

# 元基因组用于未培养微生物的研究 Metagenomics research on uncultured microorganisms

李富超, 周 婕, 秦 松

(中国科学院 海洋研究所, 山东 青岛 266071)

中图分类号: Q938

文献标识码: A

文章编号: 1000-3096(2010)05-0079-04

微生物学 100 多年的研究发展表明, 目前能够在实验室纯培养的微生物仅占其总数的不到 1%, 绝大多数微生物不可培养, 并且代谢功能未知。早在 20 世纪 80 年代, Lane 等<sup>[1]</sup>认为简单的形态学、生理性状对微生物鉴定提供的信息很少, 应重视非传统方法在微生物研究中的应用。随着人们对不可培养微生物的认识, 针对多种特殊环境(如海洋及沉积物、深海热液及冷泉、土壤、动物瘤胃及内脏、人肠道等)的微生物群落多样性, 建立了基于微生物基因组学和分子生态学(如 16S rDNA, ITS, recA, psbA 等分子标志)基础上的微生物群体基因组分析方法, 即元基因组技术, 通过直接从环境中分离基因组 DNA, 构建 BAC、Fosmid 及 Cosmid 等大片段基因组文库, 以及基于 16S rDNA 及其他特殊分子标志的小片段基因组文库。元基因组技术实现了对不同环境中不可培养微生物基因组进化的研究, 为分析特殊环境微生物多样性及其群落功能提供了新的思路。作者简要介绍了元基因组研究的常用方法和在微生物生态学及生物技术领域的应用, 尤其是在海洋微生物研究中的应用。

## 1 元基因组的研究方法

在 1986 年, Pace 等<sup>[2]</sup>首先提出了从环境样品中直接进行基因克隆, 并且从环境样品中克隆获得大片段的 DNA<sup>[3]</sup>。Healy 等<sup>[4]</sup>从木质纤维素厌氧分解池, 提取混合微生物总 DNA, 构建了 DNA 文库(zoolibraries), 从中筛选得到水解纤维素酶活性的阳性克隆。Stein 等<sup>[5]</sup>通过对海洋浮游生物元的基因组文库分析, 鉴定了一个 38.5kb 的 fosmid 克隆, 这个克隆包含了未培养古菌的核糖体小亚基的 DNA 片段。

元基因组(metagenomics)的概念是由 Handelsman 等<sup>[6]</sup>在 1998 年提出的, 最初是用来定义土壤细菌的混合基因组, 后来发展为从各种环境(如海洋、土壤、动物瘤胃及肠道等)中直接分离基因组总 DNA, 构建基因组文库, 进行序列和功能性分析, 以解析特殊环境微生物群落的多样性及结构功能关系<sup>[7]</sup>(图 1)。

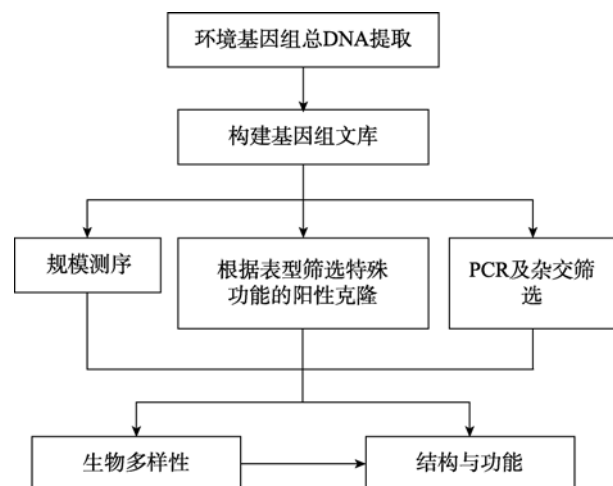


图 1 元基因组文库的建立和筛选策略

常用的构建元基因组文库的载体有 cosmid, fosmid, BAC 等, cosmid 是一种人工构建的载体, 包含  $\lambda$ -噬菌体的 cos 基因, cosmid 载体可以被包装入  $\lambda$ -噬菌体颗粒中, 用于感染大肠杆菌, 它和 fosmid 的插入片段可达 40kb 左右, 但是稳定性不好。

收稿日期: 2008-11-10; 修回日期: 2009-01-12

基金项目: 中国科学院知识创新工程项目(KSCX2-YW-G-022, 2007-1, KZCX2-YW-216); 中国大洋协会项目(DYXM-115-02-2-17)

作者简介: 李富超(1977-), 男, 山东泰安人, 硕士, 主要从事海洋微生物研究, E-mail: lifuchao@ms.qdio.ac.cn; 秦松, 通信作者, 电话: 0532-82898500, E-mail: sqin@ms.qdio.ac.cn

