

土壤宏基因组学研究方法与进展*

贺纪正^{1†} 袁超磊^{1,2} 沈菊培¹ 张丽梅¹

(1 中国科学院生态环境研究中心, 北京 100085)

(2 中国科学院研究生院, 北京 100049)

摘要 土壤微生物驱动着土壤中的物质循环和养分转化。在土壤学的研究中, 长期将土壤作为一个黑箱系统来对待, 对其中的生物组成及其参与的生化过程知之甚少。土壤中绝大部分微生物目前尚难以分离培养, 因此基于传统的培养方法对于认识土壤微生物群落组成和功能有其局限性。宏基因组学直接从环境样品中提取全部微生物的 DNA, 或通过测序探究环境中微生物的群落结构和功能(序列驱动), 或构建宏基因组文库, 筛选新的基因或生物活性物质(功能驱动), 克服了传统培养方法的缺陷, 极大地丰富了对土壤微生物多样性及其功能的认知。本文在综述土壤宏基因组学研究基本流程的基础上, 重点介绍了日益重要的第二代测序平台在土壤宏基因组学研究中的应用及其产生的海量数据的分析处理方法, 并简要探讨了宏基因组学在土壤微生物生态学中的应用。最后, 作者建议在国家层面上展开相关土壤宏基因组学研究, 调查微生物群落及其变化, 为生物资源开发、农业生产和环境保护作出应有的贡献。

关键词 宏基因组学; 研究方法; 第二代测序平台; 微生物生态学; 土壤环境

中图分类号 Q938.1⁺3 **文献标识码** A

微生物是土壤生态系统的核心, 主要充当分解者的角色, 其对有机物的分解与转化、养分循环与利用、土壤肥力形成、温室气体产生、污染环境净化等起着关键作用。土壤微生物组成复杂、类群繁多、数量巨大、功能多样, 组成了地球上最丰富的生物资源库, 也是最重要的基因资源库和代谢产物库。对土壤微生物群落多样性特征、分布格局、重要元素循环的生态过程介导等的考察, 受到土壤学、生态学和微生物学等领域学者的普遍重视, 成为当前研究的前沿和热点。

土壤是微生物的大本营, 据估计, 全球土壤中大约有 2.6×10^{29} 个原核生物细胞^[1], 但其中绝大部分是不可培养的。因此, 利用传统的分离培养方法不能完全反映土壤环境中的微生物群落组成及多样性特征。基于土壤微生物 DNA/RNA 的分子生物学技术迅速发展, 为突破这一瓶颈起到了极大的推动作用, 并在土壤微生物生态学研究中得到广泛应用。特别是近年来, 随着测序技术的发展及研究工作的深入, 土壤宏基因组学(metagenomics) 日益得到关注。

宏基因组学既是一套研究方法, 也是一个研究领域^[2]。Handelsman 等于 1998 年首先提出了宏基因组(metagenome)的概念, “meta”有“集合”或“整体”的意思, 指的是土壤微生物的全部基因组的集合^[3], 或称环境基因组^[4]。发展至今, 宏基因组学是指对环境直接获得的总 DNA 进行分析的所有研究^[5]。

基于 DNA 的宏基因组学理论上覆盖了环境样品中的全部微生物, 因此可以更加全面真实地反映微生物群落组成, 同时大大拓展了筛选新的基因或生物活性物质的来源。根据所用策略不同, 宏基因组学研究可分为序列驱动的(sequence-driven)和功能驱动的(function-driven); 序列驱动是指通过测序分析微生物群落的结构和功能, 功能驱动是指基于构建宏基因组文库筛选新基因或新物质的宏基因组学研究(图 1)。本文主要介绍了这两大类宏基因组学研究的相关方法和过程, 涉及 DNA 的提取、宏基因组克隆文库构建及筛选、宏基因组测序以及数据分析等; 并对宏基因组学在土壤微生物生态学中的应用做了简单讨论。

