

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12Q 1/68 (2006.01)



[12] 发明专利说明书

专利号 ZL 200610096705.3

[45] 授权公告日 2008 年 9 月 10 日

[11] 授权公告号 CN 100417730C

[22] 申请日 2006.10.10

[21] 申请号 200610096705.3

[73] 专利权人 东南大学

地址 210096 江苏省南京市四牌楼 2 号

[72] 发明人 施小龙 唐超 陆祖宏

[56] 参考文献

US6692915B 2004.2.17

CN1245218A 2000.2.23

Detection of mitochondrial single nucleotide polymorphisms using a primer elongation reaction on oligonucleotide microarrays. . Erdogan F, et al. Nucleic Acids Res. , Vol. 29 No. 7. 2001

用引物延伸芯片法实现对转基因水稻中质粒 pCAMBIA1301 的检测. 高秀丽等. 遗传, 第 27 卷第 2 期. 2005

审查员 苏林

[74] 专利代理机构 南京经纬专利商标代理有限公司

代理人 陆志斌

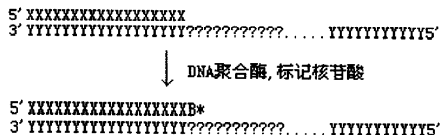
权利要求书 1 页 说明书 10 页 附图 3 页

[54] 发明名称

基于引物延伸的 DNA 测序方法

[57] 摘要

核酸序列测定方法：在被测的 DNA 模板中引入一段已知的 DNA 序列，使之固定于固相基质上；设计 1-15 个寡核苷酸 DNA 引物组合，使其含有一段与上述加入的 DNA 序列互补的序列，使该 1-15 个寡核苷酸 DNA 引物的 3' 端分别含有 0、1、2、……、14 个简并性核苷酸；根据待测核苷酸位点的位置，从上述引物组合中选择一种引物，将其与固定 DNA 模板杂交，使待测核苷酸位点挨着引物 3' 末端；用 DNA 聚合酶及用四种可分辨的标记物标记的双脱氧核苷酸底物或分别用一种或几种标记物标记的四种脱氧核苷酸或双脱氧核苷酸，在杂交到模板上的引物上进行核苷酸延伸；检测被延伸核苷酸种类，DNA 模板被测碱基的种类与延伸核苷酸种类互补。



1、一种用于确定未知核酸序列的核酸序列测定方法，其特征在于：A)在被测的 DNA 模板中引入一段已知的通用 DNA 序列，并使之固定于固相基质上；B)设计 1-15 个寡核苷酸 DNA 引物组合，使其含有一段与上述加入的通用 DNA 序列互补的序列，并使该 1-15 个寡核苷酸 DNA 引物的 3' 端分别含有 0、1、2、...、14 个简并性核苷酸；C)根据待测核苷酸位点的位置，从上述引物组合中选择一种引物，并将其与步骤 A) 中所述的固定 DNA 模板杂交，使待测核苷酸位点紧接着引物 3' 末端；D)用 DNA 聚合酶及用四种可分辨的标记物标记的双脱氧核苷酸底物或分别用一种或几种标记物标记的四种脱氧核苷酸或双脱氧核苷酸，在杂交到模板上的引物上进行一种核苷酸延伸；E)检测步骤 D) 中被延伸核苷酸种类，步骤 A) 中 DNA 模板被测碱基的种类与延伸核苷酸种类互补。

2、根据权利要求 1 所述的核酸序列测定方法，其特征在于依次用含不同个数的简并性核苷酸的引物重复步骤 A) - E) 操作，检测不同位置核苷酸。

3、根据权利要求 1 的核酸序列测定方法，其特征在于步骤 D) 所述的标记的核苷酸为用于聚合酶延伸的标记有荧光化学基团的核苷酸。

4、根据权利要求 1 所述的核酸序列测定方法，其特征在于通过生物素、地高辛连接，用于显色反应的酶或纳米金颗粒标记的抗体或亲和素来标记核苷酸。

5、如权利要求 1 所述的核酸序列测定方法，其特征在于步骤 A) 所述的固相基质是指平板基质、微孔板或微球。

6、如权利要求 1 的核酸序列测定方法，其特征在于步骤 A) 所述的被测的 DNA 模板为 PCR 产物，滚环扩增产物，DNA 质粒或酶加工后的 DNA 片断。

7、如权利要求 1 的核酸序列测定方法，其特征在于步骤 A) 所述的引入一段已知的通用 DNA 序列的方法是：通过 DNA 连接酶在待测的序列两端加入接头，该接头为和引物杂交的已知序列。

8、如权利要求 1 的核酸序列测定方法，其特征在于步骤 A) 所述的引入一段已知的通用 DNA 序列的方法是：将待测的 DNA 连接到载体中，然后转化大肠杆菌，得到单菌落克隆后进行以插入 DNA 序列两端的载体序列作为引物进行 PCR，从而使载体的一部分已知序列引入待测 DNA 序列。

9、如权利要求 1 的核酸序列测定方法，其特征在于步骤 A) 所述的引入一段已知的通用 DNA 序列的方法是：对已知部分序列信息的待测 DNA 进行测序，可直接以已知序列部分作为引物结合区，待测部分紧接着引物结合区。

基于引物延伸的 DNA 测序方法

技术领域

本发明涉及一种 DNA 序列测定方法,尤其涉及一种基于引物延伸的 DNA 测序方法。

背景技术

目前两个最常用的 DNA 测序方法是 Sanger 的双脱氧链终止法及 Maxam 和 Gilbert 的化学切割测序法,但二者都依赖于电泳分离大小不同的 DNA 片段。Sanger 测序法应用了双脱氧核苷酸随机终止聚合反应,这样产生了各种长度的延伸链,通过电泳来分离。由于终止处的核苷酸可以被确认,而每一处都有终止处,因此可以读出序列。虽然电泳技术特别是毛细管电泳的应用使得测序的自动化程度得到了很大提高,但其通量仍受到一定程度的限制,因此对于需要高额费用和高通量的大规模基因组计划或诊疗测序来说不是最适宜的方法。近年来随着基因芯片在生命科学领域中迅速发展,其高通量检测的优点被引用到 DNA 测序中,并发展或提出了一些非电泳的测序方法,例如芯片杂交测序法,焦磷酸测序法,光切割荧光核苷酸法。

