

# 新一代测序技术的发展和應用

田李<sup>1</sup> 张颖<sup>2</sup> 赵云峰<sup>1</sup>

(1. 曲阜师范大学生命科学学院, 曲阜 273165 ; 2. 山东水利职业学院, 日照 276826)

**摘要:** DNA 测序技术是人类探索生命秘密的重要研究手段。自第一代的 Sanger 测序技术诞生以来, DNA 测序技术经历了三代变革, 产生了第二代到第四代测序技术, 统称为新一代测序技术。目前, 新一代测序技术的数据产出能力呈指数增长, 而且这一技术本身也从依赖 DNA 聚合酶的生化反应转变为面向物理学中纳米技术的新领域。新一代测序对生命科学领域具有里程碑意义, 引领了科学研究模式革新和研究思维的转变。科研人员可利用新一代测序技术对基因组、转录组和表观组等诸多领域展深入的研究。分析了新一代测序技术的特点, 并对其未来的发展方向以及应用进行了展望。

**关键词:** 新一代 DNA 测序 ; 边合成边测序 ; 单分子测序 ; 纳米孔 ; 应用

DOI : 10.13560/j.cnki.biotech.bull.1985.2015.11.003

## The Next Generation Sequencing Technology and Its Applications

Tian Li<sup>1</sup> Zhang Ying<sup>2</sup> Zhao Yunfeng<sup>1</sup>

(1. College of Life Science, Qufu Normal University, Qufu 273165 ; 2. Shandong Water Polytechnic, Rizhao 276826)

**Abstract:** DNA sequencing technology is an important research method in life science. Since the birth of the Sanger sequencing technology, DNA sequencing technology has experienced three generations of change, resulting the second generation to the fourth generation sequencing technology. DNA sequencing not only has been improving its productivity in an exponential growth rate but also been evolving into a new technological territories toward physical disciplines of nanotechnology. Next generation sequencing that has landmark significance of life sciences, can improve researches in the genome, transcriptome and epigenome. This review analyzes technical characteristics of the next generation sequencers and provide prospective insights into their future development and applications.

**Key words:** next generation sequencing; sequence by synthesis ; single molecule sequencing ; nanopore ; applications

DNA 测序技术在生命科学的发展中起着越来越重要的作用。新一代测序技术是一种革命性的技术, 它的出现使得科研人员能够以相对较少的经费获得以往望尘莫及的海量 DNA 序列, 从根本上改变了人们研究生命科学的方式<sup>[1]</sup>。现阶段, 生命科学的研究已经从以往研究单一基因转变为研究整个基因组, 其中既包括了基础研究中的基因组、转录组和表观遗传, 也涉及了应用研究中的医学诊断和农作物育种等<sup>[2]</sup>。本文回顾了 DNA 测序技术的演化, 并论述了其在生命科学研究中的应用。

### 1 测序技术的发展

#### 1.1 第一代测序技术

Sanger 等在 20 世纪 70 年代中期发明了 DNA 末端终止法测序技术, 他的发明第一次为人们开启了解读生命遗传密码的大门, Sanger 本人也因此获得了 1980 年诺贝尔化学奖<sup>[3]</sup>。DNA 末端终止法测序技术的基本原理是: 通过在 DNA 聚合酶、模板、放射性同位素标记的引物、dNTP 和 ddNTP 的作用下发生延伸反应, 由于 ddNTP 的存在, 会形成长度不一的 DNA 延伸片段; 然后采用平板凝胶电泳, 用 4

收稿日期: 2015-04-10

基金项目: 国家自然科学基金项目(31501588), 中国博士后科学基金项目(2014T70621), 山东省自然科学基金项目(ZR2013CQ018)

作者简介: 田李, 博士, 副教授, 研究方向: 植物病原真菌的致病机理及比较基因组学; E-mail: tianlister@163.com

通讯作者: 赵云峰, 博士, 教授, 研究方向: 细胞生物学; E-mail: yfz667788@163.com

条电泳道来分离 4 个反应的所得产物,便可以按顺序读出相应的 DNA 序列。在那个年代,测序主要依靠手工操作,难以自动化,并且依赖电泳技术,试剂消耗也大,这些都极大限制了测序的通量。

