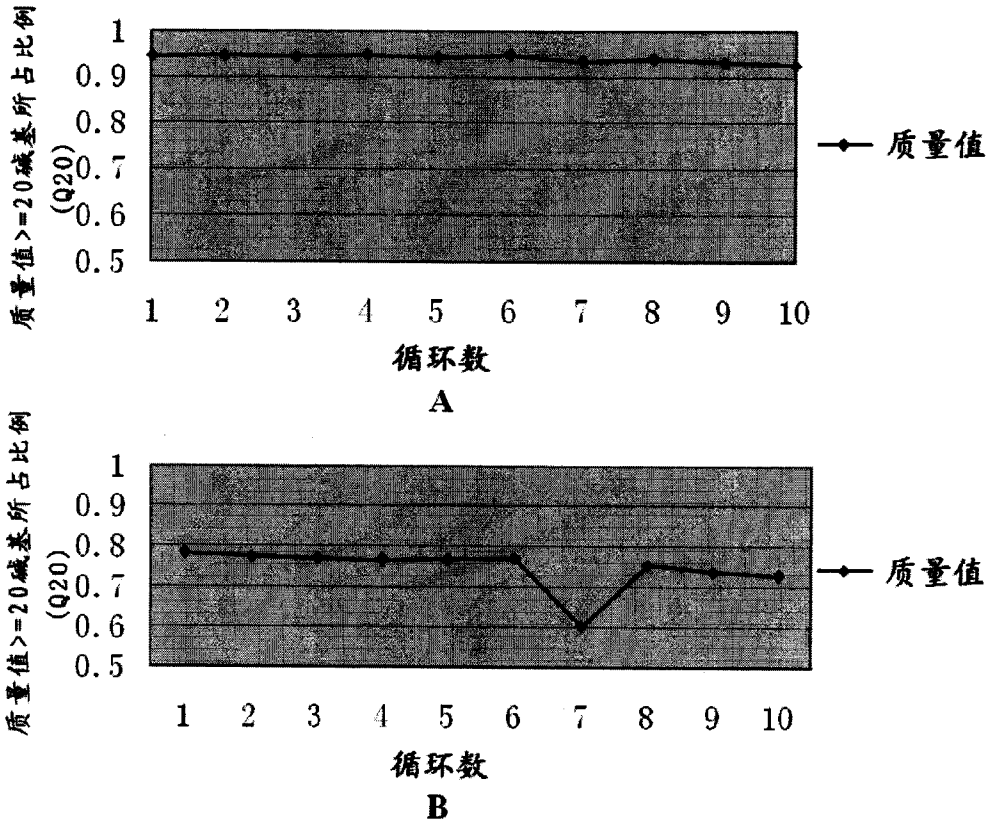


图 3

替换页(细则第26条)

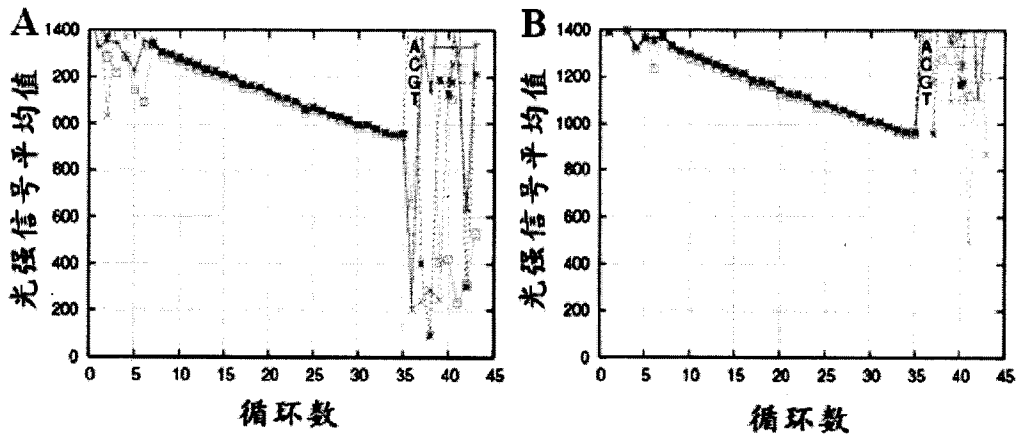


图 4

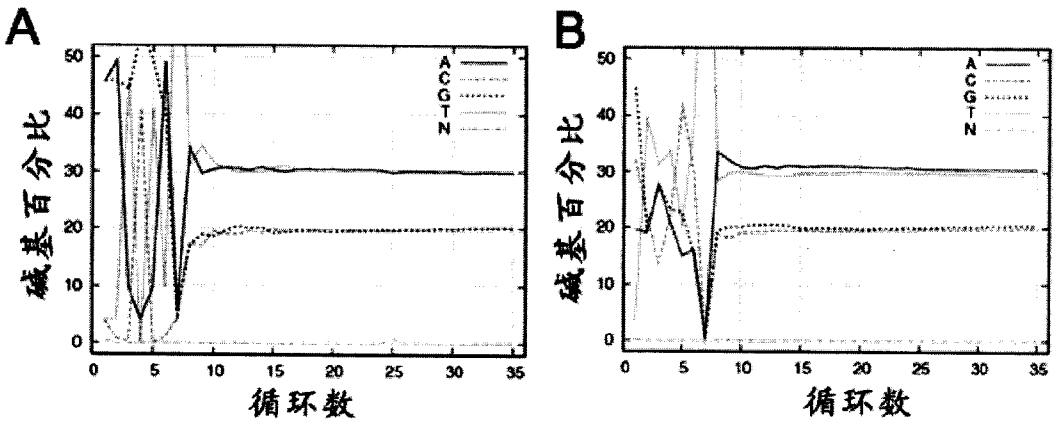


图 5

替换页(细则第26条)

