

DNA测序技术的发展历史与最新进展

解增言 林俊华 谭军 舒坤贤

(重庆邮电大学生物信息学院, 重庆 400065)

摘要: DNA测序技术是现代分子生物学研究中最常用的技术。从1977年第一代测序技术的出现, 经过30多年的发展, DNA测序技术取得重大进展, 以高通量为特点的第二代测序技术逐步成熟并商业化, 以单分子测序为特点的第三代测序技术也已经出现。介绍每一代测序技术的特点, 并重点介绍了第二代测序技术及其应用。展望新的测序技术对于未来生物学研究及人们自身健康与人类疾病等方面研究的影响。

关键词: 第一代 DNA测序技术 第二代 DNA测序技术 第三代 DNA测序技术 高通量 单分子测序

The History and Advances of DNA Sequencing Technology

Xie Zengyan Lin Junhua Tan Jun Shu Kunxian

(College of Bioinformatics, Chongqing University of Posts and Telecommunications, Chongqing 400065)

Abstract As the commonly used technology in molecular biology, DNA sequencing technology has seen great advances in more than thirty years since 1977. Next generation DNA sequencing technology characterized by high-throughput has entered the market and been growing maturely. The third generation sequencing technology has also arisen, which can sequence single DNA molecules. In this review, we introduced three generations of sequencing technologies and emphasized on the second generation sequencing technology and its applications. At the end of the review, the effect of new DNA sequencing technologies on research and medicine were pre-viewed.

Key words: First generation DNA sequencing technology Second generation DNA sequencing technology Third generation DNA sequencing technology High-throughput Single molecule sequencing technology

DNA测序技术是分子生物学研究中最常用的技术, 它的出现极大地推动了生物学的发展。成熟的DNA测序技术始于20世纪70年代中期。1977年Maxam和Gilbert^[1]报道了通过化学降解测定DNA序列的方法。同一时期, Sanger^[2]发明了双脱氧链终止法。20世纪90年代初出现的荧光自动测序技术将DNA测序带入自动化测序的时代。这些技术统称为第一代DNA测序技术。最近几年发展起来的第二代DNA测序技术则使得DNA测序进入了高通量、低成本的时代。目前, 基于单分子读取技术的第三代测序技术已经出现, 该技术测定DNA序列更快, 并有望进一步降低测序成本, 改变个人医疗的前景。现介绍DNA测序技术的发展历史及不同

发展阶段各种主要测序技术的特点。

1 DNA测序技术的出现

早在DNA测序技术出现之前, 蛋白质和RNA的测序技术就已经出现。1949年, Frederick Sanger^[3]开发了测定胰岛素两条肽链氨基末端序列的技术, 并在1953年测定了胰岛素的氨基酸序列^[4, 5]。Edman^[6]也在1950年提出了蛋白质的N端测序技术, 后来在此基础上发展出了蛋白质自动测序技术^[7]。

Sanger等^[8]在1965年发明了RNA的小片段序列测定法, 并完成了大肠杆菌5S rRNA的120个核苷酸的测定。同一时期, Holley^[9, 10]完成了酵母丙氨酸转运RNA的序列测定。

收稿日期: 2010-02-02

基金项目: 重庆邮电大学博士启动金(A2009-18)

作者简介: 解增言, 男, 博士, 讲师, 研究方向: 生物信息学; E-mail: zengyanxie@gmail.com

©1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

DNA测序技术出现的较晚,1975年 Sanger和 Coulson^[1]发明了“加减法”测定 DNA序列。1977年在引入双脱氧核苷三磷酸(ddNTP)后,形成了双脱氧链终止法,使得 DNA序列测定的效率和准确性大大提高^[2]。Maxam和 Gilbert^[1]也在1977年报道了化学降解法测定 DNA的序列。DNA序列测定技术出现后,迅速超越了蛋白质和 RNA测序技术,成为现代分子生物学中最重要的技术。