Fosmid 包含大肠杆菌 F 因子复制子的单一拷贝,是在 cosmid 的基础上进一步改进而成的,稳定性较其他高复制载体要好,如 pCC1FOS,包含了大肠杆菌 F 因子的单拷贝复制起点,又含有多拷贝的可诱导复制起点 oriV,从而保证了克隆的稳定性和高产量。BAC(在大肠杆菌中复制),包含了大肠杆菌 F 因子单一拷贝的复制起始位点,每个细胞中严格含有 1~2 个复制子,和高拷贝载体相比它含有的克隆数更稳定,平均插入片段为 80 kb,更大的插入片段可达到 150 kb<sup>[8]</sup>。

### 1.1 序列分析

包括随机测序和利用系统进化标签对阳性克隆进行测序推断该 DNA 片段的可能来源。目标基因一旦确定,可以在该基因的侧翼 DNA(flanking DNA)序列中寻找相关系统进化标签,从而将系统进化和功能基因联系起来。通过系统进化标签所指导的序列分析方法最初是由 Stein 等<sup>[5]</sup>提出的,Stein 等首先获得了与来自未培养古菌 16S rDNA 相连的基因序列,又根据该序列确定了和  $\alpha$ -变形细菌(Proteobacteria)相关的一个插入片段,侧翼 DNA 序列揭示了它是一个细菌视紫红质样的基因,基因产物是一个自动的光感受器,这个发现暗示了细菌视紫红质基因并不局限于古菌,实际上在海洋变形细菌中广泛存在<sup>[9]</sup>。

另一种序列分析方法是随机克隆测序,这种方法可以获知一个种群中功能的分布和冗余,特性的关联,基因组结构,以及基因水平转移等信息,这种方法对于大尺度群落分析特别有效。马尾藻海(Sargasso Sea)和美国黄铁矿污水(Acid mine drainage, AMD)群落中不可培养生物基因组的重建是近期具有里程碑意义的测序研究,验证了大范围测序的有效性,也丰富了人们对未知群落的认识。这些研究还将系统进化及功能联系起来,重建或部分重建了一些未培养生物的基因组。

### 1.2 功能分析

建立元基因组文库,对克隆获得的表型进行功能分析,已经鉴定了许多新的抗生素<sup>[10]</sup>、抗性基因<sup>[11]</sup>以及各种特殊的酶<sup>[4,11]</sup>等。功能分析有别于序列分析,更有可能发现新基因,获得新功能。但克隆的大多数基因往往因为不能在其他宿主细菌中表达,而无法被检测到。

在元基因组中通过寻找合适的表型来增加获得

阳性克隆的机率, Majernik 等<sup>[12]</sup>设计了一个巧妙的筛选  $\text{Na}^+(\text{Li}^+)/\text{H}^+$  逆向转运蛋白的方法,成功地从 1 480 000 个包含来自土壤 DNA 的克隆中获得了两个新的逆向转运蛋白。由于从元基因组文库中筛选表达特定生物活性克隆的频率很低,阳性率大约只有克隆总数的 0.1%左右<sup>[11]</sup>,因此需要发展更有效的筛选方法增加新的活性分子被发现的几率,PCR 以及探针杂交的方法也随之建立起来。

功能筛选需要克服异源表达的障碍,实现在巨型文库中有效探测少量阳性克隆。同时利用探针杂交,在文库中进行快速筛选,将具有同一个功能的一系列克隆装配好之后进行测序。

## 2 元基因组在微生物生态学和工业生物技术中的作用

### 2.1 共生系统研究

很多专性共生微生物不能实现实验室培养,大多在寄主组织中富集且纯度较高,一旦从寄主组织中分离出来,他们就成为用于元基因组分析的理想材料。*Buchnera aphidicola* 基因组的重建提供了该菌和蚜虫之间共生体系的进化信息和生理生化上的相互独立性以及基因组收缩和重排的机理<sup>[13]</sup>。Hughes 等<sup>[14]</sup>从深海管状蠕虫 *Riftia pachyptila* 的一种共生体变形细菌中提取总 DNA 并且构建 fosmid 文库,研究它的生理学特性,并从文库中确定了和 RubisCO 相似的基因,这个发现支持了该化能自养型细菌为寄主固定碳源的假说。尽管元基因组这个概念意味着多成员群落的混合基因组,但研究这些系统为分析复杂群落提供了很好的模型。

