

微生物学报 *Acta Microbiologica Sinica*
51(4):445-457; 4 April 2011
ISSN 0001-6209; CN 11-1995/Q
<http://journals.im.ac.cn/actamicrocn>

高通量测序技术及其在微生物学研究中的应用

秦楠¹, 栗东芳¹, 杨瑞馥^{2*}

¹ 深圳华大基因研究院, 深圳 518083

² 病原微生物生物安全国家重点实验室, 军事医学科学院微生物流行病研究所, 北京 100071

摘要: 20世纪70年代发明的核酸测序技术为基因组学及其相关学科的发展做出了巨大贡献, 本世纪初发展的以 Illumina 公司的 HiSeq 2000, ABI 公司的 SOLiD, 和 Roche 公司的 454 技术为代表的高通量测序技术又为基因组学的发展注入了新活力。本文在阐述这些技术的基础上, 着重讨论了新一代测序技术在微生物领域中的应用。

关键词: 新一代测序技术, 高通量, 微生物学, 应用, 基因组学

中图分类号: 文献标识码: A 文章编号: 0001-6209 (2011) 04-0445-13

20世纪70年代, 由 Frederick Sanger 发明的双脱氧链终止法核酸测序技术, 为科学发展做出了重要贡献, 成就了人类基因组的完成。Sanger 测序法也被称为第一代测序技术^[1], 它的原理是以 DNA 单链为模板, 进行 PCR 扩增, 扩增体系中加入的碱基为 dNTP 和荧光 ddNTP, 对得到的不同长度产物进行电泳分离和激光诱导荧光颜色区分, 经过信息转换, 获得长达 800 bp 的 DNA 链碱基组成序列。该方法已经在 PCR 产物、载体克隆测序等方面得到广泛应用, 但其成本高和测序通量低的缺陷, 限制了该方法在大规模测序中的应用。

新一代测序技术 (next-generation sequencing technology, 或被称为第二代测序技术) 以 Illumina 公司的 Solexa, ABI 公司的 SOLiD, 和 Roche 公司的 454 技术为代表^[2-4]。这些测序平台以数据产出通量高为最大特点, 以 Solexa 技术为例, 采用了该技术的 HiSeq 2000 测序仪, 一台机器在两周内就可以产出

超过 300G 的数据, 相当于把人类基因组重复测 100 遍以上。这完全改变了过去的研究模式, 给人类和动植物基因组学、转录组学、宏基因组学研究等方面带来全新的变化, 并逐步深入到微生物学研究领域中。本综述旨在介绍代表性的高通量测序技术, 同时着重阐述该技术在微生物学研究中的应用。

1 高通量测序技术简介

1.1 Illumina Genome Analyzer 和 HiSeq 2000

Illumina 公司的新一代测序仪 (包括 Genome Analyzer 及其升级版 HiSeq 2000) 利用基于单分子簇的边合成边测序技术 (Sequencing by Synthesis, SBS) 和专有的可逆终止化学反应, 可以在短时间内获得大量数据^[2]。该测序技术将基因组 DNA 的随机片段附着到光学透明的玻璃表面, 这些 DNA 片段经过延伸和桥式扩增, 形成了具有数以亿计 Clusters

基金项目: 国家科技部 973 项目 (2009CB522600); 传染病重大专项 (2008ZX10004-009)

* 通信作者。Tel: +86-10-66948595; Fax: +86-10-63815689; E-mail: ruiyfuyang@gmail.com

作者简介: 秦楠 (1980-), 男, 山东潍坊人, 微生物学博士, 研究方向为微生物基因组学, 细胞信号转导和转录调控。E-mail: qinnan@gmail.com

收稿日期: 2010-09-15; 修回日期: 2010-10-22

的 Flow cell, 每个 Cluster 就是具有数千份相同模板的单分子簇; 然后对这些模板使用可逆终止并可移除的四色荧光染料, 进行边合成边测序。这种新方法确保了高精度和真实的单碱基连续测序, 为同聚物和重复序列的测序难题提供了很好的解决方案。

测序特点: ①通量高。目前一台机器在两周内最高可产出 360G 的数据; ②准确率高。≥98.5%, 同时也有效地解决了多聚重复序列的读取问题; ③成本低。低于传统 Sanger 测序技术成本的 1%; ④ DNA 序列的读取长度不断增加, 当前单条序列读长可达到 150 bp; ⑤可以进行 Pair-end (PE) 双向测序, PE 文库插入片段大小范围可由 150 bp 到 10 kb。正确选择插入片段长度有利于高重复序列含量基因组的组装, 这进一步扩展了该技术的应用范围。

1.2 Roche GS FLX Titanium System

2005 年底, 454 公司推出了革命性的基于焦磷酸测序法的超高通量基因组测序系统——Genome Sequencer 20 System, 被《Nature》杂志以里程碑事件报道, 开创了新一代测序技术的先河^[3]。最新的 Roche GS FLX 测序仪使用了一种叫做“Pico TiterPlate”(PTP) 的平板, 含有 160 多万个由光纤组成的孔, 每个孔中载有一个乳化扩增 (Emulsion PCR) 过的 DNA 单拷贝磁珠, 还有化学发光反应所需的各种酶和底物。测序时 4 种碱基依照 T、A、C、G 的顺序循环进入 PTP 板。如果发生碱基配对, 就会释放一个焦磷酸, 经过合成和化学发光反应, 光信号被高灵敏度 CCD 捕获到, 就可以准确、快速地确定待测模板的碱基序列。

测序特点: ①速度快。一个测序反应耗时 10 h, 获得 4–6 亿个碱基对。比传统的 Sanger 测序的方法快 100 倍; ②读长长。单条序列的读长平均可达到 450 bp; ③通量高。每个反应可以得到超过 100 万个序列读长; ④准确度高。读长超过 400 bp 时, 单一读长的准确性可以超过 99%; ⑤可以进行 Pair-End 测序研究。

1.3 AB SOLiD system

AB SOLiD sequencer 是由 ABI 公司研发的新一代高通量基因测序分析系统^[4], 该技术以用四色荧光标记寡核苷酸进行连续的连接反应为基础, 能够对单拷贝扩增的 DNA 片段进行大规模高通量并行测序, 根据双碱基编码原理进行数据比对。建库过