2 第一代 DNA测序技术

传统的化学降解法、双脱氧链终止法以及它们的基础上发展来的各种 DNA测序技术统称为第一代 DNA测序技术。第一代测序技术在分子生物学研究中发挥过重要的作用,如人类基因组计划(human genome project HGP)主要基于第一代 DNA测序技术。目前基于荧光标记和 Sanger的双脱氧链终止法原理的荧光自动测序仪仍被广泛地应用。

2.1 化学降解法

在该方法中,一个末端被放射性标记的 DNA片段在5组互相独立的化学反应中分别被部分降解,其中每一组反应特异地针对某种碱基。因此生成5组放射性标记的分子,每组混合物中均含有长短不一的 DNA分子,其长度取决于该组反应所针对的碱基在原 DNA片段上的位置。最后,各组混合物通过聚丙烯酰胺凝胶电泳进行分离,再通过放射自显影来检测末端标记的分子^[12]。

化学降解法刚问世时,准确性较好,也容易为普通研究人员所掌握,因此用得较多。而且化学降解较之链终止法具有一个明显的优点,即所测序列来自原 DNA分子而不是酶促合成产生的拷贝,排除了合成时造成的错误。但化学降解法操作过程较麻烦,逐渐被简便快速的 Sanger法所代替。

2.2 双脱氧链终止法

双脱氧链终止法又称为 Sanger法^[2],该方法的原理是:核酸模板在 DNA聚合酶、引物、4种单脱氧核苷三磷酸(dNTP其中的一种用放射性 P₃₂标记)存在条件下复制时,在四管反应系统中分别按比例引入4种双脱氧核苷三磷酸(ddNTP),因为双脱氧核苷没有3'-OH,所以只要双脱氧核苷掺入链的末端,该链就停止延长,若链端掺入单脱氧核苷,链就

可以继续延长。如此每管反应体系中便合成以各自的双脱氧碱基为3端的一系列长度不等的核酸片段。反应终止后,分4个泳道进行凝胶电泳,分离长短不一的核酸片段,长度相邻的片段相差一个碱基。经过放射自显影后,根据片段3端的双脱氧核苷,便可依次阅读合成片段的碱基排列顺序。

Sanger法因操作简便,得到广泛的应用。后来在此基础上发展出多种 DNA测序技术,其中最重要的是荧光自动测序技术。

2.3 荧光自动测序技术

荧光自动测序技术基于 Sanger原理,用荧光标记代替同位素标记,并用成像系统自动检测,从而大大提高了 DNA测序的速度和准确性。

早在1985年,Smith等采用激光激发标记的荧光,并用 CCD检测,比常规电泳测序速度提高9倍,测序速度达到8 000 bp/h以上^[13]。

20世纪80年代初 Jorgenson和 Lukacs^[14 15]提出了毛细管电泳技术(capillary electrophoresis)。1992年美国的 Mathies实验室首先提出阵列毛细管电泳(capillary array electrophoresis)新方法,并采用激光聚焦荧光扫描检测装置,25只毛细管并列电泳,每只毛细管在1.5 h内可读出350 bp DNA序列分析效率可达6 000 bp/h^[16]。1995年 Walle^[17]研究组用该技术进行测序研究,使用四色荧光标记法,每个毛细管长3.5 m在9 min内可读取150个碱基,准确率约97%。

目前,应用最广泛的应用生物系统公司(applied biosystems ABI)3730系列自动测序仪即是基于毛细管电泳和荧光标记技术的 DNA测序仪。如 ABI 3730XL测序仪拥有96道毛细管,4种双脱氧核苷酸的碱基分别用不同的荧光标记,在通过毛细管时不同长度的 DNA片段上的4种荧光基团被激光激发,发出不同颜色的荧光,被 CCD检测系统识别,并直接翻译成 DNA序列。

2.4 杂交测序技术

杂交法测序是20世纪80年代末出现的一种测序方法,该方法不同于化学降解法和 Sanger法,而是利用 DNA杂交原理,将一系列已知序列的单链寡核苷酸片段固定在基片上,把待测的 DNA样品片段变性后与其杂交,根据杂交情况排列出样品的序列