其后在此技术原理的基础上产生了几次变革,主要技术上的变化有以下三点:(1)采用具有颜色的荧光染料取代了放射性同位素标记;(2)采用毛细管电泳技术取代了平板凝胶电泳技术;(3)并行化程度更高。这其中应用最广泛的是 ABI 公司的 3730 测序仪,它可以在一次运行中分析 96 个样本,读长最多可以超过 1 000 bp。这一代测序技术在人类基因组计划的后期阶段起到了关键的作用,加速了人类基因组计划的完成<sup>[4]</sup>。但是,由于其对电泳分离技术的依赖,使其难以进一步提高分析的速度和通过微型化降低测序成本,因此在 2005 年后,除了在 PCR 产物测序和病毒的基因组测序中继续发挥重要作用,其他均已较少采用。但由于其在原始数据质量(准确率达 99.999%)以及序列读长方面具有的优势,它还将与新的测序平台并存。

## 1.2 第二代测序技术

高通量测序技术进入市场,使 DNA 测序技术在 2005 年发生了重要转折,改变了测序的规模化进程。Illumina、Roche 和 ABI 公司都推出了各自的新一代 DNA 测序仪,主要技术革新有以下几点:(1)采用矩阵分析技术,实现了大规模并行化,使得矩阵上的 DNA 样本可以被同时并行分析;(2)不再采用电泳技术,使得 DNA 测序仪得以微型化,测序成本大大降低;(3)边合成边测序,测序速度大幅提高。与 Sanger 测序相比,第二代测序技术单次运行产出序列数据量大,所以又被通称为高通量测序技术。其技术原理是:首先构建 DNA 模板文库,将 DNA 固定在芯片表面或微球表面;然后通过扩增形成 DNA 簇或扩增微球;最后利用聚合酶或者连接酶进行一系列循环的反应操作,通过 CCD 相机采集每个循环反应中产生的光学事件信息,从而获得 DNA 片段的序列。

1.2.1 Illumina Genome Analyzer Illumina 公司于 2007 年以 6 亿美元收购基因测序公司 Solexa,推出了成熟商业产品 Genome Analyzer<sup>[5]</sup>。该技术利用单

链 DNA 两端的非对称接头将 DNA 片段固定在芯片表面形成寡核苷酸桥,并将该芯片放置于流通池内,完成 DNA 模板文库构建步骤。经过多个 PCR 循环扩增出大量的复制产物,每一簇复制产物都分别固定在芯片表面的特定位置上。然后,测序引物杂交到扩增产物中的接头上,开始合成测序反应。在每一轮的测序循环中, DNA 聚合酶和标记不同荧光基团的 4 种核苷酸被同时加入到流通池中,按照碱基互补配对的原则延伸一个核苷酸。此时采集荧光基团所发出的荧光图像,就可以获得模板中这一位置的 DNA 序列信息。为防止额外的延伸,每个核苷酸的 3' 羟基是被封闭起来的,然后打开 3' 端,继续进行下一轮反应并重复多次,以获得约 50 个碱基的 DNA 序列。

1.2.2 Roche 454 Genome Sequencer 该技术将固化引物的微球与单链 DNA 相结合,构建 DNA 模板文库<sup>[6]</sup>。调整微球与文库片段的比例,以保证大多数微球只能结合 1 个单链 DNA 分子。油与水溶液混合形成油包水结构乳滴,利用微乳滴 PCR 来生成扩增产物。经过多轮循环,每个微球表面都结合了大量相同的 DNA 片段。富集微球并转移到带有规则微孔阵列的微孔板上,每个微孔只能容纳 1 个微球。微孔板的其中一面可以进行测序反应,另一面则与 CCD 光学检测系统相接触。

序列测定同样采用边合成边测序<sup>[7]</sup>。三磷酸核苷结合到 DNA 链上会释放出焦磷酸,此时通过荧光素酶和 ATP 硫酰化酶产生级联反应会释放出光信号。454 利用该光学信号来进行检测。具体方法是顺次向微孔板中加入 4 种 dNTP 中的一种,监测每个微孔之中是否释放出光信号,表明该 dNTP 是否连接到 DNA 片段上,以此明确 DNA 模板上的互补碱基。

1.2.3 Life Technologies SOLiD System 与 454 类似, SOLiD 也采用微乳滴 PCR 的方法扩增 DNA 模板<sup>[8]</sup>,并将扩增微球固定在玻璃基板上形成高通量的阵列。SOLiD 采用连接反应进行边合成边测序。将通用引物与连在微球上的 DNA 文库模板杂交,然后进行一系列的连接反应。每个连接反应都发生在 DNA 延伸链和带有荧光标记的单链八核苷酸探针池中的某一探针之间。八核苷酸探针的碱基与特定的荧光颜色有明确的对应关系。经过一系列复杂的连接,酶切