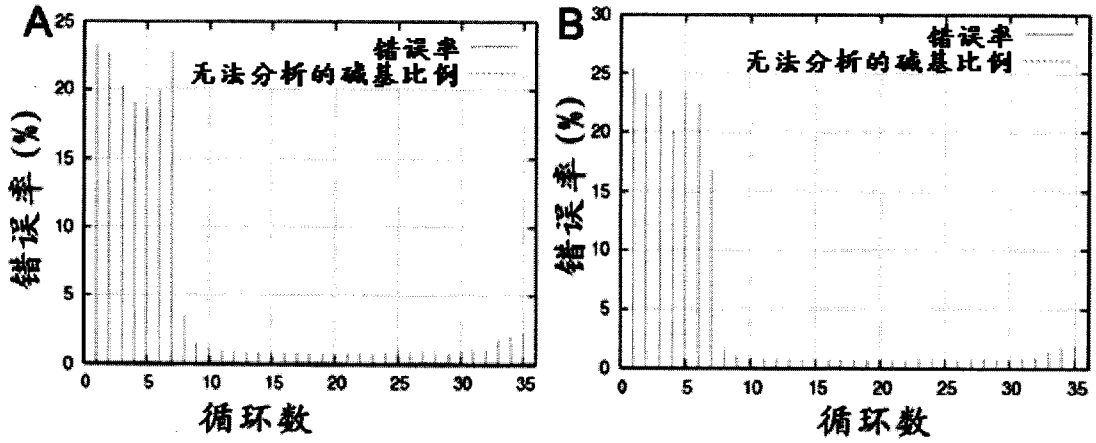


图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2012/071893

A. CLASSIFICATION OF SUBJECT MATTER

C12Q1/68（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)

CNPAT, CNKI, DWPI, SIPOABS, CPRSABS, CNABS, CNTXT, VEN, CJFD, ISI Web of Knowledge, NCBI: multiplex sequencing, adapter, index, library, sequence search on SEQ ID NO：1-18、39-40.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Richard Cronn et al., Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by synthesis technology，Nucleic Acids Research, 27.Aug.2008, Vol.36, No.19, e122.	1-29
A	David W. Craig et al., Identification of genetic variants using bar-coded multiplexed sequencing, Nature Methods, 14.Sept.2008，Vol.5, No.10, pages 887-893.	1-29
A	Andrew M. Smith et al.,Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples, Nucleic Acids Research, 11.May 2010., Vol.28, No.13.,e142.	1-29
A	US2006263789 A1 (KINCAID R) 23.Nov. 2006 (23.11.2006), see the whole document.	1-29
A	WO2005054504 A1 (UNIV KOELN) 16.Jun. 2005 (16.06.2005), see the whole document.	1-29

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

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"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

"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

"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

"&"document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02.Jun. 2012（02.06.2012）	14.Jun. 2012（14.06.2012）
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	Authorized officer ZHANG, Yanxia Telephone No. (86-10) **62411049**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2012/071893

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO03052137A2（GNOTHIS HOLDING SA） 26.6 月 2003（26.06.2003）, see the whole document	1-29

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2012/071893

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO2005054504 A1	16.06.2005	DE10356783 A1	07.07.2005
WO03052137 A2	26.06.2003	AU2002361162 A1	30.06.2003
		EP1458892 B1	03.08.2011
		DE10162536 A1	17.07.2003
		US2007020626 A1	25.01.2007
		EP1458892 A2	22.09.2004
		US8017319 B2	13.09.2011
US2006263789 A1	23.11.2006	NONE	

国际检索报告

国际申请号
PCT/CN2012/071893

A. 主题的分类

C12Q1/68（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）)

CNPAT，CNKI，DWPI，SIPOABS，CPRSABS，CNABS，CNTXT，VEN，CJFD，ISI Web of Knowledge，NCBI: 多重测序，DNA 接头，DNA 标签，文库，multiplex sequencing，adapter，index，library，基于 SEQ ID NO：1－18、39-40 的序列检索

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	Richard Cronn 等，Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by synthesis technology，Nucleic Acids Research，27.8 月 2008，第 36 卷第 19 期，e122	1-29
A	David W. Craig 等，Identification of genetic variants using bar-coded multiplexed sequencing，Nature Methods, 14. 9 月 2008，第 5 卷第 10 期，第 887-893 页	1-29
A	Andrew M. Smith 等，Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples，Nucleic Acids Research，11.5 月 2010，第 38 卷第 13 期，e142	1-29
A	US2006263789 A1（KINCAID R） 23.11 月 2006（23.11.2006）参见全文	1-29
A	WO2005054504 A1（UNIV KOELN） 16.6 月 2005（16.06.2005）参见全文	1-29
A	WO03052137A2（GNOTHIS HOLDING SA） 26.6 月 2003（26.06.2003）参见全文	1-29

☐ 其余文件在 C 栏的续页中列出。　　☒ 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	国际检索报告邮寄日期
02.6 月 2012（02.06.2012）	**14.6 月 2012 (14.06.2012)**
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 张艳霞 电话号码：(86-10) **62411049**

国际检索报告

关于同族专利的信息

国际申请号
PCT/CN2012/071893

检索报告中引用的专利文件	公布日期	同族专利	公布日期
WO2005054504 A1	16.06.2005	DE10356783 A1	07.07.2005
WO03052137 A2	26.06.2003	AU2002361162 A1	30.06.2003
		EP1458892 B1	03.08.2011
		DE10162536 A1	17.07.2003
		US2007020626 A1	25.01.2007
		EP1458892 A2	22.09.2004
		US8017319 B2	13.09.2011
US2006263789 A1	23.11.2006	无	