信息^[18]。杂交测序检测速度快,采用标准化的高密度寡核苷酸芯片能够大幅度降低检测的成本^[19],具有部分第二代测序技术的特点。但该方法误差较大,且不能重复测定。

3 第二代 DNA测序技术

随着人类基因组计划的完成,人们进入了后基因组时代,即功能基因组时代,传统的测序方法已经不能满足深度测序和重复测序等大规模基因组测序的需求,这促使了新一代 DNA测序技术的诞生。新一代测序技术也称为第二代测序技术,主要包括罗氏 454 公司的 GS FLX 测序平台、Illumina 公司的 Solexa Genome Analyzer 测序平台和 AB 公司的 SOLiD 测序平台^[20]。

第二代测序技术最显著的特征是高通量,一次能对几十万到几百万条 DNA 分子进行序列测序,使得对一个物种的转录组测序或基因组深度测序变得方便易行^[21]。

新一代测序技术将片段化的基因组 DNA 两侧连上接头,随后用不同的方法产生几百万个空间固定的 PCR 克隆阵列。每个克隆由单个文库片段的多个拷贝组成。然后进行引物杂交和酶延伸反应。由于所有的克隆都在同一平面上,这些反应就能够大规模平行进行,每个延伸反应所掺入的荧光标记的成像检测也能同时进行,从而获得测序数据。DNA 序列延伸和成像检测不断重复,最后经过计算机分析就可以获得完整的 DNA 序列信息^[22]。

3.1 454 测序技术

454 生命科学公司在 2005 年最早推出了第二代测序平台 Genome Sequencer 20 并测序了支原体 *Mycoplasma genitalium* 的基因组^[23]。并在 2007 年推出性能更优的第二代基因组测序系统—Genome Sequencer FLX System (GS FLX)。罗氏在 2005 年便与 454 公司洽谈并购事宜,2007 年完成并购,紧接着公布了 DNA 双螺旋的发现者之一——沃森 (Jim Watson) 的个人基因组,测序总花费不到一百万美元^[24-25]。454 公司在 2010 年即将推出初级版的新一代测序仪 GS Junior,预计每次反应能得到 100 000 条序列,是 GS FLX 的十分之一,读取的序列平均长度为 400 bp 准确率达 99%,更适合规模较小的实验室^[26]。454 测序技术利用了焦磷酸测序原理^[25],

主要包括以下步骤。

3.1.1 文库准备 将基因组 DNA 打碎成 300—800 bp 长的片段 (若是 snRNA 或 PCR 产物可以直接进入下一步),在单链 DNA 的 3' 端和 5' 端分别连上不同的接头。

3.1.2 连接 带有接头的单链 DNA 被固定在 DNA 捕获磁珠上。每一个磁珠携带一个单链 DNA 片段。随后扩增试剂将磁珠乳化,形成油包水的混合物,这样就形成了许多只包含一个磁珠和一个独特片段的微反应器。

3.1.3 扩增 每个独特的片段在自己的微反应器里进行独立的扩增 (乳液 PCR, emulsion PCR),从而排除了其它序列的竞争。整个 DNA 片段文库的扩增平行进行。对于每一个片段而言,扩增产生几百万个相同的拷贝。乳液 PCR 终止后,扩增的片段仍然结合在磁珠上。

3.1.4 测序 携带 DNA 的捕获磁珠被放入 PIP 板中进行测序。PIP 孔的直径 (29 μm) 只能容纳一个磁珠 (20 μm)。放置在 4 个单独的试剂瓶里的 4 种碱基,依照 T A C G 的顺序依次循环进入 PIP 板,每次只进入一个碱基。如果发生碱基配对,就会释放一个焦磷酸。这个焦磷酸在 ATP 硫酸化酶和荧光素酶的作用下,释放出光信号,并实时地被仪器配置的高灵敏度 CCD 捕获到。有一个碱基和测序模板进行配对,就会捕获到一分子的光信号;由此一一对应,就可以准确、快速地确定待测模板的碱基序列^[27]。