芯片杂交测序法利用芯片上高密度的寡核苷酸序列点阵和靶序列杂交,然后拼接出靶 DNA 的序列,需要合成非常高密度的点阵才能有效测出靶 DNA 序列,而高密度点阵的制作目前仍很困难。

焦磷酸测序法和光切割荧光核苷酸测序法的共同点是采用 DNA 聚合酶进行逐步延伸引物链,属于合成法测序 (sequencing by synthesis)。因此先对合成法测序过程作一般概述:首先制得单链模板 DNA,引物和模板 DNA 链退火得到杂交的 DNA 分子,然后在 DNA 聚合酶存在下加入核苷酸底物,每次只加入一种核苷酸底物或四种带标记的双脱氧核苷酸底物,如果加入的核苷酸和模板上对应的碱基互补,则聚合酶会将之整合到引物链。核苷酸的整合事件可以通过各种方法检测,例如焦磷酸测序是通过检测焦磷酸的释放;也可以将容易检测的标记物如发荧光的物质连接到核苷酸上,然后检测。

焦磷酸测序法 (pyrosequencing) 是通过生物催化发光来确定是否有核苷酸掺入,其原理是加入的核苷酸底物和模板碱基互补时, DNA 聚合酶催化其聚合,同时会释放出焦磷酸,焦磷酸被酶转化为 ATP, ATP 在荧光素酶的作用下会产生光子而被高灵敏光电探测器检测到。它通过底物加入顺序来确定核苷酸种类,是一种即时 (real-time) 测序法。该方法多个酶促反应和生物芯片的高通量检测结合起来对于基因组测序有一定的优势。但其存在测序长度受限 (50bp 左右) 和对多个连续同样的核苷酸不能准确测序,而且需要特定的仪器装备。

一些研究者合成了许多光切基团或化学试剂切割基团用于 DNA 测序研究。采取的路线基本是用 DNA 合成酶掺入荧光标记物后检测,然后用激光或其他化学试

剂去除荧光基团。之后加入下一个标记底物。也有研究将光切割和核苷酸 3' -OH 用化学基团保护结合起来, 这样可以实现每次只掺入一个核苷酸进入引物链。光切割荧光核苷酸法借鉴了 DNA 和多肽化学合成的思想, 但采用的催化剂是高度特异性酶, 由于 3' -OH 端被保护, 每次只能掺入一个核苷酸, 每次延伸的核苷酸的种类由模板来决定, 但我们可以通过其荧光颜色来得知每次掺入的是何种核苷酸, 因此序列可以得知。但由于 3' -OH 涉及磷酸二酯键的形成, 核苷酸的 3' -OH 非常靠近 DNA 聚合酶的活性位点, 因此对 3' -OH 的修饰非常敏感, 很少 DNA 聚合酶能够催化该底物的聚合。这些路线的缺点一是需要合成复杂的切割基团; 二是荧光基团不一定会完全切掉和脱掉, 当延伸到一定长度会产生背景干扰。光切割荧光核苷酸法是最近新发展起的新技术, 还在起步阶段, 其光切割荧光核苷酸还没有商品化, 且其准确度和长度还待进一步确定。

单碱基延伸反应 (single base extension, SBE), 又名微测序法 (minisequencing)。根据待检测样品的 SNP 位点序列设计引物 (不包括 SNP 位点) 并将其直接固定在芯片上, 以待测样品的 PCR 产物作为模板, 则寡核苷酸引物的 3' 端碱基紧挨于模板多态性碱基位点, 当模板和引物杂交后, 加以用各色荧光或者其他标记物标记的 4 种 ddNTP, 在 DNA 多聚酶的催化下与模板配对的引物直接在芯片上进行固相单碱基延伸反应, 通过检测荧光的种类得知掺入 SNP 位点的核苷酸底物的种类。DNA 多聚酶的特异性使该方法的敏感度和分辨率均较高, 但其缺点是必须使用多色荧光系统以及相应的检测系统, 而且多种染料的激发光和发射荧光的光谱往往会有较大部分的重叠, 从而干扰对荧光强度的测量精度。

发明内容

本发明的目的在于克服上述所提到的核酸测序方法中的不足, 提出一种基于生物芯片的高通量基于引物延伸的 DNA 测序方法, 具有测序价格低、准确性高和可操作性强的优点。

本发明采用如下技术方案:

一种用于确定未知核酸序列的核酸序列测定方法: A) 在被测的 DNA 模板中引入一段已知的通用 DNA 序列, 并使之固定于固相基质上; B) 设计 1-15 个寡核苷酸 DNA 引物组合, 使其含有一段与上述加入的通用 DNA 序列互补的序列, 并使该 1-15 个寡核苷酸 DNA 引物的 3' 端分别含有 0、1、2、……、14 个简并性核苷酸; C) 根据待测核苷酸位点的位置, 从上述引物组合中选择一种引物, 并将其与 A 中所述的固定 DNA 模板杂交, 使待测核苷酸位点紧接着引物 3' 末端; D) 用 DNA 聚合酶及用四种可分辨的标记物标记的双脱氧核苷酸底物或分别用一种或几种标记物标记的四种脱氧核苷酸或双脱氧核苷酸, 在杂交到模板上的引物上进行一种核苷酸延伸; E) 检测 D 中被延伸核苷酸种类, A 中 DNA 模板被测碱基的种类与延伸核苷酸种类互补。

与现有技术相比, 本发明具有如下优点:

1、高通量。本发明将 DNA 测序技术和生物芯片相结合, 可以同时芯片上

固定数千到数十万个样品，可以对数千个样品进行测序。

2、便宜。本发明只需要数套测序引物组合，引物组合的套数与测序长度相同，可以对大量的样品进行测序。

3、方便。本发明采取简并性引物延伸法测序，可以方便地测出引物一侧 8—9 个碱基的序列。如果和特异性限制内切酶结合使用可以测得更长的序列。

4、准确性高。

附图说明

图 1 是本发明的用含 0 个简并性核苷酸的引物延伸示意图。X 行代表引物链，和引物结合的区域为引物结合区(在模板链中)；Y 行代表模板链，?代表待测得序列，B*代表掺入的标记核苷酸。

图 2 是本发明的用含 1 个简并性核苷酸的引物延伸示意图。引物链中 3' 端有 1 个简并性核苷酸(N)；经过 DNA 聚合酶的聚合，和模板中第二位置的碱基(待测的碱基)互补的标记核苷酸 B*被共价掺入引物链。与 B*互补的核苷酸即为模板中第二位置?的核苷酸。

图 3 是本发明的用含 2 个简并性核苷酸的引物延伸示意图。物链中 3' 端有 2 个简并性核苷酸(NN)；经过 DNA 聚合酶的聚合，和模板中第三位置的碱基(待测的碱基)互补的标记核苷酸 B*被共价掺入引物链。与 B*互补的核苷酸即为模板中第三位置?的核苷酸。

图 4 是本发明的用含 7 个简并性核苷酸的引物延伸示意图。引物链中 3' 端有 7 个简并性核苷酸(NNNNNNN)；经过 DNA 聚合酶的聚合，和模板中第八位置的碱基(待测的碱基)互补的标记核苷酸 B*被共价掺入引物链。与 B*互补的核苷酸即为模板中第八位置?的核苷酸。

图 5 是本发明实施例 2 的芯片扫描结果图。每个方阵有 4 行，从上至下分别为 seq1, seq2, seq3 和 seq4。每个序列重复 4 次。A, C, G, T 分别代表所加的标记核苷酸，P1-P7 代表 seq5~seq11 引物(和 seq1~seq4 杂交后延伸)。

图 6 是本发明实施例 2 的荧光强度柱状图。图中 seq1、seq2、seq3 和 seq4 分别为 4 个合成的模板，与材料方法中所列序列相对应。X 轴的数字 1-7 分别和所加入的测序引物 seq5—seq11 相对应；Y 轴为扫描后所得到的荧光强度。通过比较荧光强度可以在 seq1 中读出序列：CTACCTG；在 seq2 中读出序列为：GACGCAC；在 seq3 中读出序列为：TCGTCTG；在 seq4 中读出：AGGACTA。

图 7 是本发明实施例 3 的 PCR 产物琼脂糖凝胶电泳结果图。

图 8 是本发明实施例 3 的芯片扫描结果图。

图 9 是本发明实施例 3 用 Sanger 法测序测序部分结果图。

具体实施方式

实施例 1

一种用于确定未知核酸序列的核酸序列测定方法：

A) 在被测的 DNA 模板中引入一段已知的通用 DNA 序列, 并使之固定于固相基质上; B) 设计 1-15 个寡核苷酸 DNA 引物组合, 使其含有一段与上述加入的通用 DNA 序列互补的序列, 并使该 1-15 个寡核苷酸 DNA 引物的 3' 端分别含有 0、1、2、……、14 个简并性核苷酸; C) 根据待测核苷酸位点的位置, 从上述引物组合中选择一种引物, 并将其与 A 中所述的固定 DNA 模板杂交, 使待测核苷酸位点紧接着引物 3' 末端; D) 用 DNA 聚合酶及用四种可分辨的标记物标记的双脱氧核苷酸底物或分别用一种或几种标记物标记的四种脱氧核苷酸或双脱氧核苷酸, 在杂交到模板上的引物上进行一种核苷酸延伸; E) 检测 D 中被延伸核苷酸种类, A 中 DNA 模板被测碱基的种类与延伸核苷酸种类互补。

本实施例可以依次用含不同个数的简并性核苷酸的引物重复 A-E 操作, 检测不同位置核苷酸。

上述步骤 D) 所述的标记的核苷酸为用于聚合酶延伸的标记有荧光化学基团的核苷酸。

标记核苷酸的方法是通过生物素、地高辛连接, 用于显色反应的酶或纳米金颗粒标记的抗体或亲和素来标记核苷酸。

上述步骤 A) 所述的固体基质是指平板基质、微孔板或微球。

上述步骤 A) 所述的被测的 DNA 模板为 PCR 产物, 滚环扩增产物, DNA 质粒或酶加工后的 DNA 片断。

上述步骤 A) 所述的引入一段已知的通用 DNA 序列的方法可以是下列之一:

(1) 通过 DNA 连接酶在待测的序列两端加入接头, 该接头为和引物杂交的已知序列; (2) 将待测的 DNA 连接到载体中, 然后转化大肠杆菌, 得到单菌落克隆后进行以插入 DNA 序列两端的载体序列作为引物进行 PCR, 从而使载体的一部分已知序列引入待测 DNA 序列; (3) 对已知部分序列信息的待测 DNA 进行测序, 可直接以已知序列部分作为引物结合区, 待测部分紧接着引物结合区。

下面结合图 1—图 4, 对本发明的具体实施方式作出更为详细的描述:

首先将待测的 DNA 样品固定在生物芯片基片上, 通过变性产生 DNA 单链模板。模板上有已知序列的引物结合区。待测的 DNA 序列为紧接着引物结合区 5' 方向一侧的序列, 如图 1 中所示的标有“?”的区域。