和下一引物结合的反应循环后,获取荧光图像,即可根据碱基与荧光之间的对应关系读出 DNA 序列信息。

第二代技术是目前市场上主流的 DNA 测序技术,已经广泛地应用于各项研究领域。较第一代测序技术而言,测量通量明显提高。第二代测序技术极大地推进了基因组相关研究的进展,以前让研究者望尘莫及的基因组测序工作,现在几乎每一个实验室都可以开展。但是其不足之处也日益凸显。首先,第二代测序读长较短<sup>[9]</sup>。这一缺点对后续的序列拼接,组装以及注释等生物信息学分析带来了很大困难。SOLiD 测序仪和 Illumina 公司的测序仪读取的单一序列长度一般介于 75-100 bp, Roche 公司的 454 测序仪可以达到 700 bp 的读长,相应的其通量仅仅为 0.7 Gb,因此其成本偏高。其次,第二代测序技术原理是建立在 PCR 的基础上,但是扩增后得到的 DNA 分子片段的数目和扩增前 DNA 分子片段的数目比例有相对偏差,在分析基因表达方面存在较大的弊端<sup>[10]</sup>。因此序列读长较短和需要模板扩增步骤,成为第二代测序技术最集中的弊端所在。这样就需要开发出不经扩增的单分子测序、读长超过以往的新型测序技术,第三代测序技术便应运而生。

### 1.3 第三代测序技术

第三代测序技术的技术标志就是单分子测序和长读长。这得益于物理、化学、材料等学科的发展及其与生命科学的融合。第三代测序技术通过在单一 DNA 分子组成的阵列上进行合成测序。在一个表面积限定的介质上使用单个分子,可以增加独立分析的 DNA 片段的数量,也意味着不再进行昂贵的 DNA 扩增步骤了,因此,可以使数据产出量更高,并且将进一步降低测序的成本。但同时该技术也带来了一些新的挑战,主要集中在单分子水平光学信号的检测方面。主要的问题是要降低没有参与到实际化学反应中的游离荧光分子的背景干扰。解决原则主要是将检测局限在测序反应发生的实际位置附近。下面以在商业化中应用较好的 Pacific Biosciences 公司的单分子实时测序仪 SMRT 加以阐述<sup>[11, 12]</sup>。

SMRT 单分子实时测序仪以 SMRT 芯片为载体进行测序反应。SMRT 芯片是一种带有很多零模式波导孔的金属片,在该纳米孔的底部区域锚定有 DNA 聚合酶。由于零模式波导孔直径只有几十个纳米,其直径低于光的波长,所以光线无法透射。这样就创造了一个体积很小的检测空间。测序时,被打断成许多小的片段的基因组 DNA 分散到不同的零模式波导纳米孔中。当孔底部聚合反应发生时,不同荧光标记的 dNTP 会在小孔的荧光探测区域中被 DNA 聚合酶滞留数十毫秒,在这期间,荧光标记会在激光束的激发下发出相应的荧光,根据荧光的种类就可以判定该位置核苷酸的种类。

目前, SMRT 技术的平均读长已经提升至 3 000 bp 左右,在这一点上远远优于二代测序技术,所以在序列拼接和需要跨越重复区域的 DNA 组装中有着极大优势。另外,读长的增加也使需要测序覆盖深度随之下降,进一步降低了测序的成本。但是因为单分子测序,测序中产生的任何一个错误都会被真实地记录下来,这就造成了 SMRT 测序仪最致命的问题。具体来说,测序错误可能是会出现碱基的插入和缺失错误:碱基缺失错误是由于在某些时刻碱基掺入 DNA 链的速度过快,超过了相机最大的拍摄帧数;插入错误是由于在某些时刻 DNA 聚合酶随机的选择一些 dNTP,但并未真正将这些 dNTP 掺入 DNA 链中。这些测序错误导致 SMRT 测序仪的准确性仅有 85%,相比第二代测序技术至少 99.5% 的测序准确率,确实是很大的短板。但这些错误是随机的,并不会随着读长的增加而增加。未来随着测序试剂的优化以及每个纳米孔可获得的数据量的增加,测序错误会随着测序覆盖深度的增加逐渐被降低,相信单分子测序技术可以在不断的发展过程中克服其劣势<sup>[13]</sup>。