程使用微反应板和乳液 PCR/微珠富集。多轮测序反应和每轮多次连接反应, 保证了每个碱基判读 2 遍, 增加了序列读取准确性, 使原始碱基数据准确度大于 99.94%, 而在 15 × 覆盖率时准确度可达 99.999%。测序特点: ①可制备 Mate-paired 文库测序, 插入片段范围 600 bp–10 kb; ②通量高, 每台 SOLiD™ 4 System 测序仪在 15 天内能够获得 100G 的数据量; ③采用 Primer reset 方式, 保证了较低的噪音, 失败的 Round 可以重做; ④测序时采用连接反应, 稳定性高, 准确性高, 有效地解决了多聚核苷酸序列困难读取的问题; ⑤每个 DNA 碱基检测 2 次, 这增加了序列读取的准确性; 2-base encoding 可以用来鉴别 SNP。

1.4 单分子测序技术

基因组学的研究随着 Illumina, 454, SOLiD 等第 2 代测序平台的建立有了重大的突破, 随着 2008 年 Helicos Biosciences 的第一台单分子 DNA 测序仪的研发成功, 以及同样被称为第 3 代单分子测序技术的 PacBio SMRT, Nanopore and modified forms 和 ZS Genetics TEM 的相继研发, 基因组学将迎来进一步的改革^[5–7]。

Helicos 单分子测序仪利用合成测序理论, 将样本 DNA 数以百万碱基的单链分子绑定在该仪器特有的没有背景荧光的玻璃表面, 通过加入荧光标记的核苷酸 (一次加入 4 种核苷的一种) 和聚合酶到单分子阵列中, 核苷酸会结合到它的特异性接收位点上。激光激发结合上的核苷的标记, 使标记物释放出荧光, 相机以 15 毫秒速度快速扫描整个阵列, 检测特异性结合到片段上的碱基。在此之后, 结合的核苷酸对会被移除, 同时通过重复加入标记核苷酸进行循环反应。

单分子测序技术具有操作简单和周期短等特点, 目前第 2 代测序技术, 一个细菌基因组测序完成的时间也要在 15 天左右, 而单分子测序技术省略了 PCR 的过程, 可以在几小时内测序完成并得到有效数据, 同时避免了 PCR 带来的误差。而针对微生物转录组等方面的研究, 单分子测序技术中的 RNA 直接测序技术还可以直接读取细胞中的 RNA pool, 即转录组, 避免了合成互补 DNA, 缩短了研发周期的同时也大大提高了准确率^[7]。目前, 已经有研究利用 Helicos 单分子测序仪对病毒 M13 进行了重测序^[8]。测序过程中采用了所谓的 two-pass

sequencing 的技术,即对每个碱基进行 2 次读取,取 2 次一样的数据为可用数据,大大降低了错误率,并且通过对 DNA 模板中的每个分子单独的进行检测,避免了测序过程中片段的不同步合成导致的错误。

随着第 2 代测序技术的不断升级,以及第 3 代测序技术的研发和逐步走向市场,相信不久的将来测序成本还会快速下降,机器体积更小,通量更高的大规模测序仪器将会越来越大众化。

2 在微生物学研究中的应用

2.1 细菌分类学

自从 2400 年前亚里士多德提出物种的概念后,人们一直在追求着用自然分类的手段对物种进行分类。技术的进步以及分类概念的创新始终贯穿于原核生物分类体系的构建过程中^[9]。20 世纪 60 年代发展的 DNA 杂交和 GC 含量测定,使得细菌种的概念有了一个 DNA 同源性 70% 以上这个明确的指标,随着化学分类学和 16S rDNA 测序技术的发展,生物的分类概念从过去的五界和六界系统过渡到了 Woese 的三域理论^[10-11]。随着核酸测序技术和化学分类技术的不断发展,人们又提出了数值分类和多相分类学的概念,旨在更全面地描述物种的分类^[12]。但是,这些技术或多或少都存在不足,我们就像瞎子摸象一样,对物种的描述总是局部的,难以真正实现按照自然分类的理念对物种进行分类^[9]。随着基因组测序技术的进步,人们尝试着用这些信息(如核酸序列,翻译后的蛋白质序列)进行自然分类,而且与常用的 16S rDNA 测序为基础的分类一致性很好^[13-15]。Konstantinidis 等利用基因组中的平均氨基酸标识(Average amino acid identity, AAI)分析了 175 株已经测定全基因组序列细菌的关系,结果表明,相近分类单元分类在一起,在遗传和基因方面具有明显亲缘关系^[16];该研究小组还用平均核苷酸标识(Average nucleotide identity, ANI)分析了 70 株已经测定全基因组序列细菌的关系,结果显示,ANI 在 94% 左右时,相当于传统意义上的 70% DNA 同源性的种的界限^[17];Richter 等综合利用 ANI 和四联核苷酸标识(tetranucleotide signatures),结果表明,ANI 可以取代 DNA 杂交来定义种,进一步细化为以 ANI 95-96% 为种定义的标准,同时,作者还编制了一个软件 JSpecies,综合利用四联核

苷酸标识,该参数与 ANI 直接相关,且可以帮助我们清晰的定义一个种^[18]。作者还建议,从分类角度而言,只随机测定一个物种 20% 的基因序列就可以进行上述的比较和物种分类^[18]。

高通量测序技术的发展,使我们能在短时间内测定大量物种的核酸序列,对某个菌种,我们可以测定上百株细菌的基因组序列,这就为我们进行基于全基因组信息的分类标准建立奠定了良好基础。为此,美国微生物学会建议首先将所有已分离的模式株进行测序,并建议如果用 454 技术测序的话,只需测定 20x 覆盖度即可^[19]。Wu 等人综合分析了现有的 1000 多株细菌的全基因组序列,并测定了原核生物系统发育树关键分支上的 56 株菌,充分证明了全基因组序列在细菌分类和相关功能分析中的重要作用^[20]。