为了测得模板中紧接着引物结合区 5' 方向后第一个位置的碱基的种类, 用 3' 端紧接着待测位点的引物和模板杂交, 然后用 DNA 聚合酶、标记的双脱氧核苷酸或脱氧核苷酸及反应缓冲液延伸 (见图 1)。延伸时可用四种不同标记信号的双脱氧核苷酸底物同时延伸; 也可用一种标记信号标记的四种双脱氧核苷酸或脱氧核苷酸分别进行四次延伸。延伸后检测, 根据标记信号及碱基互补配对原则 (即 A 和 T, C 和 G 互补配对), 则第一个位置的碱基种类可以确定。

为了测得模板中紧接着引物结合区 5' 方向后第二个位置的碱基的种类, 所用的引物包括两个部分, 一部分为和引物结合区的已知序列, 一部分为和未知序列结合的区。因为引物结合区 5' 方向后第一个位置的为未知序列, 其种类的可能性有四种。因此引物中和其对应的碱基也有四种。在合成引物时该位置为简并

性核苷酸, 即该位置为含核苷酸 A, G, C 和 T 的混合物, 如图 2 所示用 N 来表示。当该引物和待测模板杂交后, 用标记的核苷酸及 DNA 聚合酶延伸, 由于 DNA 聚合酶的特异性选择, 当标记底物核苷酸和第二个位点的碱基互补时, 标记底物会掺入引物链 (在图 2 中用 B* 表示), 之后用检测仪器检测, 根据标记物性质或标记核苷酸加入顺序, 就可以确定是何种碱基掺入引物链。

为了测得模板中紧接着引物结合区 5' 方向后第三个位置的碱基的种类, 用 3' 端有两个简并性核苷酸的引物杂交, 即有两个 N 的引物和模板杂交。见图 3。之后用 DNA 聚合酶及标记底物延伸, 然后检测。

如此类推, 当测第四个碱基是用 3' 端有三个简并性核苷酸的引物杂交, 测第八个碱基时用 3' 端有七个简并性核苷酸的引物杂交 (见图 4)。

由以上方法可知, 本发明可以测得引物结合区一侧特定位置碱基的种类。为了测得待测位点的碱基种类, 可用方案一: 将待测模板分成若干份, 固定在芯片的不同位置上, 或几张重复制作的芯片上。每个位置分别用含有不同个数的简并性引物杂交后延伸, 可以同时获得待测的不同位置的碱基种类的信息。也可以用方案二: 先用其中的一种引物和固定的模板杂交后延伸, 待检测后通过变性处理使引物链和模板链分离, 然后加入第二种引物杂交后延伸, 如此反复进行多次。以上提供的方法可以非常方便的测出引物结合区一侧 10 个左右的碱基序列的信息。为了测出更长的碱基序列, 可以结合限制性内切酶的酶切位点进行测序。也就是测定特定酶切位点一侧的序列, 酶切位点一般具有 4-6 个特定碱基组成的序列, 因此可以有效地增加测序的长度。

本发明提供的测序方法是将待测的双链 DNA 样品共价固定到生物芯片基质上。生物芯片基质是指平板基质 (如玻璃片、硅片、凝胶层、塑料、橡胶、陶瓷、硝酸纤维素膜、尼龙膜等)、微孔板 (如塑料板、金属板、橡胶板、玻璃板等)、和微球 (如亲和素包被的磁珠)。有多种方法可以将 DNA 样品固定至生物芯片的片基表面。首先, 将生物芯片基片进行修饰, 使其表面带有醛基、氨基、巯基、羧基、环氧基等其他活性基团, 或用链霉亲和素、亲和素等包被, 然后将用氨基、生物素、磷酸基或巯基等修饰 DNA 分子点至基片上, 将其连接到修饰过的芯片基片上; 如醛基修饰的基片可以和氨基修饰的 DNA 反应, 羧基修饰的基片在催化剂如 EDC 存在下和氨基修饰得 DNA 反应; 对于微球来说常用的为链霉亲和素包被的磁珠, 将修饰有生物素的 DNA 和微球在小离心管中混合, 放置一段时间后 DNA 会结合到磁珠表面。

被测得 DNA 序列含有一段已知序列用于和测序引物杂交。该序列的引入有多种方法: 例如通过 DNA 连接酶在待测的序列两端加入接头, 接头可作为和引物杂交的序列; 将待测的 DNA 连接到载体中, 然后转化大肠杆菌, 得到单菌落克隆后进行以插入 DNA 序列两端的载体序列作为引物进行 PCR, 从而使载体的一部分已知序列引入待测 DNA 序列。对已知序列信息的待测 DNA 进行再测序, 可直接以已知序列部分作为引物结合区, 待测部分紧接着引物结合区。

被测的 DNA 模板为 PCR 产物, 如 PCR 产物固定至芯片基质后, 通过高温或加

入变性剂如 NaOH, 甲酰胺, 尿素等变性 DNA 双链, 使没有连接到基质的链分离。被测的 DNA 模板也可为滚环扩增产物, 如将 DNA 模板用 DNA 连接酶成环, 然后用 5' 端修饰的引物和模板环杂交, 杂交后通过引物将杂交产物固定在芯片基质上, 然后加入持续合成能力很强的 DNA 合成酶如 phi29 DNA 聚合酶或 Bst DNA 聚合酶及 dNTP 底物进行滚环扩增, 也可以先将引物固定后再和模板形成的环杂交, 然后延伸或将滚环扩增产物直接固定在基质上。被测的 DNA 模板也可以为 DNA 质粒, 酶加工后的 DNA 片断等。