### 1.4 第四代测序技术

在上述第二代测序和第三代测序技术中, DNA 序列都是在荧光等发光物质的协助下,通过 DNA 聚合酶将不同的 dNTP 连接到 DNA 链上,读取此过程中释放出的不同光学信号而间接确定的。这些方法都需要昂贵的光学监测系统,并依赖 DNA 聚合酶读取碱基序列,这些项目都增加了测序的成本。因此

开发出使用生物化学试剂,直接读取 DNA 序列信息的新型测序方法是非常可取的,由此构成了第四代测序技术的主要思想。

第四代测序技术中的代表当属纳米孔测序,它不需要对 DNA 样品进行任何生物或化学方面的处理,而采用物理方法直接读出其碱基序列<sup>[14, 15]</sup>。其基本原理可概括为:单个碱基通过纳米孔通道时,就会引起通道电学性质的变化,并且由于 ATGC 这 4 种不同的碱基存在电学性质差异,使得它们穿越纳米孔时所引起的电学参数的变化量也不同。因此,不同的电学参数变化量就对应通过纳米孔的相应碱基。由此可见,第四代测序技术特点是完全抛弃了在复杂的 DNA 聚合酶的生化反应中进行 DNA 序列的读取,而是利用不同碱基的电学性质差异,通过纳米孔等直接对碱基穿过电极时的电流变化进行测量实现的。从目前的情况来看,研究人员已经在纳米孔的制造和 DNA 分子的控制上取得了一定的进展,但是目前第四代测序技术所取得的成果还都处在实验室阶段并且存在着其局限性,但是最近的研究工作表明未来新一代的 DNA 测序平台可能将在其中产生<sup>[16]</sup>。

### 1.5 测序技术的发展趋势

回顾上述四代测序平台的技术的发展,可以看出,生物化学技术和固态技术的融合推动了 DNA 测序技术的进步。现阶段,技术的融合有从生物化学手段向物理手段发展的趋势<sup>[17]</sup>,相信这一趋势将继续持续下去。下一代 DNA 测序技术将可能不再使用生物化学的方法,而物理手段纳米技术将有可能发挥更大的作用。未来基于纳米孔的 DNA 测序技术,当线性 DNA 通过纳米孔时,核苷酸序列就会被确定下来。这样可以同时实现长读长和高通量。理论上一个纳米孔结构单次测序读长可能仅仅受到线状 DNA 链的长度限制;而表面积很小的芯片上也可以容纳不计其数的纳米孔。因此,预计新一代的测序技术在具有超高通量的同时,其读长也将轻易超过以长读长闻名的第一代测序技术。

## 2 新一代测序技术的应用

### 2.1 基因组从头测序

基因组从头测序是在没有任何现有的 DNA 序列

资料的情况下,直接对某个物种的基因组进行测序。第一代测序技术在 1990 年启动的人类基因组计划和多种模式生物,如拟南芥 (*Arabidopsis thaliana*)<sup>[18]</sup>、线虫 (*Caenorhabditis elegans*)<sup>[19]</sup> 和小鼠 (*Mus musculus*)<sup>[20]</sup> 全基因组测序中起了重要的作用。但是,测序速度慢、成本高和通量低的第一代测序技术远远不能满足人们对大量生物基因组解析的需求,因此第二代测序技术出现后,人们开始选择使用新一代测序技术进行全基因组从头测序。熊猫 (*Ailuropoda melanoleura*) 基因组<sup>[21]</sup> 的从头测序是第一次完全采用第二代测序技术完成的大型物种的全基因组从头测序,标志着第二代测序技术和拼接组装技术登上了基因组从头测序的历史舞台。2011 年以来,第二代测序技术快速发展。伴随着测序所需的成本的降低和测序时间的缩短,采用第二代测序技术从头测序的全基因组犹如雨后春笋般出现,基因组学研究也迎来了革命性突破。不过第二代测序技术测序读长短,这就要求必须有足够的覆盖度才能完成基因组序列的拼接。第三代测序技术具有读长长的特点,在基因组测序中能降低测序后 contig 的数量,大大减少了后续的基因组组装的工作量,节省大量的测序成本和时间。科学家仅仅用 0.5× 的第三代测序平台的测序数据结合 38× 的第二代测序的数据,就完成了马达加斯加指猴基因组序列的组装<sup>[22]</sup>。现阶段,三代测序技术均有其优势与局限性。因此从根本上说,要完成特定物种的基因组从头测序,必须进行合理评估以选择合适的测序平台。