2.2 比较基因组学和演化基因组学

微生物基因组的演化主要包括以下 5 种机制:单碱基突变、基因组内重排、基因水平转移(基因获得)、基因缺失和重复序列拷贝数变异。通过这些突变机制,微生物获得了新特性,得以在与环境和宿主的相互斗争中生存下去^[21]。

过去的研究中,一系列分子生物学方法被设计用来获取各类基因组多态性数据,并进一步用于推测微生物的种群结构。已经得到广泛应用的方法包括:16S rRNA 分型技术、多位点序列分型(Multilocus sequence typing, MLST)、多位点串联重复序列分析(Multiple loci VNTR analysis, MLVA)同义单核苷酸多态性(Synonymous single nucleotide polymorphism, sSNP)分型技术、差异区段(Different region, DFR)分型以及基于规律成簇的间隔短回文重复(Clustered repeats interspaced short palindromic repeats, CRISPR, 在结核分枝杆菌中又被称为 Spoligotyping)的分型技术等^[22-27]。一些国际数据库被建立起来,存储和管理这些方法所产生的海量生物学数据,方便了国际不同实验室之间的标准化数据比对与交流^[28-29]。

尽管这些技术方法在微生物分型和演化历史重建方面发挥了重要作用,但其内在的缺陷是方法本身所不可避免的(图 1)。如 16S rRNA、MLST 和 CRISPR 技术,由于分辨率低,在分析亲缘关系近、变异小的物种时,就会出现图 1-B 所示的问题,将多个邻近基因型的样本混为一体。通过比对已有的少数基因组序列,找到差异位点,然后针对这些位点对其

他大量菌株进行扫描的方法,如 MLVA, DFR 等。由于考察的差异位点不全,将产生系统发育挖掘偏倚 (Phylogenetic discovery bias), 见图 1-C 和 D^[30]。sSNP 变异不导致编码蛋白的变化,因此不受自然选择压力影响,其多态性能客观的反映细菌进化情况,并且在增加所考察的 sSNP 位点数目时,可以得到

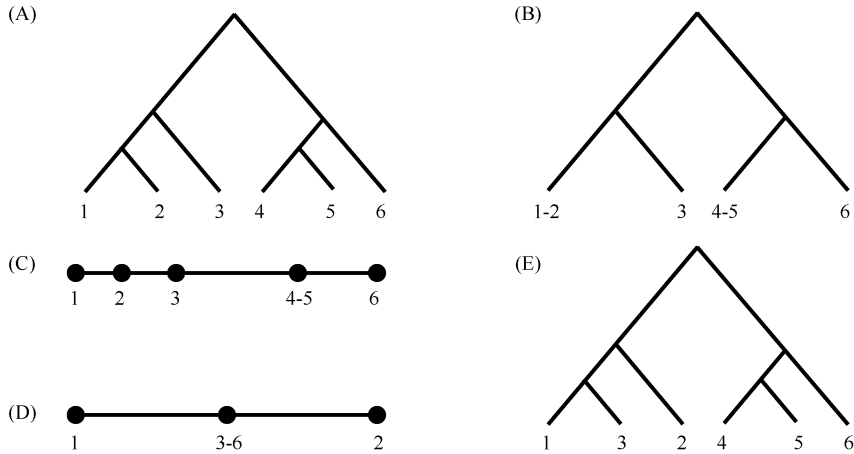


图 1 传统分子生物学方法在重建系统发育结构时可能发生的问题

Fig. 1 The potential flaws in phylogenetic tree reconstruction based on traditional molecular methods. The numbers in all panels denote the different samples. A. Real phylogeny; B. The phylogenetic tree based on methods with low resolution; C. Phylogenetic discovery bias (sample 1 and 6 were selected as reference strains); D. Lacking diversity between selected reference strains will lead to more serious phylogenetic discovery bias (sample 1 and 2 were selected as reference strains). E. High reverse or conversion mutation rate of selected targeting loci can result in wrong phylogenetic relationship. Panel C and D were cited from reference 28 with modification.

采用 Sanger 测序技术来获取全基因组序列并进行比较基因组学分析,可以弥补上述方法的不足。如通过对一种嗜甲烷菌的序列测定和分析,明确了该菌的系统发育地位和甲烷代谢相关代谢通路的变异、进化情况,丰富了对嗜甲烷菌多样性的认识^[32]。我国对采油微生物嗜热脱氮芽孢杆菌 (*Geobacillus thermodenitrificans*) 的全基因组破译,首次发现了重要的烷烃降解代谢路径,为清除石油污染带来新的思路,对于微生物采油技术的革新具有重要意义^[33]。固氮斯氏假单胞菌 (*Pseudomonas stutzeri*) 的全序列解读,发现其固氮相关基因集中在一个基因组岛上,该岛可能通过基因水平转移获得;对固氮代谢网络的分析为根系相关固氮菌在可持续发展农业中的应用提供了广阔前景^[34]。为了解析鼠疫菌在全球的传播演化,我们在世界各地选取 11 株代表性菌株进行测序,与已发表的 6 株鼠疫完成图序列一起进行比对,找到 933 个 SNP,并基于此对更大范围

比较理想的分辨率。但该方法也无法避免系统挖掘偏倚的影响。为了提高分辨率,有些方法选取了变异速率较高的靶标位点,如检测串联重复序列拷贝数变异的 MLVA 方法等。变异速率快会带来高回复和趋同突变的可能性,在构建系统发育树时,将导致拓扑结构不稳定甚至错误(图 1-E)^[31]。

的 286 株全球分离株进行了系统发育地理学和进化分析,全面揭示了鼠疫这种重要致病菌的传播历史和进化规律^[35]。

虽然 Sanger 测序法在微生物比较基因组学研究中发挥了重要作用,但其成本高、速度慢、数据产出量低的特点,阻碍了它的进一步广泛应用。与之相反,新一代高通量技术所需要的测序时间大大缩短,同时成本也越来越低,使得对所有目标样本进行全基因组测序和差异比较分析成为可能。该技术方法跨越了传统分子生物学研究方法所不能逾越的鸿沟,为比较基因组学、流行病学和微生物演化研究掀开了崭新一页。