* 国家自然科学基金项目(41025004, 41090281, 40871129)资助

† 通讯作者, E-mail: jzhe@rcees.ac.cn

作者简介: 贺纪正(1965—), 男, 研究员, 博士, 主要从事土壤分子生态学研究。E-mail: jzhe@rcees.ac.cn

收稿日期: 2011-03-18; 收到修改稿日期: 2011-05-27

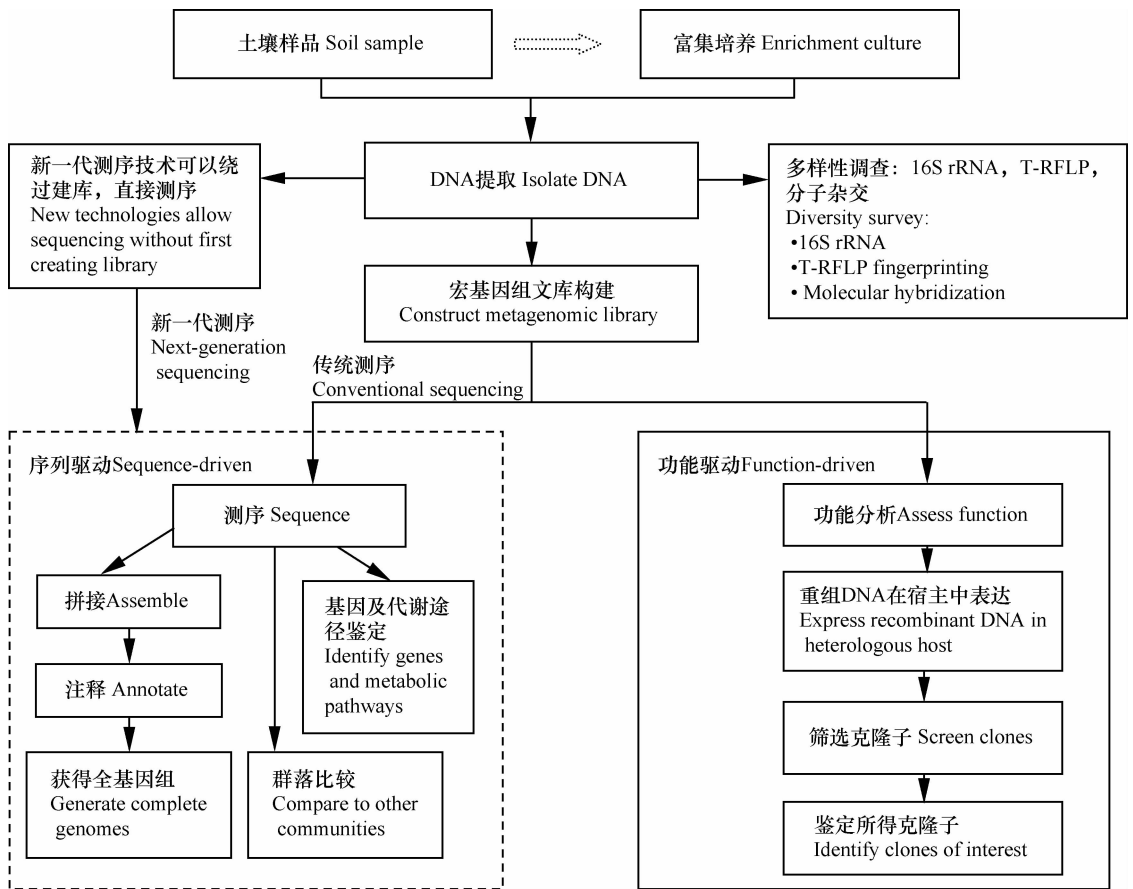


图1 宏基因组学研究流程(参考文献[2]和[6])

Fig. 1 Research procedures of soil metagenomics (according to reference [2] and [6])

1 土壤 DNA 的提取

土壤样品的采集是关键的第一步,采样需遵循操作规程,采取具有代表性的样品,缩短运输和保存的时间,使样品能更好地代表自然状态下土壤微生物的原貌^[7]。由于微生物细胞与土壤颗粒的结合以及有机质的干扰,获得高质量的土壤 DNA 有一定难度^[7]。土壤 DNA 的提取方法主要有两种:一是直接(裂解)法,即直接裂解土壤中的细胞,释放 DNA,再将 DNA 与土壤颗粒和其他杂质分离;另一种是细胞分离提取法或间接法,即先分离土壤中的微生物细胞,再裂解细胞提取 DNA^[8]。直接法提取 DNA 的效率较间接法要高,一般认为其所得 DNA 更具有代表性,因而得到更广泛的应用^[9]。目前基于直接法的土壤 DNA 提取试剂盒较多,但提取效率不尽一致^[10]。针对更为复杂的土壤环境样品,例如高有机质、高铁锰氧化物含量的土壤样品,需对 DNA 提取方法进行一些改进,以提取高质量的土

壤 DNA^[11-12]。

2 土壤宏基因组克隆文库的构建及筛选

2.1 克隆文库的构建

从土壤中提取的 DNA 可用于构建克隆文库进而筛选一些重要的生物活性分子。在建库选择载体和宿主时,应考虑到自身的研究目的以及所获得的 DNA 的量、纯度、片段大小等因素。载体的选择需要考虑载体的大小、载体拷贝数、插入片段的大小、所用宿主及筛选方法等因素^[7]。常用的克隆载体包括质粒(plasmid)、柯斯载体(cosmid)、福斯质粒(fosmid)和细菌人工染色体(bacterial artificial chromosome, BAC)等,可满足不同的插入片段大小要求。然而,小插入片段文库和大插入片段文库各有优劣^[9],应根据研究目的进行取舍。宿主的选择主要应考虑转化效率、重组载体在宿主细胞中的稳定性;若是需要表达外源基因,还应考虑表达效率、目标性状的筛选方法等^[7]。大肠杆菌(*Escherichia*

coli) 是土壤细菌基因或基因簇克隆和表达的通用宿主^[6]。利用穿梭载体或 BAC 文库可将大肠杆菌包含的文库信息转移至其他宿主如链霉菌 (*Streptomyces*) 或假单胞菌属 (*Pseudomonas*) 中^[13-14]。一些突变体也可作为宿主来对文库进行功能筛选,例如利用苜蓿根瘤菌 (*Sinorhizobium meliloti*) *bdhA* 突变体筛选聚 3-羟基丁酸酯 (poly-3-hydroxybutyrate) 分解基因^[15]。

2.2 宏基因组文库的筛选

土壤宏基因组具有高度的复杂性,需要高通量、高灵敏度的方法来筛选和鉴定文库中的有用基因或生物活性分子。宏基因组文库的筛选策略也可以分为序列驱动的和功能驱动的^[16]。

序列驱动的筛选是指根据已知的基因或基因表达产物的保守序列设计 PCR 引物或探针,通过 PCR 扩增或杂交筛选阳性克隆子,以获得某一类结构或功能类似的蛋白质中的新分子或其编码基因^[17-18]。序列分析法对鉴定新的基因有一定的局限性,但它的一个优点是不必依赖外源基因在宿主中的表达^[9]。

功能驱动的筛选主要通过化学或生物化学的手段从克隆文库中检测表达目标活性物质的转化子。绝大部分新发现的生物催化剂或活性化合物都是利用这种方式筛选得到的^[6]。目前,功能驱动的筛选主要发展出了三种策略^[16]。一种是直接检测具有特殊表型的克隆子。通过加入化学染料、不溶的或含发色团的底物等,制成选择性培养基,接种菌株后,阳性克隆产生的活性物质会引发显色反应或产生透明圈或荧光。此外,有些宿主菌株(通常是突变株)需要异源基因的功能互补才能在某种选择性条件下生长,也就是说,宿主细胞的生长,需要外源基因的转入并表达出相应的活性物质。第二种方法依据的就是这个原理。

第三种方法——诱导基因表达法 (induced gene expression)——则更为巧妙,并且发展出了不同的变体^[19],可以快速大量地筛选克隆文库。最早提出的是底物诱导基因表达法 (substrate-induced gene expression screening, SIGEX);代谢相关基因或酶基因往往在有底物存在的条件下才表达,反之则不表达,SIGEX 就是利用这个原理来筛选目的代谢基因^[20]。用于克隆宏基因组 DNA 的表达载体含有一个操纵子“陷阱”,即在克隆位点下游包含有一个 *gfp* (绿色荧光蛋白) 基因。这样,当土壤 DNA 片段连接到载体上并受底物诱导而表达时,其下游的 *gfp*