### 2.2 基因组重测序

基因组重测序是针对已知基因组序列的物种而言,重新测序的对象是该物种具有不同性状的其他个体。通过基因组重测序并进行差异信息分析,人们能够快速地进行很多有意义的研究,具有重大的科研价值和产业价值。具体来说主要有以下几点:(1) 在群体水平研究物种的进化历史和对环境的适应性。对种内具有不同表型的个体进行基因组重测序,可以在全基因组水平上找到群体内个体间的 DNA 差异,包括大量的 SNPs 和结构变异 (structure variations, SVs) 等变异信息,而这些差异可能与这些个体的表型差异存在关联性,从而明确基因组是

如何进化以使物种适应不同环境等问题。Lam 等<sup>[23]</sup>对 14 株栽培大豆和 17 株野生大豆进行了全基因组重测序,通过比较分析,鉴定出了栽培大豆中获得以及丢失的 18 万多个遗传变异位点,且栽培大豆相对于野生大豆有着更低水平的遗传多样性,这可能与人类的选择有关。(2) 基因组重测序可以在全基因组水平扫描出与动植物重要性状相关的变异位点,是育种研究中迅速有效的新方法。Zheng 等<sup>[24]</sup>采用基因组重测序技术,对 950 份世界范围内的水稻栽培种进行了遗传分析,鉴定出 18 个与粒重和开花期相关联的候选基因,为水稻的进一步遗传育种提供了理论基础。(3) 遗传突变、适应进化和表型筛选是创造出带有优良性状突变体的有力工具,基因组重测序技术有利于突变位点的定位和鉴定。Ashelford 等<sup>[25]</sup>对一个拟南芥突变体的回交系进行基因组重测序,成功鉴定出在 AtNFXL-2 基因中引起该突变表型的 SNP 位点。

### 2.3 转录组测序

转录组测序(RNA-seq)是从总 RNA 中富集出单链 mRNA 经反转录得到双链 cDNA,而后对其进行高通量测序分析。第二代测序技术发展后, RNA-seq 在新基因发现、可变剪切位点识别、基因表达和小 RNA 测序及其靶标 mRNA 的识别上都有重要应用。而第三代测序技术拥有实时测序的特点,可以直接对 RNA 进行测序,免除了将 RNA 转变成 DNA 的过程,更加促进了 RNA-seq 的发展。下面将逐一作出阐述。

2.3.1 mRNA 测序 Chen 等<sup>[26]</sup>采用 RNA-seq 对飞蝗(*Brugia malayi*)的转录组进行了测序,对获得的 21.5 Gb 的序列进行了拼接,共得到 7 万多转录本,由此鉴定出 11 490 个蝗虫蛋白的编码基因,从基因组范围内全面解析了飞蝗的核心基因集。Li 等<sup>[27]</sup>使用 RNA-seq 分析了玉米叶片的转录组,得到约 120 Mb 的转录组数据,结合玉米基因组序列,预测了基因的结构和可变剪接事件。结果显示,大部分玉米基因存在不同的 mRNA 可变剪接事件,这表明可变剪接事件比预期的更常见。这些数据为研究远比预期复杂的玉米转录调节机制提供了广泛的依据。

2.3.2 基因表达分析 随着测序技术的进步,科学

家们越来越多的采用数字基因表达谱(digital gene expression, DGE)技术进行基因差异表达分析。该技术的基本原理是将 mRNAs 反转录成 cDNAs,然后将 cDNAs 进行双酶切,使得一条 mRNA 对应一个相应的短 DNA 标签,而后采用高通量测序和分析流程,经过生物信息分析比较不同样本间各种标签条数,找出差异的表达标签,从而明确差异基因表达。

Wang 等<sup>[28]</sup>利用数字基因表达谱技术分析野生型棉花和它的突变体基因表达情况发现,在野生型和突变体之间,磷酸酶基因、纤维素合成酶基因和脱氢酶基因表达差异水平最大,而上述基因都参与了棉纤维细胞的发育过程,从而证实了在纤维早期发育中基因转录调控的高度复杂性。Hao 等<sup>[29]</sup>首先对红豆杉通过 RNA-seq 技术对其转录组进行了从头测序组装,并基于生物信息学分析和同源蛋白的搜索,鉴定出 2 万多个红豆杉单一基因序列;然后使用数字基因表达谱技术分析了根、茎和叶 3 种组织中基因差异表达情况,鉴定出一批红豆杉组织特异性基因和紫杉烷生物合成途径的重要基因。