第一,对全部目标样本全基因组序列之间进行比对,可以提供极高分辨率和准确的系统发育结构,并避免了系统发育挖掘偏倚问题。2010 年 1 月《Science》的一篇报道选取了 63 株抗药性金黄色葡萄球菌 (methicillin-resistant *Staphylococcus aureus*,

MRSA),其中包括20株7个月内,在同一医院的不同病房中分离到的菌株。尽管这些菌株间差异极小,使用全基因组测序和比对还是能够将其一一分开。通过构建菌株间的系统发育关系,可以追踪到院内传染的详细过程。该研究为阻断MRSA的院内感染提供了宝贵信息^[36]。另一项研究对95株A族链球菌(Group A *Streptococcus*, GAS)进行了高通量测序和比对分析,揭示了GAS的每次流行都是由不同于以前流行的菌株所导致的,而不是同一菌株的死灰复燃;并建立起菌株基因型和病人表型之间的关系^[37]。

第二,通过大样本的全基因组测序和比对,可以全面的发现自然压力选择下,变异规律特殊的基因位点,这些位点往往联系着重要的表型。在伤寒沙门氏菌(*Salmonella Typhi*)的研究中,26个基因被鉴定出受选择压力,其中半数编码表面暴露、输出或分泌相关蛋白,这可能跟病原菌改变抗原表型,逃避人体免疫系统压力有关^[38]。MRSA研究中发现的38个趋同进化突变位点里,有10个跟已知的抗药性机制相关^[36]。通过SNP在基因组上的非随机分布,研究者在GAS中鉴定出22个受正向选择的基因,其中包含多个已知的毒力因子,如*ropB*, *emm3*, *covR*和*covS*等^[37]。在29株艰难梭菌(*Clostridium difficile*)的测序研究中,使用dN/dS分析鉴定出的12个受正向选择的编码序列,这些序列编码应激调控蛋白和表面蛋白,提示了宿主免疫系统为该物种进化提供了选择压力^[39]。以上这些研究中均发现了一系列潜在的受选择基因,这对于各物种进一步的功能研究具有重要意义。

第三,对尽量多样化的微生物基因组进行测序,可以更准确的建立整个微生物界的种群结构,确立各个物种在演化中的地位,并能够充分挖掘微生物资源,为人类服务。目前,已经有超过1200种细菌的基因组被完整测序^[40]。这些测序对象的选择都是针对研究者的不同目的,如广受关注的致病菌等。结果导致对微生物基因组多样性的认知上产生严重偏倚^[41]。德国GCMCC研究所(German Collection of Microorganisms and Cell Cultures)和美国JGI研究所(Joint Genome Institute)共同发起了古细菌和真细菌基因组百科全书计划(Genomic Encyclopedia of Bacteria and Archaea, GEBA)。在第一步56株菌的测序和分析中,发现了1768个全新的蛋白家族,对

46个以前认为非同源的蛋白家族间建立了联系。另外也大大丰富了对非编码区域的认识,如在嗜盐粘细菌(*Haliangium ochraceum*)的807个CRISPR位点中,发现了目前已知最大的CRISPR元件,包含382个spacers。在接下来的几年中,他们将对更多的可培养微生物进行测序,建立完整、平衡的微生物生命之树^[20]。

2.3 转录组学

转录组学(transcriptomics),是一门在RNA水平上研究细胞中基因转录的整体情况及转录调控规律的学科。研究转录组最为广泛的方法是利用微阵列(microarray)技术检测有机体基因组中基因的表达。近年在微阵列技术基础上改进的瓦片阵列(tiling array)技术,使用了覆盖全基因组、相互交叠的探针,能够更精细的揭示RNA世界的状态和变化情况。该技术已经被成功应用到多种细菌的研究中^[42-45]。但除了存在背景干扰,饱和度,探针密度和质量等影响实验准确度的因素外,微阵列技术的最大缺点是无法进行*de novo*转录组研究^[46]。芯片探针设计要依赖于已有的参考基因组序列,不能发现每株菌特有的转录序列以及这些序列的表达水平变化。而使用RNA-seq方法,对全基因组cDNA进行高通量测序,可以从根本上解决这个问题。

RNA-seq在细菌中的研究和应用相对滞后于真核生物,这是因为细菌总RNA准备液中含量达80%以上的rRNA和tRNA^[47],直接进行反转录和测序会使结果产生很大的偏倚。与真核mRNA不同,大部分细菌mRNA没有poly-A尾,因此不能够通过杂交固定poly-T的方法直接对mRNA进行富集^[48]。因此在对细菌进行RNA-seq之前,必须选择适当的技术方法,尽量去除掉rRNA等干扰,从而间接的使mRNA富集起来。目前已应用于实际研究的方法包括:①使用包被了16S和23S rRNA的保守区域寡核苷酸探针的磁珠进行rRNA捕获;②使用专性消化rRNAs的新型核酸外切酶对RNA进行降解;③选择性反转录rRNA,利用核糖核苷酸酶及脱氧核糖核苷酸酶I消化去除反转录出的cDNA;④利用抗体捕获技术,去除与特定蛋白结合的rRNAs^[49]。

通过上述富集方法,可以获得足够的mRNA进行反转录和高通量测序,从而揭开了RNA-seq在细菌中应用的序幕:2009年,Yoder-Himes等使用Illumina测序方法,在*Burkholderia cenocepacia*中发

现了 13 种未知 ncRNA。并且发现尽管基因组相似度很高,来源不同的 *B. cenocepacia* 两菌株在调节反应上相差甚远,这或许可以解释他们栖息环境和致病潜力的差异^[50]。Passalacqua 等同时使用 Illumina 和 ABI SOLiD 两种测序技术,研究炭疽芽孢杆菌不同生长阶段和芽孢形成过程中的转录改变,两种技术的结果表现出很好的相关性。通过该研究,对炭疽芽孢杆菌的全基因组转录起始位点和操作子结构进行了预测,发现许多以前未被注释的基因组区域也表现出显著的转录活性^[51]。在伤寒沙门氏菌的 RNA-seq 研究中,Perkins 等发现了 40 个新的 ncRNA 序列,根据 RNA-seq 的结果对基因组序列的注释信息进行了校正,并找到 OmpR 操纵子的一些新成员^[52]。Cynthia 等对幽门螺杆菌的转录组测序研究发现,操作子内部存在数百个与已知基因方向相反的转录起始点,这说明其转录水平上存在的高度复杂性,并为其他物种的转录调控研究提供了新思路^[53]。