基因也一起表达,则筛选具有荧光信号的菌株即可得到阳性克隆^[16]。这个方法的优点在于高通量,而且不需要对底物进行修饰;而不足之处是转录调控元件可能被底物以外的其他因素激活,因而得到假阳性^[6, 21]。类似地,Williamson 等^[22]提出了代谢物调节表达法 (metabolite-regulated expression, METREX),该方法将宏基因组片段与生物感应器转化到同一个细胞中,此生物感应器可以检测到能引起群体感应的信号小分子,当宏基因组片段表达出的小分子量超过一定阈值,就可以诱导生物感应器中绿色荧光蛋白的表达,从而显示阳性克隆。2010 年,Uchiyama 和 Miyazaki^[23]报道了产物诱导基因表达法 (product-induced gene expression, PIGEX);其原理则是:阳性克隆生成的所欲筛得的酶可以将底物转化为特定产物,而该产物可以激活报告基因 (*gfp*) 的表达,再通过检测荧光信号而显示阳性。

与序列驱动的筛选方法不同,功能驱动的筛选不必依赖于已知的基因信息,因此可能得到新的基因和产物;但其受制于外源基因在宿主细胞中的表达,并且因检测手段的局限,筛选工作量大、效率低,往往分析大量克隆仅有几个具有活性^[24]。鉴于此,发展高通量、高灵敏度、快速简便的筛选方法或体系十分必要。虽然如此,利用这两类方法,已经得到了大量重要的生物活性物质^[6, 16, 25]。而在实际工作中,序列驱动和功能驱动也可以结合使用。例如,陈旭玉等^[26]先用特异的引物对土壤宏基因组文库进行 PCR 筛选,再进一步通过功能筛选(复筛),得到了两个对稻瘟菌有明显抑制作用的克隆子。此外,一些新的方法如微阵列技术 (microarray)、稳定性同位素标记技术 (stable isotope probing, SIP)、荧光原位杂交技术 (fluorescent in situ hybridization, FISH) 也已用于筛选和鉴定文库中的有用基因^[4, 7]。

3 宏基因组测序及数据分析

从环境样品中提取的 DNA 也可用于测序,以考察微生物群落结构、功能、代谢调节、进化及其与各种环境因子的关系。目前对宏基因组测序,采用的有传统的 Sanger/鸟枪法^[27-28]和新一代的测序技术^[29-30]。由于其高通量、快速、低成本的特点,第二代甚至第三代测序技术已经迅速发展起来^[31]并得到了越来越普遍的应用,传统的 Sanger 法在大规模测序上有被取代的趋势。下面主要介绍应用广泛

的第二代测序技术。

3.1 第二代测序技术原理

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 ABI 公司的 SOLiD 测序平台。测序过程可分为两步,即先进行模板的扩增再进行测序。大致步骤为:将片断化的基因组 DNA 两端连上特殊的接头,然后通过不同的方式将每个片段结合在微珠或芯片上,形成几百万个可以同时进行反应的微型反应池(有效反应体积 $10^{-9} \sim 10^{-12}$ L),每个单链 DNA 片段在其微型反应池中通过 PCR 扩增产

生大量拷贝形成单克隆“DNA 簇群”(Polony),以作为测序的模板;之后通过酶延伸反应或寡核苷酸连接反应同时对这几百万个微型反应池中的模板进行测序^[32-33]。三种测序平台在固定 PCR 克隆阵列的方式和后续的测序原理上略有不同,表现也各有优劣(表 1)。454 测序平台虽然测序成本较高,但读长长,在环境 DNA 样品的宏基因组测序中应用较多;Illumina Solexa 和 SOLiD 平台读长较短,但数据量大,性价比高,目前多用于基因组重建和深度再测序^[34-36];选择不同平台时应考虑到实验目的和实验室条件。

表 1 三种第二代测序平台比较

Table 1 Comparison of three second-generation sequencing platforms

平台 Platform	Roche 454	Illumina Solexa	ABI SOLiD	参考文献 Refs
上市时间 Time to market	2005 年	2007 年	2007 年	[32]
仪器价格 Price(10^4 dollars)	50	54	59.5	[37]
测序成本 Run cost(dollars Mb ⁻¹)	50	4	2.5	[38]
单次反应数据量 Gb per run	0.4	20	50	[32]
读长 Read length(bases)	400	50 × 2	50	[32, 37]
优点 Pros	读长长,运行速度快	性价比高,是目前应用最广泛的测序平台	准确度高	[32, 37]
缺点 Cons	测序成本高,均聚物重复序列区(homopolymer repeats)错误率较高	适应样品种类较少	运行时间长	[37]

3.2 第二代测序数据处理

相对传统的 Sanger 法,第二代测序技术不需要构建克隆文库,可以获得大量的序列数据,但是现期第二代测序主要的问题也就在于数据处理困难。由于序列片段较短,环境样品通常具有高度的复杂性,很难确定某条序列来自于哪个物种,并且提取的土壤 DNA 中也许并没有包含该物种的全部基因组,或者可参考的数据库中该物种的信息不全甚至没有(如新的物种),因此面对海量的短序列,拼接和进一步的分析具有相当大的挑战^[2, 33, 39]。这就需要相应的生物信息学的方法和软件以及配套的计算能力有所突破。一般而言,宏基因组的数据处理包括序列拼接和基因征集(gene calling),接下来分析生物多样性,进行基因注释;在此基础上,可以获得一些重要的宏基因组信息,如序列组成(GC 含量、基因组大小等)、物种组成、功能组成和群落特