2.3.3 小 RNA 测序及其靶标 mRNA 的识别 Guo 等<sup>[30]</sup>采用高通量 RNA-seq 测序,分析了常规条件下和 H<sub>2</sub>O<sub>2</sub> 胁迫处理条件下的水稻幼苗的 miRNAs 组。通过生物信息学分析发现,有 7 个 miRNAs 家族在 H<sub>2</sub>O<sub>2</sub> 胁迫处理条件下呈现出明显的差异表达。这些 miRNAs 的靶基因参与了包括养分运输、转录调控、细胞增殖和细胞程序化凋亡不同的代谢过程和细胞周期调控,说明多样化的 miRNAs 形成了一个复杂的植物氧化应激反应的调控网络。除此之外,在水稻中还发现了 32 个尚未鉴定出的 miRNAs,并且首次发现了一个前体位于植物外显子小 RNA,说明植物也可以使用某些外显子作为 miRNA 的来源。

明确了全基因组范围内的 miRNAs 组后,鉴定 miRNAs 的靶标 mRNAs 可以对其生物学功能展开详细的研究。随着测序技术的发展,现在可以采用 RNA-seq 技术用于 miRNAs 的靶标 mRNAs 配对关系的发现,这一方法被称为降解组测序。其基本原理是:在植物体内大多数的 miRNAs 剪切 mRNA 的位点是两者互补区域的第 10 位核苷酸,经剪切后靶 mRNA 产生了 3' 剪切片段和 5' 剪切片段;其中 3' 剪切片段含有 5' 单磷酸基团,可用于下游高通量测

序;而含有帽子结构的 5' 剪切片段和含有 5' 帽子结构的尚未降解的 mRNA 缺少 5' 单磷酸基团,无法进入下游的高通量测序。因此,对 3' 端降解片段进行高通量测序并进行比较分析后,可以直观地发现在某个 mRNA 的某个位点上会出现一个波峰,该 mRNA 便是 miRNAs 的靶标 mRNA,波峰位置便是候选的 miRNA 剪切位点。Zhou 等<sup>[31]</sup>利用降解组测序在水稻中鉴定 miRNAs 靶基因,发现了 87 个 miRNAs 的 177 个靶标 mRNAs。这些靶标 mRNAs 在水稻的基因表达调控中发挥重要作用,构成了复杂的调节网络。

**2.3.4 第三代测序技术与 RNA 测序** 利用第三代测序平台,可以免除将 RNA 转变成 cDNA 的步骤,实现 RNA 的直接测序<sup>[32]</sup>。这是因为第三代测序平台为单分子测序平台,将 DNA 聚合酶换为反转录酶便可对 RNA 直接进行测序,利用该技术已成功对酿酒酵母的 RNA 进行了直接测序<sup>[33]</sup>。

## 2.4 新一代测序技术对表观遗传学的贡献

表观遗传学是研究在非基因序列改变前提下,DNA 甲基化和组蛋白修饰等所导致的基因表达水平变化。而随着测序技术的发展,产生了表观基因组学,它是在基因组水平上对表观遗传学改变的研究。DNA 甲基化修饰、组蛋白修饰是表观基因组学的重要研究内容。

**2.4.1 DNA 甲基化修饰** 亚硫酸氢盐可以使 DNA 中没有发生甲基化的胞嘧啶转变为尿嘧啶,甲基化的胞嘧啶则可以保持不变。利用上述原理,对亚硫酸氢盐处理过的基因组测序并且与未经处理的序列相比较,就可以得到全基因组范围内单碱基分辨率水平的甲基化图谱,这就是全基因组甲基化测序技术。

Xiang 等<sup>[34]</sup>利用全基因组甲基化测序技术,对家蚕的 2 个个体进行了测序,得到了家蚕丝腺的甲基化图谱,共找出 17 万个甲基化位点,其中绝大部分位于 GC 岛,0.11% 的胞嘧啶发生了甲基化修饰。在这些甲基化位点中,基因内部的甲基化占了很大一部分;而在基因启动区域、rDNA 区域和转座元件区域甲基化程度很低。说明在高等生物中发挥重要调控作用的启动子区甲基化、核糖体 rDNA 甲基化