与其它第二代测序平台相比,454 测序法的突出优势是较长的读长,目前 GS FLX 测序系统的序列读长已超过 400 bp。虽然 454 平台的测序成本比其他新一代测序平台要高很多,但对于那些需要长读长的应用,如从头测序,它仍是最理想的选择。

3.2 Solexa 测序技术

Illumina 公司的新一代测序仪 Genome Analyzer 最早由 Solexa 公司研发,利用合成测序 (Sequencing by Synthesis) 的原理,实现自动化样本制备及大规模平行测序。

Genome Analyzer 技术的基本原理是将基因组 DNA 打碎成约 100—200 个碱基的小片段,在片段

的两个末端加上接头(adapter)。将DNA片段变成单链后通过接头与芯片表面的引物碱基互补而使一端被固定在芯片上。另外一端随机和附近的另外一个引物互补,也被固定住,形成桥状结构。通过30轮扩增反应,每个单分子被扩增大约1000倍,成为单克隆的DNA簇,随后将DNA簇线性化。在下一步合成反应中,加入改造过的DNA聚合酶和带有4种荧光标记的dNTP。在DNA合成时,每一个核苷酸加到引物末端时都会释放出焦磷酸盐,激发生物发光蛋白发出荧光。用激光扫描反应板表面,在读取每条模板序列第一轮反应所聚合上去的核苷酸种类后,将这些荧光基团化学切割,恢复3'端黏性,随后添加第二个核苷酸。如此重复,直到每条模板序列都完全被聚合为双链。这样,统计每轮收集到的荧光信号结果,就可以得知每个模板DNA片段的序列^[28]。

Genome Analyzer系统需要的样品量低至100 ng,文库构建过程简单,减少了样品分离和制备的时间,配对末端读长可达到2×50 bp,每次运行后可获得超过20 GB的高质量过滤数据,且运行成本较低,是性价比较高的新一代测序技术。

3.3 SOLID测序技术

SOLID全称为Supported Oligo Ligation Detection,是ABI公司于2007年底推出的全新测序技术,目前已发展到SOLID3 Plus与454和Solexa的合成测序不同,SOLID是通过连接反应进行测序的。其基本原理是以四色荧光标记的寡核苷酸进行多次连接合成,取代传统的聚合酶连接反应。具体步骤包括:

3.3.1 文库准备 SOLID系统能支持两种测序模板:片段文库(fragment library)或配对末端文库(mate paired library)。片段文库就是将基因组DNA打断,两头加上接头,制成文库。该文库适用于转录组测序、RNA定量、mRNA研究、重测序、3',5'-RACE、甲基化分析及ChIP测序等。配对末端文库是将基因组DNA打断后,与中直接头连接,环化,然后用EcoP15酶切,使中直接头两端各有27 bp的碱基,最后加上两端的接头,形成文库。该文库适用于全基因组测序、SNP分析、结构重排及拷贝数分析等。

3.3.2 扩增 SOLID用的是与454技术类似的乳

液PCR对要测序的片段进行扩增。在微反应器中加入测序模板、PCR反应元件、微珠和引物,进行乳液PCR(emulsion PCR)。PCR反应结束后,磁珠表面就固定有拷贝数目巨大的同一DNA模板的扩增产物。

3.3.3 微珠与玻片连接 乳液PCR完成之后,变性模板,富集带有延伸模板的微珠,微珠上的模板经过3'修饰,可以与玻片共价结合。SOLID系统最大的优点就是每张玻片能容纳更高密度的微珠,在同一系统中轻松实现更高的通量。含有DNA模板的磁珠共价结合在SOLID玻片表面,SOLID测序反应就在SOLID玻片表面进行。每个磁珠经SOLID测序后得到一条序列。