征等^[33, 39]。

宏基因组数据分类分析的一个关键步骤是根据相似度或序列组成对序列进行归类(bin-ning)^[19, 33],即将样品序列归到正确的类或 OTU,以与其来源联系起来。基于相似度的方法将序列与数据库比较,然后依据相似度来对序列归类。相关的生物信息学工具有 Metagenome Analyzer (MEGAN)^[40-41]、CARMA^[42]、Phymm^[43]以及 SOrt-ITEMS^[44]等。这类方法的一个缺陷是依赖于现有的数据库,而环境中绝大多数微生物的序列仍是未知,因此宏基因组数据中最多可达 90% 的序列由于缺乏参考序列而不能被鉴定^[41]。基于组成的方法则是分析序列自身的特征,如 GC 含量、密码子使用率或寡核苷酸频率。相关工具有 PhyloPythia^[45]、TETRA^[46]、TACO^[47]、GSOM^[48]和 S-GSOM^[49]等。

由于前述种种困难,目前直接对土壤宏基因组

测序的报道鲜见,主要的障碍在于序列的拼接;而针对特定目标片段(如 SSU rRNA 基因、某些功能基因)的大规模测序,已经有了相当的研究^[30, 50-52]。以 16S rDNA 序列分析为例,拿到高通量测序数据后,先去除低质量的序列,再计算 OUT,进行相似性搜索,分析群落组成(优势种、丰度、 α -多样性等);若是比较不同环境样品,则可计算 β -多样性,进行主成分分析(PCA)、聚类分析等,考察环境因子对微生物群落的影响。

比较不同宏基因组的相关指标有助于了解微生物群落与无机环境间的相互关系。对宏基因组的比较最早见于 Tringe 等的报道^[53],研究指出,从不同环境中获得的微生物群落的鸟枪法测序序列,经过比较分析后表现出生境特异的指纹图谱,这为解释和诊断各种环境因素提供了新的思路。随后,由于宏基因组数据的持续积累以及算法和软件的发展,相关报道迅速增加^[41, 54-55]。最近,一些基于网络的宏基因组注释平台,如 metagenomics RAST server^[56]、IMG/M server^[57]以及 METAREP^[58],也相继被开发出来。将上传的宏基因组数据与已有蛋白质和核苷酸数据库进行比对,可以分析不同环境中微生物群落的结构和功能^[19]。

4 基于宏基因组的微生物生态学分析

筛选克隆文库,主要目的是获得一些新颖的生物活性物质或基因以用于生产实践,在此不做详述。土壤微生物生态学研究的是微生物群落的结构与功能,及其与无机环境间的相互影响,采用的则是对 DNA 序列进行获取和分析的方法。基于宏基因组的微生物生态学研究手段主要可以分为三类^[2],分述如下。

4.1 16S/18S rDNA 或其他标记基因调查

16S 和 18S rDNA 分别编码原核生物和真核生物核糖体小亚基 RNA 分子,由于其高度的保守性,长期以来被用于分析物种间的亲缘关系和物种多样性^[59]。通常先提取土壤总 DNA,再利用通用引物扩增 16S 或 18S rDNA,建库后测序,根据 16S 或 18S rDNA 序列分析系统发育关系和生物多样性。或者采用指纹图谱的方法,如变性梯度凝胶电泳(denaturing gradient gel electrophoresis, DGGE)和末端限制性片段长度多态分析(terminal restriction fragment length polymorphism, T-RFLP)等。例如,He 等分别利用 16S 和 18S rRNA 基因调查了澳大利

亚森林土壤中细菌和真菌群落^[60-61]。但是,也有人认为将 rDNA 作为判断 OTU 的标记基因存在缺陷:rDNA 会发生水平基因转移,并且细菌内 16S rDNA 的拷贝数会有不同,这就会导致对群落中个体数目或 OTU 中成员数目的计算产生偏差^[33]。因此,一些功能基因也被提出作为标记基因替代或补充 rDNA 分析。常用的功能基因包括 RNA 聚合酶 β 亚基基因(*rpoB*)以及参与碳、氮转化的基因如甲烷氧化酶基因(*pmoA*)、氨单加氧酶基因(*amoA*)、亚硝酸盐还原酶基因(*nirS/nirK*)和光合基因(*pufM*)等^[33]。

上述建库或指纹图谱的方法相对费时费力,并且覆盖的物种有限。新一代的测序技术的出现,可以绕过建库的过程,直接对扩增出的土壤 16S 或 18S rDNA 进行测序,使大规模的土壤微生物多样性调查得以实现。

4.2 大尺度 DNA 测序

大尺度 DNA 测序既可以是获得大规模的 16S/18S rDNA 或其他特定功能基因的序列;也可以是将所得到的环境 DNA 全部测序,以考察群落的物种组成和基因组成,明确群落的结构和功能及其与环境间相互关系。

最早的大规模测序工作是 Venter 等^[28]对 Sargasso Sea 环境基因组的研究。应用鸟枪法对 Sargasso Sea 表层水样的基因组文库进行测序,研究人员发现了大量新的基因和物种,大大扩展了对海洋微生物多样性的认识。值得一提的是,在这些序列里,发现了一个古菌的氨单加氧酶基因(*amoA*),这直接开启了一个新的研究方向,即对氨氧化古菌的研究^[62-64]。随后,Tyson 等^[27]同样应用鸟枪法对一个群落结构相对简单的嗜酸生物膜中微生物种群结构和代谢进行了调查,为了解该微生物群落在极端环境中的生存机制提供了重要参考。2007 年至 2008 年间进行的海洋宏基因组学调查(the *Sorcerer II Expedition*)也产生了丰富的数据和研究成果,大大促进了海洋宏基因组学的发展^[65]。此外,宏基因组学在研究微生物与宿主间的关系以及环境病毒学方面也得到了良好的应用^[33]。

由于时空变异以及管理方式的差别,土壤从微观尺度到宏观尺度都存在着巨大的异质性,相应的土壤微生物群落也迥然不同^[66]。Buée 等^[30]利用第二代测序技术(454 焦磷酸测序)对森林土壤进行调查,揭示出了极高的真菌生物多样性。研究者先从