和转座子区的甲基化未在昆虫中进化出来,家蚕甲基化谱的成功绘制为解析昆虫类的表观遗传调控提供了重要资料。

第三代测序技术对 DNA 聚合酶的工作状态进行了实时监测,聚合酶每合成一个碱基都要消耗一个时间段,而当 DNA 模板的碱基带有甲基化等修饰时,聚合酶的速度就会慢下来。通过这一原理就可以判断 DNA 模板的这个位置是否存在甲基化修饰,为表观遗传学研究开辟了一条新路<sup>[35]</sup>。

**2.4.2 组蛋白修饰 染色质免疫共沉淀 (Chromatin immunoprecipitation, ChIP)** 是研究体内 DNA 与蛋白相互作用的一种方法,开始往往用在解析转录因子在基因组范围内的结合位点上。近年来,将该技术与新一代测序技术相结合后产生了染色质免疫共沉淀-测序 (ChIP-seq) 技术,在表观遗传学中发挥了重要作用。先通过 ChIP 富集与特定组蛋白修饰相结合的 DNA 片段,然后进入高通量测序流程,最后将获得的所有 DNA 序列标签定位到基因组上,从而获得不同修饰的组蛋白在全基因组范围内的 DNA 结合区段信息。Wang 等<sup>[36]</sup>采用 ChIP-seq 技术,对玉米幼苗的 4 种组蛋白修饰 (H3K4me3、H3K27me3、H3K36me3 和 H3K9ac) 进行了详尽的研究,表明其中 3 种组蛋白修饰 (H3K4me3、H3K9ac 和 H3K36me3) 正调控基因表达;而组蛋白修饰 H3K27me3 负调控基因表达。

## 3 小结

DNA 测序技术的发展已经成为生物学领域最前沿的领域之一。从测序技术上来看,已经商业化的前三代测序技术由于之间功能上的互补性,它们将长期共存;而第四代测序技术指明了未来测序技术的发展方向。从应用方面来看,快速而廉价的 DNA 测序能力将使基因组学成为研究生物学问题的常规方法,引领我们开辟一系列新的研究领域。

## 参考文献

- [1] 周晓光,任鲁风,李运涛,等. 下一代测序技术:技术回顾与展望[J]. 中国科学:生命科学,2010,40(1):23-37.
- [2] 岳桂东,高强,罗龙海,等. 高通量测序技术在动植物研究领域中的应用[J]. 中国科学:生命科学,2012,42(2):107-124.