3.3.4 连接测序 SOLID连接反应的底物是8碱基单链荧光探针混合物。探针的5'端用4种颜色的荧光标记,探针3'端第1、2位碱基是ATCG 4种碱基中的任何两种碱基组成的碱基对,共16种碱基对,因此每种颜色对应着4种碱基对。3—5位是随机的3个碱基。6—8位是可以和任何碱基配对的特殊碱基。单向SOLID测序包括5轮测序反应,每轮测序反应含有多次连接反应,得到原始颜色序列。

SOLID序列分析软件根据“双碱基编码矩阵”把碱基序列转换成颜色编码序列,然后与SOLID原始颜色序列进行比较。由于双碱基编码规则中一种颜色对应4种碱基对,前面碱基对的第二个碱基是后面碱基对的第一个碱基,所以一个错误颜色编码就会引起连锁的解码错误,改变错误颜色编码之后的所有碱基。SOLID序列分析软件可以对测序错误进行自动校正,最后“解码”成原始序列。因为SOLID系统采用了双碱基编码技术,在测序过程中对每个碱基判读两遍,从而减少原始数据错误,提供内在的校对功能,得到的原始碱基数据的准确度大于99.94%,而在15X覆盖率时的准确度可以达到99.999%,是目前新一代基因分析技术中准确度最高的^[29-30]。

超高通量是SOLID系统最突出的特点,目前SOLID3单次运行可产生50 GB的序列数据,相当于17倍人类基因组覆盖度。

3种新一代测序技术各有所长,它们各种特点的对比如表1。

表 1 三种第二代测序技术对比

测序技术	454	Solexa	SOLID
上市时间	2005	2007	2007
价格（万 美 元，2007年）	50	45	59
单次反应数据量（G）	0.4	20	50
读长（bp）	400	50×2	50
优势	长读长	低测序成本，高性价比	高通量，高精度

3.4 第二代测序技术的应用

3.4.1 从头测序（de novo sequencing）对于基因组未被测序过的生物，其基因组测序需要从头测序。由于受测序读取长度的限制，新一代测序技术中只有 454 技术能独立完成复杂基因组如真核生物基因组的从头测序工作。Solexa 和 SOLiD 技术只能完成简单生物如细菌的基因组的从头测序。在复杂基因组的从头测序中，将 Solexa/SOLiD 与 454 技术或传统的 Sanger 测序技术结合，分别利用它们的高通量和较长读长的优势，可以大大降低测序成本，提高测序速度。如最近完成的黄瓜全基因组测序，就联合运用了 Solexa 技术和 Sanger 技术，得到平均 72.2 倍覆盖度的黄瓜基因组序列，其中 3.9 倍是用 Sanger 技术得到的，68.3 倍是用 Solexa 的 GA 测序平台得到的^[31]。

3.4.2 重测序 如果对照一个参考基因组，新一代测序技术可以短时间内非常轻松的完成一个基因组的重测序。2008 年，由中国、美国和英国共同启动的千人基因组计划（1000 Genomes Project）是人类基因组计划的延续，也是迄今为止最大的基因组重测序计划。该计划打算测序大约 1 200 个全世界不同国家的人类个体的基因组。454、AB 和 Illumina 共同加入该计划^[32]。

3.4.3 SNP 研究 SNP 全称是 Single Nucleotide Polymorphism，意即单核苷酸多态性，是指不同个体的基因组上单个核苷酸的变异，包括替换、缺失和插入。SNP 在基因组中分布相当广泛，一般来说，SNP 是指变异频率大于 1% 的单核苷酸变异。SNP 可以作为新的遗传标记，人体许多表型差异，对药物或疾

病的易感性等等都可能与 SNP 有关。研究 SNP 是人类基因组计划走向应用的重要一环，因为 SNP 将提供一个强有力的工具，用于高危群体的发现、疾病相关基因的鉴定、药物的设计和测试以及生物学的基础研究等^[33-34]。