6片具有不同植被类型的林区取样,每个地点采集8个土样。分别对这48个土样用试剂盒提取DNA后,将每个DNA产物稀释两个倍数,获得96个样品,用带有寡核苷酸标记的引物分别对其核糖体第一内转录间隔区(ITS-1)进行扩增,然后将96个PCR产物根据6个林区合并成6个扩增子文库,再将它们等摩尔混合后在454平台上进行焦磷酸测序,因为不同的样品PCR时,引物中带的寡核苷酸标记不同,因此测序后可以将来自不同地点的序列区分开来。接下来是进行序列分析。得到的原始数据经过质量筛选后,最终产生166 350条序列,平均读长为252 bp。确定OTU后,作者进一步建立稀释曲线、计算多样性和丰富度指标。同时将序列分类结果用MEGAN建树,分析系统发育关系。结果表明81%的OTUs属于子囊菌门(Ascomycota)和担子菌门(Basidiomycota),其中26个最丰富的OTU囊括了73%的ITS-1序列。不同植被类型的样点,真菌的组成在门水平差异显著,而在科、属、种的水平比较类似,反映出真菌群落结构随环境变化的精细结构。该研究体现了高效的454测序在研究森林生态系统真菌群落时空动态中的应用。

对宏基因组的分析,可以有两种角度:一是考察一个群落的成员微生物(基因组),即以基因组为中心的分析(genome-centric analysis);另一种是以基因为中心的分析(gene-centric analysis),即从组成该群落的基因入手。前者将宏基因组序列拼接成完整或接近完整的基因组,进而研究群落中的微生物组成及其多样性和功能;后者考察群落中包含哪些基因以阐明群落的功能或其对环境的响应,在某个群落中出现频率更高的基因通常对该群落具有积极的作用,例如表层海水中变形菌视紫质基因的分布较在其他生境中广泛,白蚁后肠中降解纤维素的酶的基因较其他生境多^[2,5]。

4.3 杂交或微阵列方法

通过设计特定的探针与环境样品中的核酸进行杂交来考察其中的微生物或感兴趣的基因也是宏基因组学中一种重要的研究思路,下面主要介绍荧光原位杂交技术和基因芯片技术。荧光原位杂交技术是在原位杂交技术(in situ hybridization, ISH)的基础上发展而来的,其基本原理是用荧光标记的寡核苷酸探针特异地和互补核酸序列(16S/18S rDNA或其他基因)结合,然后通过检测荧光信号可以对目标DNA进行定性或定量分析,以反映微

生物组成及功能。FISH的具体过程包括以下几步:①样品固定;②预处理样品;③用相应的探针进行杂交;④洗掉未结合的探针;⑤封固、呈像及结果分析^[67]。在应用过程中应注意探针和标记方法的设计以及荧光染料的选择。由于FISH技术具有快速、准确、原位等优点,已在微生物生态学各个领域研究中得到了广泛的应用。张伟等^[68]利用16S rRNA特异性探针,结合DAPI全细胞染色和样品稀释,采用荧光原位杂交技术分析喀斯特山地土壤硫酸盐还原菌(SRB)的数量和空间原位分布状况。结果显示所研究土壤剖面各层均有SRB检出,平均为 $(2.7 \pm 1.2) \times 10^7$ 个 g^{-1} 干土,并且SRB沿剖面深度有一定变化。表明FISH能同时对土壤中的SRB进行定性和定量分析,是一种快速有效的检测手段。

微阵列技术,又称基因芯片,是一种重要的生物芯片。它也是采用核酸分子杂交的工作原理,应用已知核酸序列作为探针与互补的靶核苷酸序列杂交,然后通过信号检测对样品(如宏基因组)进行定性或定量分析^[69]。基因芯片采用集约化和平面处理方法,在微小片基上高密度而有序地排列大量基因片段或核苷酸片段,形成DNA微矩阵,从而能够一次分析大量的基因,快捷地探测环境微生物基因组,了解环境样品中微生物群落结构及其基因表达图谱和代谢途径^[4]。大体过程是先通过自动化系统将高密度的核酸样品(cDNA或寡核苷酸等)点涂或压印在载体(或称芯片,通常为显微镜载玻片或滤膜)上进行固定,再将适当来源的带有荧光标记的复合核酸混合物(探针)与制好的微阵列进行杂交,然后用激光扫描仪检测荧光标记、用数字成像软件分析每个点发出的信号即可获得基因表达谱,将实验样品的基因表达谱与对照组比较就可以进行差异分析^[70]。2007年,He等^[71]报道了首个用于研究微生物群落功能活性和生物地球化学循环微生物过程的广谱基因芯片——GeoChip;最新的版本GeoChip 3.0含有约28 000个探针,涵盖了292个基因家族中约57 000个基因,而这些基因涉及碳、氮、磷、硫循环,能量代谢、抗生素抗性、重金属抗性和有机污染物降解等功能^[72]。利用GeoChip,研究人员对多种生境中的微生物群落进行了调查^[73-75],证明其是一种强大的、高通量的宏基因组学分析工具。但是,基因芯片灵敏度较低^[76],要加强对土壤低丰度微生物的检测能力,必须提高微阵列技术的灵敏性和特异性。

5 宏基因组学在土壤微生物生态学中的应用与展望

如前所述,环境存在大量未知或不可培养的微生物,宏基因组学由于直接研究环境中的总 DNA,为开发新的生物活性物质、发现新的基因和物种、研究微生物群落结构与功能、微生物对环境变化的响应与反馈、微生物群落演替与进化、微生物区域分布与生物地理学、土壤肥力形成与变化机理以及恢复生态学等开辟了一条新的途径,为解释和解决一些重大农业和环境问题提供重要依据。

由于极高的重复率,利用传统的纯培养技术筛选到新的生物活性物质的机率显著下降^[24];而宏基因组学采用的是非培养的方法,因此突破了该瓶颈,一系列新的生物活性物质(酶和抗生素等)利用该方法得以发现。在发现新的基因和物种、并进一步调查微生物群落结构和功能方面,宏基因组学也取得了显著进展。对宏基因组学数据的比较则为揭示微生物群落与环境因子之间的相互关系提供了有力的途径。