本测序过程涉及到引物和模板的退火、聚合酶延伸反应来检测核苷酸是否掺入。在本发明中待测的 DNA 样品作为引物、聚合酶延伸的模板。任何已知的或适用的 DNA 聚合酶可以被用; 在模板为 RNA 时, 聚合酶为反转录聚合酶。底物核苷酸为任何可以用于聚合酶延伸的核苷酸, 例如脱氧核苷酸 dATP, dCTP, dGTP, dTTP, 双脱氧核苷酸; 双脱氧核苷酸是指 3' -OH 不存在或被修饰, 它可以被聚合酶掺入到引物链中但不能继续延长, 如 ddATP, ddCTP, ddGTP, ddTTP。核苷酸被标记以方便核苷酸掺入的检测; 一种或多种核苷酸被用, 如 Cy3, Cy5, TexasRed, FRET, FITC 等标记的核苷酸。核苷酸的标记及检测方法如一些文献所示, 主要是一些荧光染料, 如荧光素、花青素、罗丹明等, 也可以通过生物素、地高辛等连接纳米金颗粒标记的抗体或亲和素等来标记核苷酸。因为荧光标记和荧光检测是研究最为成熟的方法, 为了方便起见, 荧光为本发明主要标记及检测方法, 但不排除其他标记和检测方法。将引物和固定在基质的模板杂交是将引物溶解在溶液如 2XSSC 或 TE 溶液中, 使终浓度为 1 微摩尔到 100 微摩尔都可。溶有引物的溶液加到已固定有模板的基质上, 在湿盒或微量离心管中孵育 5 分钟至数小时, 孵育温度一般在 37 °C 或其它温度。之后用洗涤液洗涤, 以洗掉未杂交上的引物。延伸反应体系包括: DNA 聚合酶及酶缓冲液如 klenow DNA 聚合酶, Taq DNA 聚合酶等, 标记的核苷酸, 有时还加入 BSA, 单链结合蛋白等。延伸温度在 37 °C-72 °C 都可, 延伸时间可以为半分钟至数分钟。延伸反应可通过将芯片放入洗涤液或加入 50mM EDTA 溶液终止, 洗涤后用芯片扫描仪或其它检测仪器检测。

实施例 2: 对人工合成的 DNA 模板进行测序

以下 DNA 序列由 invitrogen 公司合成:

Seq1: 5' NH₂-(T)₂₅ TT CTC AGG TAG **AGT TGG AGC TCA GGA** CAT GGC AT
 Seq2: 5' NH₂-(T)₂₅ TT ACG TGC GTC **AGT TGG AGC TCA GGA** CTG GGC TG
 Seq3: 5' NH₂-(T)₂₅ TT GTC AGA CGA **AGT TGG AGC TCA GGA** ATA CTT AT
 Seq4: 5' NH₂-(T)₂₅ TT AGT CCT **AGT TGG AGC TCA GGA** TCC TTG GA
 Seq5: 5' -CTGAGCTCCA ACT3'
 Seq6: 5' -TGAGCTCCA ACTN 3'
 Seq7: 5' -GAGCTCCA ACTNN 3'
 Seq8: 5' -GAGCTCCA ACTNNN3'
 Seq9: 5' -AGCTCCA ACTNNNN3'

Seq10: 5' -GCTCCA**ACTNNNNN** 3'

Seq11: 5' -CTCCA**ACTNNNNN** 3'

Seq1-seq4 作为 DNA 模板, seq5-seq11 为测序引物。模板序列中倾斜加粗的部分为引物结合区, 下划线部分为待测序列, NH₂ 表示为氨基修饰。N 为简并性核苷酸, 在化学合成 DNA 模板时该位置同时加入四种合成 DNA 的底物。

Cy5-dATP, Cy5-dCTP, Cy5-dGTP and Cy5-dUTP 购自 Perkin Elmer 公司。Therminator DNA 聚合酶购自 NEB 公司。聚丙烯酸 (Poly(acrylic acid, sodium salt)) 购自 Sigma 公司。

1、基片的制备

(1) 玻片的清洁: 用 1M 的 NaOH 浸泡玻片 1 小时, 双蒸水冲洗干净, 再用重铬酸 (重铬酸钾+浓硫酸) 浸泡 12 小时, 双蒸水多次冲洗去净残酸。

(2) 氨基片的处理: 先用 95%乙醇洗涤玻片, 吹干; 用 3%的硅烷乙醇溶液 (3ml 氨基硅烷加 95%乙醇至终体积 100ml) 浸泡玻片 30 分钟或超声波浸泡 15 分钟; 用 95%乙醇洗净, 再用双蒸水洗净, 吹干, 110°C 烘 30 分钟。

(3) 羧基片的处理: 氨基硅烷化的玻片用 Poly(acrylic acid, sodium salt) 2mg/ml 水溶液 (配制: Poly(acrylic acid, sodium salt) 35wt% solution in water 572ul 加双蒸水至 100ml) 浸泡 20 分钟, 双蒸水洗涤三次, 吹干。

2、模板的固定: 5' 端氨基修饰的模板溶解在 0.1 M Mes 缓冲液 (2- (N-吗啉代) 乙磺酸), pH 5.1, 溶液中含有 5 mM EDC (1-乙基-3 (3-二甲基氨基丙基)-碳二亚胺), 使模板的终浓度为 5 微摩尔。通过点样仪将氨基修饰的模板点至羧基处理的玻片上, 每个模板重复点四次, 四个模板为一方阵, 共点 28 个方阵。在室温下放置于封闭的湿盒中三个小时。然后用去离子水洗涤 5 分钟, 然后用 2 × SSC/0.2% SDS 溶液洗涤 5 分钟, 再用 0.2 × SSC 洗涤 5 分钟, 最后用蒸馏水洗涤 5 分钟后用氮气吹干。

3、聚合酶介导的引物延伸反应: 终浓度为 50 μM 的引物 (溶于 2 × SSC/0.2% SDS) 加到含有测序模板的点阵上, 每 4 个方阵加一种引物, 分别加入 seq5-seq11 和模板退火。退火反应在 42° C 反应 30 分钟。然后用 2 × SSC/0.2% SDS 洗液洗涤后吹干。四个加有相同引物的方阵分别加入含有 0.4 μM Cy5-dATP, Cy5-dCTP, Cy5-dGTP and Cy5-dUTP 延伸反应液。延伸反应液包括: Cy5 修饰的核苷酸底物, 0.5U Therminator DNA 聚合酶, 聚合酶缓冲液。反应在 60° C 反应 2 分钟。用 50mM EDTA 终止反应, 洗涤后吹干。然后用芯片扫描仪扫描。