新一代测序技术利用其高通量、低成本的优势，对较多的个体测序，很容易可以得到大量的 SNP 位点。千人基因组计划即可得到众多的人类基因组 SNP 位点。其它的在 SNP 研究中的应用如利用 Solexa 测序技术对线虫 CB4858 品系进行重测序，寻找线虫（*Caenorhabditis elegans*）基因组中的 SNP 位点^[35]，以及寻找橄榄型油菜（*Brassica napus*）转录组（transcriptome）中的 SNP 位点^[36]等。

3.4.4 转录组及表达谱分析 基因表达谱（gene expression profile）指细胞在特定的条件下表达的所有基因。分析基因表达谱是了解组织或器官行使功能的分子基础及环境和病变影响生物的分子机制等的重要手段。以往的基因表达谱分析主要依靠基因芯片技术，该技术需要依赖已知的基因序列来设计探针，通过荧光标记和杂交，根据荧光的强度计算表达量的多少，误差较大，而且无法检测未知基因的表达量。

第二代测序技术可对单个细胞样品中的所有 RNA 即整个转录组进行整体测序，对每个细胞中表达 1—50 000 个拷贝 mRNA 的基因都能够检测，该技术还可以检测以前没发现过的基因或新的转录本（transcript），定量测定基因的表达模式^[37]。如 Morata 等^[38]利用 Solexa 技术对小鼠的大脑、肝脏和骨骼肌进行了 RNA 测序，对测得的每条序列进行计数从而获得每个特定转录本的表达量，能检测到丰度非常低的转录本。分析测得的序列，至少有 3 500 个基因拥有不止一种剪切形式，其中有接近 10% 是从未被报道过的 RNA 的剪切形式。同年 Sugarbaker^[39-40]利用 mRNA 深度测序对恶性胸腺瘤和对照样品进行比较，发现了肿瘤细胞基因组中存在的 15 个不同的点突变。

3.4.5 小分子 RNA（miRNA）研究 第二代测序技术还被广泛应用于小分子 RNA（miRNA、siRNA、mRNA）或非编码 RNA（ncRNA）表达研究。非编码的小分子 RNA 参与了许多重要的生物发育过程^[41]。它们

的序列长度很短,只有 18—40 个核苷酸,正好在新一代高通量测序技术的读长范围内。第二代测序方法还能发现新的小分子 RNA。如利用 Illumina 的新一代测序技术对人胚胎干细胞发育前后的分析,获得了 334 个已知的和 104 个新发现的小 RNA 的表达谱^[42]。

3.4.6 转录调控研究 (ChIP-seq) 染色体免疫共沉淀 (chromatin immunoprecipitation, ChIP) 技术是研究蛋白-DNA 相互作用的重要方法。以往的研究需结合芯片技术 (即 ChIP-on-chip 或 ChIP-chip), 因为芯片技术基于杂交原理, 需根据已知的序列设计探针, 对于未知的序列无能为力。把 ChIP 技术和第二代测序技术结合, 即 ChIP-sequencing (ChIP-seq), 可以在基因组水平上很方便的检测某种蛋白所结合的 DNA 序列, 全面了解蛋白与 DNA 的相互作用。如 2009 年 Ouyang 等^[43] 利用 ChIP-seq 技术发现, 在小鼠胚胎干细胞中, 大约有 65% 的基因表达是由 12 个转录因子调控的。

4 第三代 DNA 测序技术

第二代测序技术以其高通量、低成本的优势, 很快得到了广泛的应用。最近, 以单分子测序为特点的第三代 DNA 测序技术已经出现, 如生物科学公司 (BioScience Corporation) 的 Heliscope 单分子测序仪 (Heliscope Single Molecular Sequencer)^[44-45] 以及正在研制的太平洋生物科学公司 (Pacific Biosciences) 的单分子实时 DNA 测序技术 [Single Molecule Real Time (SMRT) DNA sequencing technology]^[46] 和牛津纳米孔技术公司 (Oxford Nanopore Technologies Ltd) 的纳米孔单分子测序技术^[47] 等^[48-49]。