在利用宏基因组学进行微生物生态学研究时,有必要充分结合生态学尤其是群落生态学和系统生态学的相关理论和方法。例如沿着一系列环境梯度,如温度、降水,进行土壤宏基因组学的调查,有助于揭示环境因子乃至气候变化对微生物的影响以及微生物对此作出的响应。根际微生物对植物生长具有重要意义,将植物生长状况与根际微生物宏基因组状况联系起来可以更好地理解植物-微生物相互关系。群落生态学中的种间关联或群落排序分析或许可以用于考察生物地球化学循环中元素耦合的问题,例如碳氮耦合的关键在于生物的碳氮比(C:N)^[77],弄清楚宏基因组中碳循环相关基因(如参与甲烷代谢的基因)与氮循环相关基因(如硝化/反硝化相关基因)的关联(比例及其变化),有助于了解碳氮循环耦合的强弱,解释生物群落间碳氮比变化并进一步厘清元素循环的库和流。

通过全面系统地调查具有代表性的土壤样品,建立典型土壤微生物基因库数据集,有利于了解土壤微生物的组成(包括代表类群、特征物种、各成员之间的比例等)、可能的功能及代谢网络,以及这些要素随环境的变化。在此基础上,可以进一步探讨不同类型微生物群落的演化特点和驱动因子,阐明元素循环及其耦合的微观机理。应用于生产实践

上,相关理论和进展则有利于我们优化微生物功能的执行效率、提高作物养分供给并进一步减少能耗和对生态环境的负面影响。国外在土壤宏基因组学方面已经开展了有力的研究^[2, 78],而我国还只有初步和零星的报道。鉴于其巨大的科学意义和应用前景,笔者建议有关部门将土壤宏基因组学的研究纳入国家重点研究计划:通过启动相关重大项目,铺开若干关系学科发展和国民生产的研究方向,培养具有竞争力的宏基因组学研究队伍;从国家层面上开展协作攻关,充分探索和发掘我国丰富的微生物资源,取得具有原创性和自主知识产权的重要研究成果;加强宏基因组学领域的国际交流与合作,以求在日益激烈的全球科技竞争中占有一席之地,为微生物生态学及系统生物学学科发展做出有影响力的贡献,并为解决日益严重全球生态和环境问题发挥自己应有的力量。

参考文献

- [1] Whitman W B, Coleman D C, Wiebe W J. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, 95(12): 6 578—6 583
- [2] The National Academies. *The new science of metagenomics: revealing the secrets of our microbial planet*. Washington DC: National Academies Press, 2007
- [3] Handelsman J, Rondon M R, Brady S F, et al. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry & Biology*, 1998, 5(10): R245—R249
- [4] 贺纪正, 张丽梅, 沈菊培, 等. 宏基因组学(Metagenomics)的研究现状和发展趋势. *环境科学学报*, 2008, 28(2): 209—218. He J Z, Zhang L M, Shen J P, et al. Advances and perspectives of metagenomics (In Chinese). *Acta Scientiae Circumstantiae*, 2008, 28(2): 209—218
- [5] Hugenholtz P, Tyson G W. Microbiology: Metagenomics. *Nature*, 2008, 455(7212): 481—483
- [6] 沈菊培, 张丽梅, 郑袁明, 等. 土壤宏基因组学技术及其应用. *应用生态学报*, 2007, 18(1): 212—218. Shen J P, Zhang L M, Zheng Y M, et al. Methodology and application of soil metagenomics (In Chinese). *Chinese Journal of Applied Ecology*, 2007, 18(1): 212—218
- [7] 钮旭光, 韩梅, 韩晓日. 宏基因组学: 土壤微生物研究的新策略. *微生物学通报*, 2007, 34(3): 576—579. Niu X G, Han M, Han X R. Metagenomics: New strategy for soil microbiology research (In Chinese). *Microbiology*, 2007, 34(3): 576—579
- [8] Robe P, Nalin R, Capellano C, et al. Extraction of DNA from soil. *European Journal of Soil Biology*, 2003, 39(4): 183—190
- [9] Daniel R. The metagenomics of soil. *Nature Reviews Microbiology*, 2005, 3(6): 470—478