结果: 以 Seq1-seq4 作为 DNA 模板, 当用 seq5 作为测序引物和它们杂交后, 用 DNA 聚合酶延伸标记的核苷酸, 在 seq1 中可以被聚合酶掺入的碱基为 C; seq2 为 G; seq3 为 T; seq4 为 A; 从扫描后结果如图 6 可以看出, 在 seq1 中 X 轴为 1 的 A、C、G、T 四个荧光强度中 C 的强度最高。在 seq2 中 X 轴为 1 的 A、C、G、T 四个荧光强度中 G 的强度最高。在 seq3 中 X 轴为 1 的 A、C、G、T 四个荧光强

度中 T 的强度最高。在 seq4 中 X 轴为 1 的 A、C、G、T 四个荧光强度中 A 的强度最高。

当 seq6 作为测序引物和 Seq1-seq4 杂交后, X 轴为 2 中荧光强度最强的分别为: seq1: T; seq2: A; seq3: C; seq4: G。

当 seq7 作为测序引物和 Seq1-seq4 杂交后, X 轴为 3 中荧光强度最强的分别为: seq1: A; seq2: C; seq3: G; seq4: G。

当 seq8 作为测序引物和 Seq1-seq4 杂交后, X 轴为 4 中荧光强度最强的分别为: seq1: C; seq2: G; seq3: T; seq4: A。

当 seq9 作为测序引物和 Seq1-seq4 杂交后, X 轴为 5 中荧光强度最强的分别为: seq1: C; seq2: C; seq3: C; seq4: C。

当 seq10 作为测序引物和 Seq1-seq4 杂交后, X 轴为 6 中荧光强度最强的分别为: seq1: T; seq2: A; seq3: T; seq4: T。

当 seq11 作为测序引物和 Seq1-seq4 杂交后, X 轴为 7 中荧光强度最强的分别为: seq1: G; seq2: C; seq3: G; seq4: A。

由结果可知本方法可以准确地测出引物一侧数个碱基的种类。

实施例 3: 对 T7 噬菌体 DNA 文库克隆进行测序

材料: T7 噬菌体 DNA 购自上海生物工程公司; Tsp509 I 购自 NEB 公司; pUC118 载体、感受态细菌及转化试剂盒、T4DNA 连接酶购自大连宝生物公司; klenow DNA 聚合酶购自 fermentas 公司; 以下序列订购自 invitrogen 公司: M13 前引物:

NH₂-CAGGAAACAGCTATGAC (seq12); M13 后引物: GTAAAACGACGGCCAGT (seq13);

测序引物: GCATGCCTGCAGGTCAATT (seq14); CATGCCTGCAGGTCAATTN (seq15);

ATGCCTGCAGGTCAATTNN (seq16); ATGCCTGCAGGTCAATTNNN (seq17);

TGCCTGCAGGTCAATTNNNN (seq18);

实验步骤:

1、将 T7 噬菌体 DNA 用 Tsp509 I 进行酶切, 反应体系为:

T7 噬菌体 DNA 2 μg

Tsp509 I (10U/μl) 0.5 μl

Tsp509 I 反应缓冲液 (10×) 2 μl

H₂O 17.5 μl

65°C, 2 小时;

酶切产物用天为公司的 PCR 纯化试剂盒纯化。

2、用 klenow DNA 聚合酶补齐粘性末端为平端, 反应体系为:

纯化的 T7 噬菌体 DNA 酶切片断 20 μl

klenow DNA 聚合酶 (10U/μl) 0.5 μl

dNTP (10mM) 0.5 μl

37°C, 2 小时; 然后 70°C, 15 分钟失活。

DNA 聚合酶补齐的产物用天为公司的 PCR 纯化试剂盒纯化。

3、将 T7 DNA 平端片断和 pUC118 连接

纯化的 T7 噬菌体 DNA 酶切片断 20 μ l

T4 DNA 连接酶 1 μ l

T4 DNA 连接酶反应缓冲液 (10 \times) 2 μ l

16 $^{\circ}$ C, 2 小时; 然后 70 $^{\circ}$ C, 15 分钟失活。

(4) 将连接的载体转化细菌: 方法按大连宝生物转化试剂盒提供的步骤进行转化。

(5) 将转化的细菌进行平板培养: 将转化的细菌涂布在 LB 固体培养基 (含 50 μ g/ml 氨苄青霉素) 过夜培养。

(6) 挑一单菌落进行菌液 PCR: 挑单个菌落在 LB 液体培养基 (含 50 μ g/ml 氨苄青霉素) 培养 4 个小时, 取 0.5 μ l 加入 PCR 反应液中:

菌液 0.5 μ l

dNTP (10mM) 1 μ l

Taq DNA 聚合酶反应缓冲液 (10 \times) 5 μ l

Taq DNA 聚合酶 (5U/ μ l) 1 μ l

M13 前引物 (10 μ M) 2 μ l

M13 后引物 (10 μ M) 2 μ l

H₂O 38.5 μ l

反应条件: 94 $^{\circ}$ C 30 秒, 54 $^{\circ}$ C 30 秒, 72 $^{\circ}$ C 30 秒, 35 个循环。反应完毕用天为公司的 PCR 纯化试剂盒纯化 PCR 产物。在纯化的最后一步中用去离子水代替洗脱液。用 1%的琼脂糖凝胶进行电泳进行检测。

(7) 模板的固定: 5' 端氨基修饰的模板溶解在 0.1 M Mes 缓冲液 (2- (N-吗啉代) 乙磺酸), pH 5.1, 溶液中含有 5 mM EDC (1-乙基-3 (3-二甲基氨基丙基)-碳二亚胺)。将模板点至羧基处理的玻片上, 每个模板重复点 4 次, 四个模板为一方阵, 共点 28 个方阵。在室温下放置于封闭的湿盒中三个小时。然后用去离子水洗涤 5 分钟, 然后用 2 \times SSC/0.2% SDS 溶液洗涤 5 分钟, 再用 0.2 \times SSC 洗涤 5 分钟, 最后用蒸馏水洗涤 5 分钟后用氮气吹干。将固定有 PCR 产物的基片浸入沸水中 10 分钟, 变性 DNA 双链, 从而单链模板。