斯坦福大学的科学家最近利用 Heliscope 单分子测序仪, 用了 48 000 美元的试剂和 4 个星期的时间, 对一名白人男子的基因组进行了测序。测序的覆盖度达 28 倍, 覆盖了 90% 的人类参考基因组。序列读长 24—70 bp 平均读长为 32 bp 并鉴定出 280 万个 SNP 位点和 752 个拷贝数变异^[50]。

目前, 我国也启动了第三代测序技术的研究。2009 年 12 月, 中科院北京基因组研究所与浪潮成立“中科院北京基因组研究所—浪潮基因组科学联合实验室”, 利用各自的优势联合研发国产第三代测序仪。第一台样机预计 2013 年问世, 届时有望缓

解测序仪核心技术受制于国外公司的现状。

5 展望

DNA 测序技术经过 30 多年的发展, 目前已经到了第三代, 三代测序技术有各自的优势。第一代测序技术虽然成本高, 速度慢, 但是对于少量的序列来说, 仍是最好的选择, 所以在以后的一段时间内仍将存在; 第二代测序技术刚刚商用不久, 正在逐渐走向成熟; 第三代测序技术有的刚刚出现, 有的则正在研制, 相信很快便可进行商业化运作。可以预见, 在未来的几年里会出现三代测序技术共存的局面。

随着新的测序技术的出现, 大规模测序的成本迅速下降, 花费 1 000 美元测一个人的基因组的目标相信很快就可以实现。届时, 对于遗传病的诊治将变得简单、快速, 并能从基因组水平上指导个人的医疗和保健, 从而进入个人化医疗的时代。

同时, 生物学研究的进展将会更多地依赖于测序技术的进步, 不同领域的科学家花很少的钱就可以对自己熟悉的物种基因组进行测序, 从而更好地指导试验设计, 取得更多新发现。

参考文献

- [1] Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 1977; 74(2): 560-564
- [2] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; 74(12): 5463-5467
- [3] Sanger F. The terminal peptides of insulin. *Biochem J* 1949; 45(5): 563-574
- [4] Sanger F, Thompson HJ. The amino acid sequence in the glycyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochem J* 1953; 53(3): 353-366
- [5] Sanger F, Thompson HJ. The amino acid sequence in the glycyl chain of insulin. II. The investigation of peptides from enzymic hydrolysates. *Biochem J* 1953; 53(3): 366-374
- [6] Edman P. Method for determination of the amino acid sequence in peptides. *Acta Chem Scand* 1950; 4: 283-293
- [7] Niall HD. Automated Edman degradation: the protein sequenator. *Methods Enzymol* 1973; 27: 942-1010
- [8] Sanger F, Brownlee GG, Barrel BG. A two-dimensional fractionation procedure for radioactive nucleotides. *J Mol Biol* 1965; 13(2): 373-398
- [9] Holley RW, Agar J, Everett GA, et al. Structure of a ribonucleic acid. *Science* 1965; 147: 1462-1465