- [10] Rajendhran J, Gunasekaran P. Strategies for accessing soil metagenome for desired applications. *Biotechnology Advances*, 2008, 26(6): 576—590
- [11] He J Z, Xu Z H, Hughes J. Pre-lysis washing improves DNA extraction from a forest soil. *Soil Biology and Biochemistry*, 2005, 37(12): 2 337—2 341
- [12] Zhang L M, Liu F, Tan W F, et al. Microbial DNA extraction and analyses of soil iron-manganese nodules. *Soil Biology and Biochemistry*, 2008, 40(6): 1 364—1 369
- [13] Martinez A, Kolvek S J, Yip C L, et al. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Applied and Environmental Microbiology*, 2004, 70(4): 2 452—2 463
- [14] Courtois S, Cappellano C M, Ball M, et al. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology*, 2003, 69(1): 49—55
- [15] Wang C, Meek D J, Panchal P, et al. Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Applied and Environmental Microbiology*, 2006, 72(1): 384—391
- [16] Simon C, Daniel R. Achievements and new knowledge unraveled by metagenomic approaches. *Applied Microbiology and Biotechnology*, 2009, 85(2): 265—276
- [17] Knietzsch A, Bowien S, Whited G, et al. Identification and characterization of coenzyme B12-dependent glycerol dehydratase and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Applied and Environmental Microbiology*, 2003, 69(6): 3 048—3 060
- [18] Daniel R. The soil metagenome—a rich resource for the discovery of novel natural products. *Current Opinion in Biotechnology*, 2004, 15(3): 199—204
- [19] Simon C, Daniel R. Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology*, 2011, 77(4): 1 153—1 161
- [20] Uchiyama T, Abe T, Ikemura T, et al. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*, 2005, 23(1): 88—93
- [21] Galvao T C, Mohn W W, de Lorenzo V. Exploring the microbial biodegradation and biotransformation gene pool. *Trends in Biotechnology*, 2005, 23(10): 497—506
- [22] Williamson L L, Borlee B R, Schloss P D, et al. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Applied and Environmental Microbiology*, 2005, 71(10): 6 335—6 344
- [23] Uchiyama T, Miyazaki K. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Applied and Environmental Microbiology*, 2010, 76(21): 7 029—7 035
- [24] 冯美琴. 宏基因组学的研究进展. *安徽农业科学*, 2008, 36(2): 415—416, 479. Feng M Q. Research progress of metagenomics (In Chinese). *Journal of Anhui Agricultural Sciences*, 2008, 36(2): 415—416, 479
- [25] Ferrer M, Beloqui A, Timmis K N, et al. Metagenomics for mining new genetic resources of microbial communities. *Journal of Molecular Microbiology and Biotechnology*, 2009, 16(1/2): 109—123
- [26] 陈旭玉, 周亚奎, 余贤美, 等. 土壤宏基因组文库的构建及拮抗稻瘟菌克隆子的筛选. *西北农业学报*, 2009, 18(1): 165—169. Chen X Y, Zhou Y K, Yu X M, et al. Construction of metagenomic library from soil and screening of clones with antagonism to the rice blast fungus (*Magnaporthe grisea*) (In Chinese). *Acta Agriculturae Boreali-Occidentalis Sinica*, 2009, 18(1): 165—169
- [27] Tyson G W, Chapman J, Hugenholtz P, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 2004, 428(6978): 37—43
- [28] Venter J C, Remington K, Heidelberg J F, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 2004, 304(5667): 66—74
- [29] Leininger S, Urich T, Schloter M, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, 2006, 442(7104): 806—809
- [30] Buée M, Reich M, Murat C, et al. 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist*, 2009, 184(2): 449—456
- [31] Rusk N. Cheap third-generation sequencing. *Nature Methods*, 2009, 6(4): 244—245
- [32] 解增言, 林俊华, 谭军, 等. DNA 测序技术的发展历史与最新进展. *生物技术通报*, 2010(8): 64—70. Xie Z Y, Lin J H, Tan J, et al. The history and advances of DNA sequencing technology (In Chinese). *Biotechnology Bulletin*, 2010(8): 64—70
- [33] Wooley J C, Godzik A, Friedberg I. A primer on metagenomics. *PLoS Computational Biology*, 2010, 6(2): e1000667. doi:10.1371/journal.pcbi.1000667
- [34] Melum E, May S, Schilhabel M B, et al. SNP discovery performance of two second-generation sequencing platforms in the NOD2 gene region. *Human Mutation*, 2010, 31(7): 875—885
- [35] Walter N, Bottomly D, Laderas T, et al. High throughput sequencing in mice: A platform comparison identifies a preponderance of cryptic SNPs. *BMC Genomics*, 2009, 10(1): 379
- [36] Marklund S, Carlborg O. SNP detection and prediction of variability between chicken lines using genome resequencing of DNA pools. *BMC Genomics*, 2010, 11(1): 665
- [37] Metzker M L. Sequencing technologies—The next generation. *Nature Reviews Genetics*, 2010, 11(1): 31—46
- [38] 张闻, 何永蜀, 陈元晓. 新一代 DNA 测序技术. *国际遗传学杂志*, 2009, 32(5): 341—344. Zhang W, He Y S, Chen Y X. New-generation DNA sequencing technology (In Chinese). *International Journal of Genetics*, 2009, 32(5): 341—344
- [39] Raes J, Foerstner K U, Bork P. Get the most out of your metagenome: Computational analysis of environmental sequence data.

- Current Opinion in Microbiology, 2007, 10(5): 490—498
- [40] Huson D H, Auch A F, Qi J, et al. MEGAN analysis of metagenomic data. *Genome Research*, 2007, 17(3): 377—386
- [41] Huson D, Richter D, Mitra S, et al. Methods for comparative metagenomics. *BMC Bioinformatics*, 2009, 10(Suppl 1): S12
- [42] Krause L, Diaz N N, Goesmann A, et al. Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Research*, 2008, 36(7): 2 230—2 239
- [43] Brady A, Salzberg S L. Phymm and PhymmBL: Metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods*, 2009, 6(9): 673—676
- [44] Monzoorul Haque M, Ghosh T S, Komanduri D, et al. SOrT-ITEMS: Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics*, 2009, 25(14): 1 722—1 730
- [45] McHardy A C, Martin H G, Tsirigos A, et al. Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods*, 2007, 4(1): 63—72
- [46] Teeling H, Waldmann J, Lombardot T, et al. TETRA: A web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics*, 2004, 5: 163
- [47] Diaz N N, Krause L, Goesmann A, et al. TACO: Taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. *BMC Bioinformatics*, 2009, 10: 56
- [48] Chan C-K K, Hsu A L, Tang S-L, et al. Using growing self-organising maps to improve the binning process in environmental whole-genome shotgun sequencing. *Journal of Biomedicine & Biotechnology*, 2008: Article No. : 513 701
- [49] Chan C K K, Hsu A L, Halgamuge S K, et al. Binning sequences using very sparse labels within a metagenome. *BMC Bioinformatics*, 2008, 9: 215
- [50] Acosta-Martinez V, Dowd S, Sun Y, et al. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biology and Biochemistry*, 2008, 40(11): 2 762—2 770
- [51] Lee O O, Wang Y, Yang J, et al. Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. *ISME Journal*, 2011, 5(4): 650—664
- [52] Kip N, Dutilh B E, Pan Y, et al. Ultra-deep pyrosequencing of *pmoA* amplicons confirms the prevalence of *Methylomonas* and *Methylocystis* in Sphagnum mosses from a Dutch peat bog. *Environmental Microbiology Reports*, 2011, doi:10.1111/j.1758—2 229.2011.00260.x
- [53] Tringe S, von Mering C, Kobayashi A, et al. Comparative metagenomics of microbial communities. *Science*, 2005, 308: 554—557
- [54] Rodriguez-Brito B, Rohwer F, Edwards R. An application of statistics to comparative metagenomics. *BMC Bioinformatics*, 2006, 7:162
- [55] Mitra S, Gilbert J A, Field D, et al. Comparison of multiple metagenomes using phylogenetic networks based on ecological indices. *ISME Journal*, 2010, 4(10): 1 236—1 242
- [56] Meyer F, Paarmann D, D'Souza M, et al. The metagenomics RAST server-A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 2008, 9: 386
- [57] Markowitz V M, Ivanova N N, Szeto E, et al. IMG/M: A data management and analysis system for metagenomes. *Nucleic Acids Research*, 2008, 36 (Database issue): D534—538
- [58] Goll J, Rusch D B, Tanenbaum D M, et al. METAREP: JCVI metagenomics reports-An open source tool for high-performance comparative metagenomics. *Bioinformatics*, 2010, 26 (20): 2 631—2 632
- [59] Zvolinski M D. DNA sequencing: Strategies for soil microbiology. *Soil Science Society of America Journal*, 2007, 71(2): 592—600
- [60] He J Z, Xu Z H, Hughes J. Molecular bacterial diversity of a forest soil under residue management regimes in subtropical Australia. *FEMS Microbiology Ecology*, 2006, 55(1): 38—47
- [61] He J Z, Xu Z H, Hughes J. Analyses of soil fungal communities in adjacent natural forest and hoop pine plantation ecosystems of subtropical Australia using molecular approaches based on 18S rRNA genes. *FEMS Microbiology Letters*, 2005, 247(1): 91—100
- [62] Könneke M, Bernhard A, De la Torre J, et al. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, 2005, 437(7058): 543—546
- [63] He J Z, Shen J P, Zhang L M, et al. Quantitative analyses of the abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea of a Chinese upland red soil under long-term fertilization practices. *Environmental Microbiology*, 2007, 9(9): 2 364—2 374
- [64] Shen J P, Zhang L M, Zhu Y G, et al. Abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea communities of an alkaline sandy loam. *Environmental Microbiology*, 2008, 10(6): 1 601—1 611
- [65] PLoS biology: Ocean metagenomics collection. Available from: <http://www.ploscollections.org/article/browseIssue.action;jsessionid=8C418C12C5434557CCD29C4B006A4C37.amba02?issue=info%3Adoi%2F10.1371%2Fissue.pcol.v06.i02>. Cited on 19th, May, 2011
- [66] Singh B K, Campbell C D, Sorenson S J, et al. Soil genomics. *Nature Reviews Microbiology*, 2009, 7(10): 756
- [67] 呼庆, 齐鸿雁, 张洪勋. 荧光原位杂交技术及其在微生物生态学中的应用. *生态学报*, 2004, 24(5): 1 048—1 054. Hu Q, Qi H Y, Zhang H X. Fluorescence in situ hybridization (FISH) and its applications in microbial ecology (In Chinese). *Acta Ecologica Sinica*, 2004, 24(5): 1 048—1 054
- [68] 张伟, 刘丛强, 刘涛泽, 等. 荧光原位杂交在喀斯特山地土壤硫酸盐还原菌检测中的应用. *微生物学通报*, 2008, 35(8): 1 273—1 277. Zhang W, Liu C Q, Liu T Z, et al. Application of fluorescence in situ hybridization to detecting Karst mountainous soil sulfate-reducing bacterium (In Chinese). *Microbiology*, 2008, 35(8): 1 273—1 277