(8) **聚合酶介导的引物延伸反应:** 终浓度为 50 μ M 的引物 (溶于 2 \times SSC/0.2% SDS) 加到含有测序模板的点阵上, 每 4 个方阵加一种引物, 分别加入测序模板 seq14-seq18 和模板退火。退火反应在 42 $^{\circ}$ C 反应 30 分钟。然后用 2 \times SSC/0.2% SDS 洗液洗涤后吹干。四个加有相同引物的方阵分别加入含有 0.4 μ M Cy5-dATP, Cy5-dCTP, Cy5-dGTP 和 Cy5-dUTP 延伸反应液。延伸反应液包括: Cy5 修饰的核苷酸底物, 0.5U Therminator DNA 聚合酶, 聚合酶缓冲液。反应在 60 $^{\circ}$ C 反应 2 分钟。用 50mM EDTA 终止反应, 洗涤后吹干。然后用芯片扫描仪扫描。

结果:

1、PCR反应产物用琼脂糖凝胶电泳鉴定结果：

T7噬菌体DNA用Tsp509 I 进行酶切，产生AATT的粘性末端；然后插入pUC118载体中后转化并挑取克隆，然后用M13引物扩增插入载体中的序列。T7DNA插入的部位及插入两侧序列如下：

NH2-CAGGAAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTC ↓ **A**
ATTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN**AATT ↓ GACCTGCAGGCATGC**AAGCTTGGC**ACTGGCCGTCGT**
TTTAC

划线部分为M13引物；箭头为T7DNA在载体中插入部位；划框的部分为测序引物结合区； ↓ AATTNNNNNNNNNNNNNNNNNNNNNNNNNAATT ↓ 为插入序列。

参照图7，1-8道为PCR产物，分子量为330bp，第9道为分子量marker：从下至上依次为100，250，500，750，1000，2000bp。

2、用含简并性引物延伸测序结果

用测序引物seq14-seq18分别和模板杂交后分别用Cy5-dATP, Cy5-dCTP, Cy5-dGTP and Cy5-dUTP延伸，结果见图9。可以从图中读出序列：AACAC；加上已知的AATT酶切位点则序列为：AATTAACAC；模板链则为：GTGTTAATT。

参照图8，PCR产物被固定在基片上，每个点阵中重复点4次。P1-P5分别为测序引物，分别与seq14-seq18相对应。A，C，G，T分别代表所加的标记核苷酸。

3、用sanger法测序验证测序结果。

将PCR产物纯化后送至invitrogen公司用sanger法测序。该PCR产物序列如下：
CCGGGGATCCTCTAGAGT**CAATT**GGACAAAATGCCAGCACTTCCGGCTAAAGGTA**ACTTGAACCTCCGT**
GACATCTTAGAGTCGGACTTCGCGTTCGCGTAACGCCAAATCAATACGACTCACTATAGAGGGACAAAC
TCAAGGTCATTTCGAAGAGTGGCCTTTATGATTGACCTTCTTCCGGTTAATACGACTCACTATAGGAGA
ACCTTAAGGTTTA**ACTTTAAGACCCTTAAGTGTTAATTGACC** (seq19)