随着 RNA-seq 在细菌中的逐步应用,大量新的转录元件被发现,一系列新的转录调控规律被揭示出来,许多基因组注释得以修正。这些使得高精度转录组图谱成为可能,极大地有利于我们对微生物转录和后转录活动的研究。

2.4 宏基因组学

宏基因组学(meta-genomics)又称为环境基因组学,生态基因组学,或者群落基因组学,这是一门直接研究自然状态下微生物群落(包含了可培养的和不可培养的细菌、真菌和病毒等基因组的总和)的学科。1998 年威斯康辛大学植物病理学部门的 Handelsman 等在研究土壤微生物时,最早提出了“宏基因组学”这一概念。其想法源于将来自环境中的大量混杂物种的基因集在某种程度上作为一个单个基因组进行研究分析。后来伯克利分校的研究人员 Chen 和 Pachter 博士将宏基因组学定义为“应用现代基因组学的技术直接研究自然状态下的微生物的有机群落,而不需要在实验室中分离单一的菌株”的科学。“宏”,即“meta-”,具有更高层组织结构 and 动态变化的含义:第一,meta 意味着宏基因组学是一种试图从总体水平上研究的生物学,超越了微生物群落研究中只专注于某物种的基因以及这些基因如何影响其他物种;第二,meta 意味着需要开发新的算法用于最大程度上去解释从未完全描绘过

的微生物群落的基因组组成和群落的生命活动是如何复杂。

宏基因组学目前还是一门新兴学科,但由于其革命性的方法,克服了大多数微生物群落的不可培养性和基因组的多样性问题,取得了一系列的丰硕成果。一般来说,宏基因组学研究采用相似步骤:从特殊的环境中提取微生物的 DNA;DNA 用于直接测序、分析或者构建 BAC 文库用于测序以及功能筛选。较早的几项宏基因组研究是采用传统的 Sanger 测序技术,涉及的领域包括酸性矿山废水(acid mine drainage, AMD)^[54]、藻海(Sargasso Sea)^[55]、高等白蚁肠道^[56]等。尽管以上这些研究确立了宏基因组学在科学研究领域中的地位,但是由于测序方法低通量的原因,限制了所能研究的样品的物种复杂度。利用新一代测序技术对来自 AMD^[57]、海洋^[58-60]、土壤^[61-62]、珊瑚礁^[63]等环境样本的研究,获得了更加完整的 DNA 样品碱基序列。这不但有利于更深入的挖掘环境微生物的种类,还把对环境微生物的认识从物种水平上升到基因功能研究的水平,拓展了人们对环境微生物的认识。因此,新一代高通量测序技术更好地满足了宏基因组学研究的需求,并且促使宏基因组学研究领域获得爆发式地发展。

除了对环境微生物的关注外,另一个宏基因组学的研究热点是人类微生物组(human microbiome)。人体有许多高度特异的环境,如皮肤^[64]、口腔、鼻腔和肠道等,可供各种微生物生存,这些微生物与人体健康状态息息相关。目前研究最多的是寄居在人体胃肠道中的微生物,最早通过 16S rDNA 分类的方法^[65-67],来检测胃肠道物种丰度和多样性,从而对人体环境中微生物的组成和多样性的概况获得大致的了解。但是 16S rDNA 方法只能考察基因组中很小的一个片段的多态性,无法了解该环境下基因的整体组成和变异情况,所以现在很多研究单位倾向于采用 454/Solexa 等新一代测序技术对环境中的全部 DNA 进行全基因组测序分析^[68-72]。深圳华大基因研究院与欧盟合作,利用 Illumina 技术进行深度测序,在人类肠道宏基因组研究领域取得了突破性进展^[73]。该项研究成果收集了 124 个来自于欧洲人肠道菌群的样本,测序后产出近 6 千亿的碱基序列。经过序列组装和基因注释分析,从中获得 330 万个非冗余的人体肠道宏基因

组的参考基因,约是人自身基因的150倍。这个基因集中包含了绝大部分目前已知的人体肠道微生物基因,但更多的是未知微生物的基因。从这个基因集中可以估计人肠道中存在约1000到1150种细菌,平均每人体内约含有160种优势菌种,并且这些细菌是绝大部分个体所共有的。这些研究不仅将对人类自身健康与体内微生物的微妙关系做出科学的诠释,还有助于预防、监控和干预由肠道菌群引发的肥胖、肠炎和糖尿病等疾病。基于肠道宏基因组的个体化医疗时代已不遥远。

2.5 单细胞测序

单细胞测序(Single-cell sequencing)指从环境中分离单个细胞,通过全基因组扩增(Whole Genome Amplification, WGA)获取足够的DNA总量,从而进行全基因组测序的技术。由于自然环境中的微生物99%以上不能被提纯培养,使得无法直接进行测序,而单个微生物细胞中DNA含量稀少,无法满足测序的需求。单细胞测序技术克服了这些障碍,提供了全新的不依靠培养而研究微生物遗传多样性和进化历史的方法^[74]。同时,对单个细胞的全基因组测序还能够检测自然生境中微生物种类并进行微生物群体中细胞与细胞之间的变异研究^[75-77]。大规模范围的应用单细胞测序还能辅助宏基因组学研究,弄清具有遗传多样性的细胞如何在微生物群体中共存的机制。

单细胞测序的第一步是单细胞的分离。目前已报道的单细胞分离技术主要有:微流体技术(microfluidics)^[78],梯度稀释法(Serial dilution)^[79],显微操作技术(micromanipulation)^[75],荧光激活细胞分类(FACS, fluorescence-activated cell sorting)^[80]。基于流式细胞仪的FACS法目前使用的较多,它可以使单个细胞准确迅速的进入标准的96或384孔板,然后以高能量激光照射高速流动状态下被荧光色素染色的单细胞或微粒,测量其产生的散射光和发射荧光的强度,从而对细胞(或微粒)的物理、生理、生化、免疫、遗传、分子生物学性状及功能状态等进行定性或定量检测,并根据发射光的荧光强度和波长将发光颗粒亚群分开并可实现单克隆分选。