- [69] 金敏, 李君文. 基因芯片技术在环境微生物群落研究中的应用. 微生物学通报, 2008, 35(9): 1 466—1 471. Jin M, Li J W. Microarray application in environmental microbial community research (In Chinese). Microbiology, 2008, 35(9): 1 466—1 471
- [70] 萨姆布鲁克 J, 拉塞尔 D W. 黄培堂. 译. 分子克隆实验指南. 北京: 科学出版社, 2002: 1 775. Sambrook J, Russell D W. Molecular cloning: A laboratory manual. New York: Cold Spring Harbor Laboratory Press, 2001: A10. 2
- [71] He Z, Gentry T J, Schadt C W, et al. GeoChip: A comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME Journal, 2007, 1(1): 67—77
- [72] He Z, Deng Y, Van Nostrand J D, et al. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. ISME Journal, 2010, 4(9): 1 167—1 179
- [73] Van Nostrand J D, Wu W-M, Wu L, et al. GeoChip-based analysis of functional microbial communities during the reoxidation of a bioreduced uranium-contaminated aquifer. Environmental Microbiology, 2009, 11(10): 2 611—2 626
- [74] Xie J, He Z, Liu X, et al. GeoChip-based analysis of the functional gene diversity and metabolic potential of microbial communities in acid mine drainage. Applied and Environmental Microbiology, 2011, 77(3): 991—999
- [75] Wang F, Zhou H, Meng J, et al. GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. Proceedings of the National Academy of Sciences, 2009, 106(12): 4 840—4 845
- [76] Zhou J, Thompson D. Challenges in applying microarrays to environmental studies. Current Opinion in Biotechnology, 2002, 13(3): 204—207
- [77] Gruber N, Galloway J. An earth-system perspective of the global nitrogen cycle. Nature, 2008, 451(7176): 293—296
- [78] Vogel T M, Simonet P, Jansson J K, et al. TerraGenome: A consortium for the sequencing of a soil metagenome. Nature Reviews Microbiology, 2009, 7(4): 252—252

METHODS FOR AND PROGRESS IN RESEARCH ON SOIL METAGENOMICS

He Jizheng^{1†} Yuan Chaolei^{1,2} Shen Jupei¹ Zhang Limei¹

(1 Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China)

(2 Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract Soil microorganisms are a driving force in material recycling and nutrient transformation in soil. However, for a long time, soil has been treated as a “black box” system, wherein microbial diversity and biochemical processes remain to be explored. Since most of the soil microorganisms are still quite hard to be isolated for culture, traditional culture methods are quite limited in helping reveal compositions and functions of soil microbial communities. The metagenomic method is able to explore the structures and functions (sequence-driven approach) of soil microbial community and to screen bioactive materials and new genes (function-driven approach) through extracting all microbial DNAs direct from environment samples and then sequencing or constructing clone library, thus breaking through the bottleneck of the traditional methods and greatly enriching the knowledge about soil microbial biodiversity and functions. While reviewing main procedures of the metagenomic technique, the paper focuses on the introduction to application of the next-generation sequencing (NGS) technologies in metagenomic research and processing of the huge volume of data it may produce. The new process on soil microbial ecology with the metagenomic technique is then discussed. And in the end, the authors propose that research projecjes on soil metagenomics should be launched at the national level to explore soil microorganism communities and their variation, so as to contribute to the causes of bioresource exploitation, agricultural production and environment protection.

Key words Metagenomics; Research methods; Next-generation sequencing; Microbial ecology; Soil environment