- [ 3 ] Sanger F. Sequences, sequences, and sequences [ J ]. Annual Review of Biochemistry, 1988, 57 ( 1 ) : 1-29.
- [ 4 ] McPherson JD, Marra M, Hillier LD, et al. A physical map of the human genome [ J ]. Nature, 2001, 409 ( 6822 ) : 934-941.
- [ 5 ] Fedurco M, Romieu A, Williams S, et al. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies [ J ]. Nucleic Acids Research, 2006, 34( 3 ) : e22.
- [ 6 ] Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors [ J ]. Nature, 2005, 437 ( 7057 ) : 376-380.
- [ 7 ] Ronaghi M, Karamohamed S, Pettersson B, et al. Real-time DNA sequencing using detection of pyrophosphate release [ J ]. Analytical Biochemistry, 1996, 242 ( 1 ) : 84-89.
- [ 8 ] Holt RA, Jones SJM. The new paradigm of flow cell sequencing [ J ]. Genome Research, 2008, 18 ( 6 ) : 839-846.
- [ 9 ] Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology [ J ]. Trends in Genetics, 2008, 24 ( 3 ) : 142-149.
- [ 10 ] Torres TT, Metta M, Ottenwälder B, et al. Gene expression profiling by massively parallel sequencing [ J ]. Genome Research, 2008, 18 ( 1 ) : 172-177.
- [ 11 ] Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules [ J ]. Science, 2009, 323 ( 5910 ) : 133-138.
- [ 12 ] Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing [ J ]. Human Molecular Genetics, 2010, 19 ( R2 ) : R227-R240.
- [ 13 ] 张得芳, 马秋月, 尹佟明, 夏涛. 第三代测序技术及其应用 [ J ]. 中国生物工程杂志, 2013, 33 ( 5 ) : 125-131
- [ 14 ] Yang J, Ferranti DC, Stern LA, et al. Rapid and precise scanning helium ion microscope milling of solid-state nanopores for biomolecule detection [ J ]. Nanotechnology, 2011, 22 ( 28 ) : 285310.
- [ 15 ] Marshall MM, Yang J, Hall AR. Direct and transmission milling of suspended silicon nitride membranes with a focused helium ion beam [ J ]. Scanning, 2012, 34 ( 2 ) : 101-106.
- [ 16 ] 陈文辉, 罗军, 赵超. 固态纳米孔: 下一代 DNA 测序技术——原理、工艺与挑战 [ J ]. 中国科学: 生命科学, 2014, 44 ( 7 ) : 649-662.
- [ 17 ] 任鲁风, 于军. 解读生命密码的基本手段——DNA 测序技术的前世今生 [ J ]. 生命科学, 2012, 24 ( 12 ) : 1357-1362.
- [ 18 ] Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* [ J ]. Nature, 2000, 408 ( 6814 ) : 796.
- [ 19 ] Hillier LDW, Marth GT, Quinlan AR, et al. Whole-genome sequencing and variant discovery in *C. elegans* [ J ]. Nature Methods, 2008, 5 ( 2 ) : 183-188.
- [ 20 ] Powell K, Abbott A, Check E. Mouse genome : The real deal [ J ]. Nature, 2002, 420 ( 6915 ) : 456-456.
- [ 21 ] Li R, Fan W, Tian G, et al. The sequence and de novo assembly of the giant panda genome [ J ]. Nature, 2010, 463 ( 7279 ) : 311-317.
- [ 22 ] Perry GH, Reeves D, Melsted P, et al. A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar [ J ]. Genome Biology and Evolution, 2012, 4 ( 2 ) : 126-135.
- [ 23 ] Lam HM, Xu X, Liu X, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection [ J ]. Nature Genetics, 2010, 42 ( 12 ) : 1053-1059.
- [ 24 ] Zheng LY, Guo XS, He B, et al. Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*) [ J ]. Genome Biology, 2011, 12 ( 11 ) : R114.
- [ 25 ] Ashelford K, Eriksson ME, Allen CM, et al. Full genome re-sequencing reveals a novel circadian clock mutation in *Arabidopsis* [ J ]. Genome Biol, 2011, 12 ( 3 ) : R28.
- [ 26 ] Chen MX, Ai L, Xu MJ, et al. Identification and characterization of microRNAs in *Trichinella spiralis* by comparison with *Brugia malayi* and *Caenorhabditis elegans* [ J ]. Parasitology Research, 2011, 109 ( 3 ) : 553-558.
- [ 27 ] Li P, Ponnala L, Gandotra N, et al. The developmental dynamics of the maize leaf transcriptome [ J ]. Nature Genetics, 2010, 42( 12 ) : 1060-1067.
- [ 28 ] Wang QQ, Liu F, Chen XS, et al. Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant [ J ]. Genomics, 2010, 96 ( 6 ) : 369-376.
- [ 29 ] Hao DC, Ge G, Xiao P, et al. The first insight into the tissue specific *taxus* transcriptome via Illumina second generation sequencing [ J ]. PLoS One, 2011, 6 ( 6 ) : e21220.
- [ 30 ] Guo W, Wu G, Yan F, et al. Identification of novel *Oryza sativa*

- miRNAs in deep sequencing-based small RNA libraries of rice infected with Rice stripe virus [ J ] . PLoS One, 2012, 7 ( 10 ) : e46443.
- [ 31 ] Zhou M, Gu L, Li P, et al. Degradome sequencing reveals endogenous small RNA targets in rice ( *Oryza sativa* L. ssp. indica ) [ J ] . Frontiers in Biology, 2010, 5 ( 1 ) : 67-90.
- [ 32 ] Uemura S, Aitken CE, Korfach J, et al. Real-time tRNA transit on single translating ribosomes at codon resolution [ J ] . Nature, 2010, 464 ( 7291 ) : 1012-1017.
- [ 33 ] Ozsolak F, Platt AR, Jones DR, et al. Direct RNA sequencing [ J ] . Nature, 2009, 461 ( 7265 ) : 814-818.
- [ 34 ] Xiang H, Zhu J, Chen Q, et al. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map [ J ] . Nature Biotechnology, 2010, 28 ( 5 ) : 516-520.
- [ 35 ] Song CX, Clark TA, Lu XY, et al. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine [ J ] . Nature Methods, 2012, 9 ( 1 ) : 75-77.
- [ 36 ] Wang X, Elling AA, Li X, et al. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize [ J ] . The Plant Cell, 2009, 21 ( 4 ) : 1053-1069.

( 责任编辑 李楠 )