提取到单细胞后,要对细胞中的DNA进行全基因组扩增。这一步的常用方法包括散在重复序列PCR(IRS-PCR, interspersed repeat sequence

PCR)^[81]、接头引物PCR(LA-PCR, linker adaptor PCR)^[82]、T-PCR(Tagged-PCR)^[83]、简并寡核苷酸引物PCR(DOP-PCR, degenerate oligonucleotide primed PCR)^[84]、引物延伸预扩增PCR(PEP-PCR, primer extension preamplification PCR)^[85]和多重置换扩增(MDA, multiple displacement amplification)^[86]等。DOP-PCR和PEP-PCR的基本原理都是通过其随机引物与基因组DNA多处退火从而使大部分基因组序列得到扩增^[87]。而目前使用最广泛的MDA法主要基于随机引物、 ϕ 29 DNA聚合酶的高持续性(持续性指的是在聚合酶离开模板链之前,加入到DNA链上进行连接的平均碱基数)和序列替换能力。随机引物首先结合到模板链上取代对应的互补链,前期扩增的产物能够在替换掉的互补链上产生引物并继续替代进行扩增,MDA法能够忠实的复制整个基因组DNA,扩增出10-100 kb大小的片段。

完成全基因组扩增后,就可以利用高通量测序技术,实现测序和后续分析了。在微生物领域已经有一些研究证实了单细胞测序技术的可靠性。如对海洋藻青菌(marine cyanobacterium *Prochlorococcus*)的单细胞测序证实了单细胞测序*de novo*组装的可靠性和准确性^[87];对苛养木杆菌的1000个细胞进行DNA扩增并进行后续的建库和测序,通过对WGA法和标准Cs-Cl方法提取的DNA库测序进行比较发现:WGA所获得的基因组无偏差的覆盖了整个基因组,保证了后续研究的准确性。

单细胞测序技术的更广泛的应用还面临着一些问题:(1)难以准确有效的保证大批量的细胞同时被分离,限制了高通量筛选的需求;(2)难以处理DNA污染;(3)难以保证WGA过程中只针对一个细胞进行扩增。虽然FACS法针对其他3种分离方法有一定的优势,但是面对研究的日益广泛和量的增多,针对不同样品的高通量提取方法都有待进一步的开发。而WGA中目前应用最广泛的MDA法,首先必须依靠随机引物来进行初始聚合反应,这就导致在反应过程中可能同时对目标DNA和污染DNA进行同时扩增,这就导致在研究一个新物种时,很可能将非目标DNA拼接到目标序列中;扩增过程中还可能会导致基因组重排或断裂,从而影响后续的非连续区域染色体的组装;而在模板DNA非常少的情况下,使得基因组每个区域的扩增呈不均一性,从而导致测序建库过程

中的偏差。

3 展望

高通量测序技术领域的发展,使得核酸序列的测定速度与成本价格的比值,以超过摩尔定律的速度迅速增加。因此,全基因组(转录组)序列测序技术将不可避免地像 PCR 技术那样,走入每一个分子生物学实验室,成为微生物学家们的常规研究手段。而高通量测序技术所带来的基因组学和转录组学研究的变革,会为蛋白质组学和代谢组学奠定基础,它们将共同成为系统微生物学发展的基石。但问题和挑战依然存在。首先,随着核酸序列数量上的跨越式累积,生物信息学分析将面临巨大的挑战,适用于常规实验室使用的新统计学方法和分析软件的开发成为当前的迫切需求。另外,海量数据的深入挖掘工作会发现用传统生物学理论难以解释的生命规律,对传统理论的颠覆和新理论的提出与建立将成为不可避免的工作。

在《物种起源》的“学说的难点”一章中,达尔文曾经这样写道:“对于每一轻微变异或个体差异的原因,我们是深刻无知的……”^[88]。利用新一代高通量测序技术,我们可以发现基因组和转录组上所有可能的细微差异;用单碱基的精确度,去探寻微生物生存和演化过程中留下的痕迹,从而更深入的理解微生物世界。

致谢 感谢崔玉军、覃俊杰、戴文魁和才必华在资料收集和文章撰写中提供的帮助。

参考文献

- [1] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 1977, 74: 5463-5467.
- [2] Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. A large genome center's improvements to the Illumina sequencing system. *Nature Methods*, 2008, 5: 1005-1010.
- [3] Meyer M, Stenzel U, Hofreiter M. Parallel tagged sequencing on the 454 platform. *Nature Protocols*, 2008, 3: 267-278.
- [4] Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genetics*, 2008, 24: 133-141.
- [5] Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature nanotechnology*, 2009, 4, 265-270.
- [6] Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Real-time DNA sequencing from single polymerase molecules. *Science*, 2009, 323: 133-138.
- [7] Ozsolak F, Platt AR, Jones DR, Reifengerger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. Direct RNA sequencing. *Nature*, 2009, 461: 814-818.
- [8] Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z. Single-molecule DNA sequencing of a viral genome. *Science*, 2008, 320, 106-109.
- [9] Rossello-Mora R. Updating prokaryotic taxonomy. *Journal of bacteriology*, 2005, 187: 6255-6257.
- [10] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 1990, 87: 4576-4579.
- [11] Staley JT. Universal species concept: pipe dream or a step toward unifying biology? *Journal of industrial microbiology & biotechnology*, 2009, 36: 1331-1336.
- [12] Uilenberg G, Goff WL. Polyphasic taxonomy. *Annals of the New York Academy of Sciences*, 2006, 1081: 492-497.
- [13] Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. *Journal of bacteriology*, 2005, 187: 6258-6264.