- [10] Holley RW, Everett GA, Madison JT, et al. Nucleotide sequences in the yeast alanine transfer ribonucleic acid. *J Biol Chem* 1965; 240: 2122-2128.
- [11] Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 1975; 94(3): 441-448.
- [12] Maxam AM, Gilbert W. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* 1980; 65(1): 499-560.
- [13] Smith LM, Fung S, Hunkapiller MW, et al. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res* 1985; 13(7): 2399-2412.
- [14] Jorgenson JW, Lukacs KD. Free zone electrophoresis in glass capillaries. *Clin Chem* 1981; 27(9): 1551-1553.
- [15] Jorgenson JW, Lukacs KD. Capillary zone electrophoresis. *Science* 1983; 222(4621): 266-272.
- [16] Huang XC, Quesada MA, Mathies RA. DNA sequencing using capillary array electrophoresis. *Anal Chem* 1992; 64(18): 2149-2154.
- [17] Woolley AT, Mathies RA. Ultra-high-speed DNA sequencing using capillary electrophoresis chips. *Anal Chem* 1995; 67(20): 3676-3680.
- [18] Dmanac R, Labat J, Brakner J, et al. Sequencing of megabase plus DNA by hybridization: theory of the method. *Genomics* 1989; 4(2): 114-128.
- [19] 邱超, 孙含丽, 宋超. DNA测序技术发展历程及国际最新动态. 硅谷, 2008(17): 127-129.
- [20] Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26(10): 1135-1145.
- [21] Schuster SC. Next generation sequencing transforms today's biology. *Nat Methods* 2008; 5(1): 16-18.
- [22] 于亮. 放眼未来, 看新一代测序. 生物通, 2009 68: 2-4.
- [23] Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; 437(7057): 376-380.
- [24] Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massive parallel DNA sequencing. *Nature* 2008; 452(7189): 872-876.
- [25] <http://www.454.com/>.
- [26] <http://www.gs.junior.com/>.
- [27] 于亮. 新一代测序技术之三国时代(中). *Roche/454. 生物通*, 2009 68: 8-10.
- [28] 于亮. 新一代测序技术之三国时代(上). *illumina. 生物通*, 2009 68: 5-7.
- [29] 于亮, 吴青. 新一代测序技术之三国时代(下). *ABI. 生物通*, 2009 68: 11-14.
- [30] <http://solid.appliedbiosystems.com/>.
- [31] Huang S, Li R, Zhang Z, et al. The genome of the cucumber *Cucumis sativus* L. *Nat Genet* 2009; 41(12): 1275-1281.
- [32] http://www.1000genomes.org/Page_Php.
- [33] Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409(6822): 928-933.
- [34] Bertram M, Pajin A. Variety is the spice of eukaryotic life. *Nat Rev Microbiol* 2007; 5(9): 660-661.
- [35] Hillier LW, Marth GT, Quinlan AR, et al. Whole genome sequencing and variant discovery in *C. elegans*. *Nat Methods* 2008; 5(2): 183-188.
- [36] Trick M, Long Y, Meng J, et al. Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnol J* 2009; 7(4): 334-346.
- [37] Morozova O, Hirst M, Marra MA. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* 2009; 10: 135-151.
- [38] Moravzavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008; 5(7): 621-628.
- [39] Sugarbaker DJ, Richards WG, Gordon GJ, et al. Transcriptome sequencing of malignant pleural mesothelioma tumors. *Proc Natl Acad Sci USA* 2008; 105(9): 3521-3526.
- [40] 滕晓坤, 肖华胜. 基因芯片与高通量 DNA 测序技术前景分析. *中国科学*, 2008 38(10): 891-899.
- [41] Huttenhofer A, Schatner P, Polacek N. Non-coding RNAs: hope or hype? *Trends Genet* 2005; 21(5): 289-297.
- [42] Morin RJ, O'Connor MD, Griffith M, et al. Application of massive parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res* 2008; 18(4): 610-621.
- [43] Ouyang Z, Zhou Q, Wong WH. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc Natl Acad Sci USA* 2009; 106(51): 21521-21526.
- [44] Harris TD, Buzby PR, Babcock H, et al. Single molecule DNA sequencing of a viral genome. *Science* 2008; 320(5872): 106-109.
- [45] <http://www.helicosbio.com/>.
- [46] <http://www.pacificbiosciences.com/>.
- [47] <http://www.nanoporetech.com/>.
- [48] Rusk N. Cheap third-generation sequencing. *Nature* 2009; 46(4): 244-245.
- [49] 孙海汶, 王秀杰. DNA测序技术发展及其展望. *e-Science 技术*, 2009 6: 24-26.
- [50] Pushkarev D, Neff NF, Quake SR. Single molecule sequencing of an individual human genome. *Nat Biotechnol* 2009; 27(9): 847-852.