可以查出末端部分序列为：GTGTTAATT。与用含简并性引物延伸测序结果一致。

由图10可以读出靠近末端的序列为：GTGTTAATTGACC。

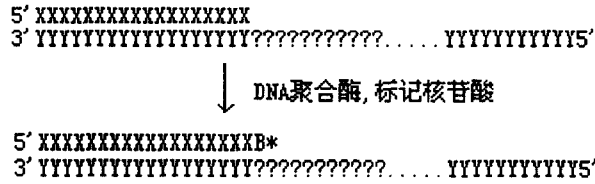


图 1

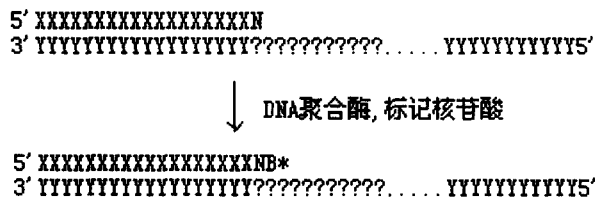


图 2

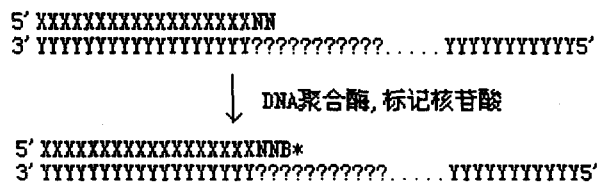


图 3

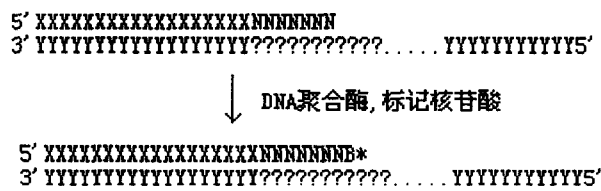


图 4

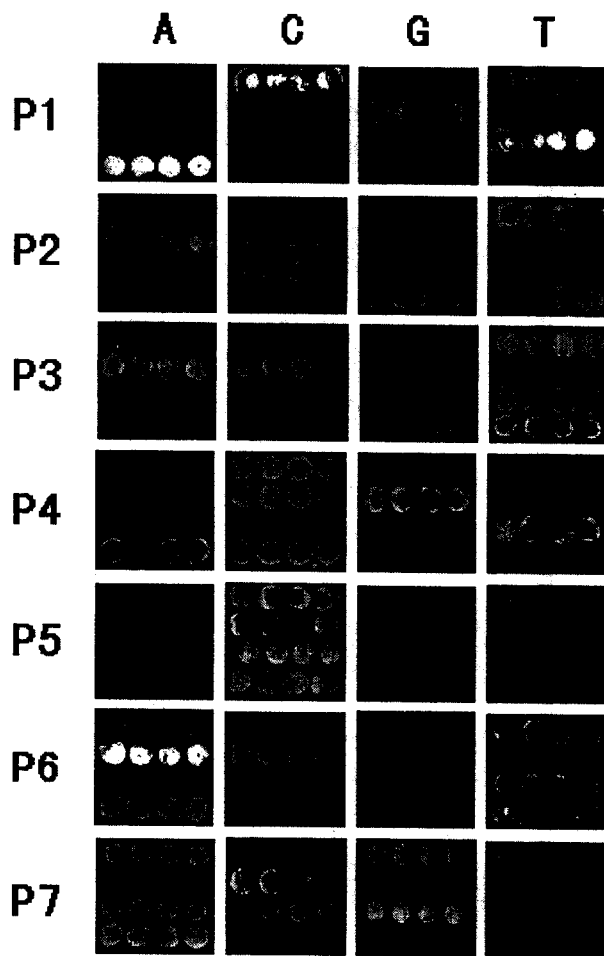


图 5

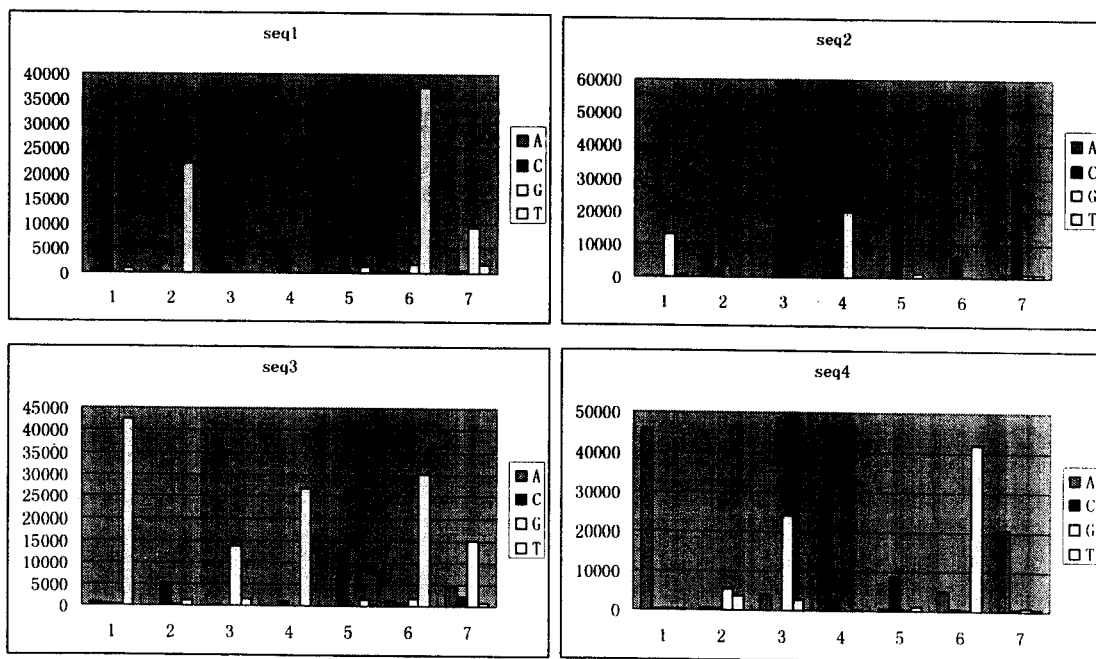


图 6



图 7

	A	C	G	T
P1				
P2				
P3				
P4				
P5				

图 8

220 230 240
TTTAACTTTAAGACCCCTTAAGTGTTAATTGACC

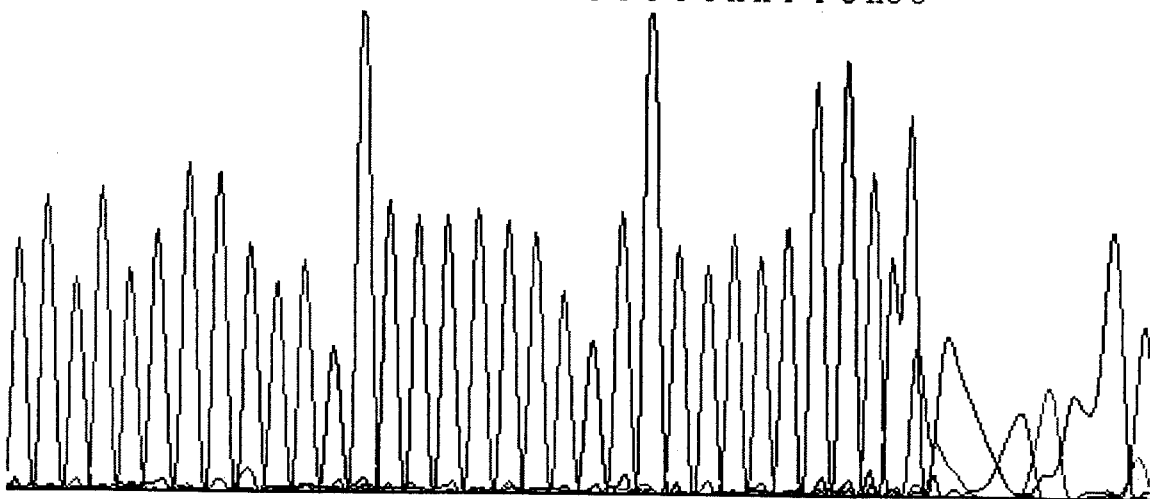


图 9