- [14] Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era : advancements and challenges ahead. *Current opinion in microbiology*, 2007, 10: 504-509.
- [15] Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106: 19126-19131.
- [16] Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. *Journal of bacteriology*, 2005, 187: 6258-6264.
- [17] Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102: 2567-2572.
- [18] Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106, 19126-19131.
- [19] Riley M, Buckley M. Large Scale Sequencing: the future of genomic sciences? A Report from the American Academy of Microbiology. <http://www.asm.org>, 2009.
- [20] Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D'Haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng JF, Lucas S, Kerfeld C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk HP, Eisen JA. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, 2009, 462: 1056-1060.
- [21] Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. *Nature reviews, Microbiology*, 2008, 6: 431-440.
- [22] Cui Y, Li Y, Gorge O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV, Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS One*, 2008, 3: e2652.
- [23] Li Y, Cui Y, Hauck Y, Platonov ME, Dai E, Song Y, Guo Z, Pourcel C, Dentovskaya SV, Anisimov AP, Yang R, Vergnaud G. Genotyping and phylogenetic analysis of *Yersinia pestis* by MLVA: insights into the worldwide expansion of Central Asia plague foci. *PLoS One*, 2009, 4: e6000.
- [24] Li Y, Dai E, Cui Y, Li M, Zhang Y, Wu M, Zhou D, Guo Z, Dai X, Cui B, Qi Z, Wang Z, Wang H, Dong X, Song Z, Zhai J, Song Y, Yang R. Different region analysis for genotyping *Yersinia pestis* isolates from China. *PLoS One*, 2008, 3: e2166.
- [25] Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, 95: 3140-3145.
- [26] Roumagnac P, Weill FX, Dolecek C, Baker S, Brisse S, Chinh NT, Le TA, Acosta CJ, Farrar J, Dougan G, Achtman M. Evolutionary history of *Salmonella typhi*. *Science*, 2006, 314: 1301-1304.
- [27] Woese CR. Bacterial evolution. *Microbiological Reviews*, 1987, 51: 221-271.
- [28] Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajj SA, Allix C, Aristimuno L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes HM, Guttierrez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Mokrousov I, Narvskaja O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasolofoa-Razanamparany V, Rasolonavalona T, Rossetti ML, Rusch-Gerdes S, Sajduda A, Samper S, Shemyakin IG, Singh UB, Somoskovi A, Skuce RA, van Soolingen D, Streicher EM, Suffys PN, Tortoli E, Tracevska T, Vincent V, Victor TC, Warren RM, Yap SF, Zaman K, Portaels F, Rastogi N, Sola C. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiology*, 2006, 6: 23.
- [29] Le Fleche P, Hauck Y, Onteniente L, Prieur A, Denoeud F, Ramiise V, Sylvestre P, Benson G, Ramiise F, Vergnaud G. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiology*, 2001, 1: 2.
- [30] Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'Ren JM, Simonson TS, Kachur SM, Leadem RR,

- Cardon ML, Van Ert MN, Huynh LY, Fraser CM, Keim P. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101: 13536-13541.
- [31] Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria : DNA sequencing in *mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One*, 2009, 4: e7815.
- [32] Dunfield PF, Yuryev A, Senin P, Smirnova AV, Stott MB, Hou S, Ly B, Saw JH, Zhou Z, Ren Y, Wang J, Mountain BW, Crowe MA, Weatherby TM, Bodelier PL, Liesack W, Feng L, Wang L, Alam M. Methane oxidation by an extremely acidophilic bacterium of the phylum *Verrucomicrobia*. *Nature*, 2007, 450: 879-882.
- [33] Feng L, Wang W, Cheng J, Ren Y, Zhao G, Gao C, Tang Y, Liu X, Han W, Peng X, Liu R, Wang L. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104: 5602-5607.
- [34] Yan Y, Yang J, Dou Y, Chen M, Ping S, Peng J, Lu W, Zhang W, Yao Z, Li H, Liu W, He S, Geng L, Zhang X, Yang F, Yu H, Zhan Y, Li D, Lin Z, Wang Y, Elmerich C, Lin M, Jin Q. Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105: 7564-7569.
- [35] Morelli G, Song Y, Mazzoni CJ, Eppinger MPR, Wagner DM, Feldkamp M, Kusecek B, Vogler AJ, Li YYC, Thomson NR, Jombart T, Leblois R, Lichtner P, Rahalison L, Petersen JMDFB, Keim PS, Wirth T, Ravel J, Yang RE C, Achtman M. Phylogenetic diversity and historical patterns of pandemic spread of *Yersinia pestis*. *Nature genetics*, 2010, doi: 10.1038.
- [36] Harris S, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day NPJ, Lindsay JA, Edgeworth JD, De Lencastre H, Parkhill J, Peacock SJ, Bentley SD. Evolution of MRSA during hospital transmission and intercontinental spread. *Science*, 2010, 327: 469-474.
- [37] Beres SB, Carroll RK, Shea PR, Sitkiewicz I, Martinez-Gutierrez JC, Low DE, McGeer A, Willey BM, Green K, Tyrrell GJ, Goldman TD, Feldgarden M, Birren BW, Fofanov Y, Boos J, Wheaton WD, Honisch C, Musser JM. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107: 4371-4376.
- [38] Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M, Dougan G. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. *Nature Genetics*, 2008, 40: 987-993.
- [39] He M, Sebahia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, Holt KE, Seth-Smith HM, Quail MA, Rance R, Brooks K, Churcher C, Harris D, Bentley SD, Burrows C, Clark L, Corton C, Murray V, Rose G, Thurston S, van Tonder A, Walker D, Wren BW, Dougan G, Parkhill J. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107: 7527-7532.
- [40] Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic acids research*, 2010, 38: D346-354.
- [41] Hugenholtz P. Exploring prokaryotic diversity in the genomic era. *Genome Biology*, 2002, 3, REVIEWS0003.
- [42] McGrath PT, Lee H, Zhang L, Iniesta AA, Hottes AK, Tan MH, Hillson NJ, Hu P, Shapiro L, McAdams HH. High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. *Nature Biotechnology*, 2007, 25: 584-592.
- [43] Selinger DW, Cheung KJ, Mei R, Johansson EM, Richmond CS, Blattner FR, Lockhart DJ, Church GM. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nature Biotechnology*, 2000, 18: 1262-1268.
- [44] Rasmussen S, Nielsen HB, Jarmer H. The transcriptionally active regions in the genome of *Bacillus subtilis*. *Molecular Microbiology*, 2009, 73: 1043-1057.
- [45] Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet - Revillet H, Balestrino D, Loh E, Gripenland J,

- Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Regnault B, Coppee JY, Lecuit M, Johansson J, Cossart P. The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature*, 2009, 459: 950-956.
- [46] Bloom JS, Khan Z, Kruglyak L, Singh M, Caudy AA. Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics*, 2009, 10: 221.
- [47] Condon C. Maturation and degradation of RNA in bacteria. *Current opinion in microbiology* 2007, 10: 271-278.
- [48] Deutscher MP. Degradation of stable RNA in bacteria. *Journal of Biological Chemistry*, 2003, 278: 45041-45044.
- [49] Sorek R, Cossart P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nature reviews, Genetics*, 11: 9-16.
- [50] Yoder-Himes DR, Chain PS, Zhu Y, Wurtzel O, Rubin EM, Tiedje JM, Sorek R. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106: 3976-3981.
- [51] Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH. Structure and complexity of a bacterial transcriptome. *Journal of bacteriology*, 2009, 191: 3203-3211.
- [52] Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR, Dougan G. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genetics*, 2009, 5: e1000569.
- [53] Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, 2010, 464: 250-255.
- [54] Edwards KJ, Bond PL, Gihring TM, Banfield JF. An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. *Science*, 2000, 287: 1796-1799.
- [55] Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 2004, 304: 66-74.
- [56] Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernandez M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, 2007, 450: 560-565.
- [57] Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC, Jr, Rohwer F. Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics*, 2006, 7: 57.
- [58] Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F. The marine viromes of four oceanic regions. *PLoS Biology*, 2006, 4: e368.
- [59] Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. Microbial population structures in the deep marine biosphere. *Science*, 2007, 318: 97-100.
- [60] Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103: 12115-12120.
- [61] Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R, Rohwer F, Jackson RB. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and environmental microbiology*, 2007, 73: 7059-7066.
- [62] Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, Prosser JI, Schuster SC, Schleper C. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, 2006, 442: 806-809.

- [63] Wegley L, Edwards R, Rodriguez-Brito B, Liu H, Rohwer F. Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environmental microbiology*, 2007, 9: 2707-2719.
- [64] Fierer N, Hamady M, Lauber CL, Knight R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105: 17994-17999.
- [65] Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI. The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101: 15718-15723.
- [66] Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science*, 2005, 307: 1915-1920.
- [67] Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102: 11070-11075.
- [68] Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. Metagenomic analysis of the human distal gut microbiome. *Science*, 2006, 312: 1355-1359.
- [69] Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. *PLoS Biology*, 2007, 5: e177.
- [70] Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 2006, 444: 1027-1031.
- [71] Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, Taylor TD, Noguchi H, Mori H, Ogura Y, Ehrlich DS, Itoh K, Takagi T, Sakaki Y, Hayashi T, Hattori M. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Research*, 2007, 14: 169-181.
- [72] Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. *Nature*, 2009, 457: 480-484.
- [73] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Jian M, Zhou Y, Li Y, Zhang X, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 2010, 464: 59-65.
- [74] Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome research*, 2001, 11: 1095-1099.
- [75] Kvist T, Ahring BK, Lasken RS, Westermann P. Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Applied microbiology and biotechnology*, 2007, 74: 926-935.
- [76] Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA, Quake SR. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104, 11889-11894.
- [77] Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H, Saw JH, Senin P, Yang C, Chatterji S, Cheng JF, Eisen JA, Sieracki ME, Stepanauskas R. Assembling the marine metagenome, one cell at a time. *PLoS One*, 2009, 4: e5299.
- [78] Marcy Y, Ishoey T, Lasken RS, Stockwell TB, Walenz BP, Halpern AL, Beeson KY, Goldberg SM, Quake SR. Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS genetics*, 2007, 3: 1702-1708.
- [79] Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, Church GM. Sequencing genomes from single cells by polymerase cloning. *Nature biotechnology*, 2006, 24: 680-686.
- [80] Stepanauskas R, Sieracki ME. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104: 9052-9057.
- [81] Nelson DL, Ledbetter SA, Corbo L, Victoria MF,

- Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proceedings of the National Academy of Sciences of the United States of America*, 1989, 86: 6686-6690.
- [82] Ludecke HJ, Senger G, Claussen U, Horsthemke B. Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. *Nature*, 1989, 338: 348-350.
- [83] Grothues D, Cantor CR, Smith CL. PCR amplification of megabase DNA with tagged random primers (T-PCR). *Nucleic acids research*, 1993, 21: 1321-1322.
- [84] Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics*, 1992, 13: 718-725.
- [85] Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. Whole genome amplification from a single cell: implications for genetic analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 1992, 89: 5847-5851.
- [86] Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Chang HW, Watson D, Brodie EL, Hazen TC, Keller M. Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Applied and environmental microbiology*, 2006, 72: 3291-3301.
- [87] Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. Whole genome amplification and de novo assembly of single bacterial cells. *PLoS One*, 2009, 4: e6864.
- [88] Darwin C. *The Origin of Species*. London: Penguin Classics, 1859, 226.

Next-generation sequencing technologies and the application in microbiology— A review

Nan Qin¹, Dongfang Li¹, Ruifu Yang^{2*}

¹ BGI-Shenzhen, Shenzhen 518083, China

² State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China

Abstract: Since its invention in 1970s, nucleic acid sequencing technology has contributed tremendously to the genomics advances. The next-generation sequencing technologies, represented by HiSeq 2000 from Illumina, SOLiD from Applied Biosystems and 454 from Roche, re-energized the application of genomics. In this review, we first introduced the next-generation sequencing technologies, then, described their potential applications in the field of microbiology.

Keywords: next-generation sequencing technology, high-throughput, microbiology, application, genomics

(本文责编: 张晓丽)

Supported by the State Key Development Program for Basic Research of China (2009CB522600) and by the National Key Program for Infectious Diseases of China (2008ZX10004-009).

* Corresponding author. Tel: +86-10-66948595; Fax: +86-10-63815689; E-mail: ruifyang@gmail.com

Received: 15 October 2010/Revised: 22 October 2010