### 2.2 元基因组与化学生态学结合

与竞争和合作相关的基因是依赖于生态系统成员之间的关联以及外界资源的匮乏来发挥功能,所以仅仅依靠序列是难以识别的。而结合功能基因组和化学生态学来识别具有生物活性和生态功能的小分子可以帮助人们获得更多的信息。

基于功能性元基因组来发现小分子主要集中于对抗生素的研究上(表 1),序列分析可以用来筛选小分子,如编码聚酮合成酶(PKS)基因。Seow 等<sup>[15]</sup>设计了和 PKS 基因高度保守片段杂交的引物,直接从土壤中扩增了新的 PKS 基因簇,这也是第一次通过 PCR 扩增的方法直接从环境样品中克隆获得 PKS 基

因。Courtois 等<sup>[16]</sup>构建了土壤 DNA 的穿梭黏粒文库，得到了 5 000 个克隆，并从中发现了 11 个新的 I 型聚

酮合酶(PKS I)的基因，同时通过 HPLC 技术分离了两种新的化合物。

表 1 基于元基因组技术发现的酶和化合物

酶和化合物	筛选和策略	载体	文献
Polyketide synthase	Sequence-based	cosmid	[16]
Agarase	Function-driven	cosmid	[17]
Amidase	Sequence-based	cosmid	[17]
Long-chain <i>N</i> -acyltyrosine antibiotics	Function-driven	cosmid	[18]
Indirubin	Function-driven	BAC	[19]
Turbomycin	Function-driven	BAC	[20]

### 2.3 功能基因及酶的筛选

新工业用酶的发现是元基因组研究的又一大贡献(表 1)。通过构建特定环境基因组文库，对文库进行序列和功能筛选，获得表达特定酶活性的克隆。张鹏等<sup>[21]</sup>从自制堆肥样品中提取未培养细菌的总 DNA，用科斯质粒 pWEB TNC 为载体构建元基因组的 cosmid 文库，筛选文库获得表达木聚糖酶活性的克隆，通过序列分析证实为新的木聚糖酶基因，这也是细菌木聚糖酶基因资源开发利用的一条新途径。赵广存等<sup>[22]</sup>通过构建牛瘤胃未培养细菌的元基因组文库，得到 10 个表达  $\alpha$ -葡萄糖苷酶活性的克隆，其中一个新的纤维素酶基因，推测这个基因可能是来自噬纤维菌属。

### 2.4 元基因组用于解析微生物参与的生物地球化学循环过程

元基因组技术可以用来解析微生物群落与功能的关系，如微生物参与的生物地球化学循环过程。对一个群落中的成员进行特定的功能分析，构建微生物代谢的整体模型。这类模型是关于生物如何分担工作量和维持群落营养和能量支出平衡的，并通过遗传和化学的方法来进行检测。Tyson 等<sup>[23]</sup>构建了黄铁矿污水(AMD)微生物群落的基因组文库，并且获得了具有高覆盖率的序列，以 G+C 含量、16S rRNA、tRNA 合成酶基因为标准鉴别序列的来源，几乎重建了 *Leptospirillum* group II 和 *Ferroplasma* type II 的全基因组，通过基因功能的分析，发现微生物群落为适应这个极端环境，具有固碳、固氮以及能量传递等特殊的代谢网络系统。

AMD 群落的研究为以后的研究者提供了一个可借鉴的分析模型，虽然对一些复杂的生理及系统进化群落来说，鉴别 DNA 片段的起源，赋予其功能

可能是困难的<sup>[24]</sup>，但是这种方法对所有的群落都是有用的。

相对 AMD 群落来说，马尾藻海更为复杂，群落中成员的系统进化关系还不清楚。Venter 等<sup>[25]</sup>进行了有史以来最大的元基因组项目，该研究小组测得了 1 600 Mbp 的序列，从中发现了 120 万个新基因，未知功能基因 794 061 个，能量传递相关的 69 718 个，视紫质类似的光受体基因 782 个，大量的磷酸盐相关的基因，有助于磷循环的研究，包括磷的获得以及转运机制等，该项工作的成果被称为一座金矿，对于光合作用、能量传递、污染物控制等研究，以及了解海洋生物的多样性及复杂性提供了基础数据。

紧接着马尾藻海元基因组项目，Yooseph 等<sup>[26]</sup>通过 Sorcerer II 全球海洋取样(Global Ocean Sampling, GOS)，进行了更大规模的元基因组研究，对获得的环境基因组利用鸟枪法进行测序(shotgun sequencing)，组合了超过 770 万条 GOS 序列，预测超过 612 万个编码蛋白，发现一些原本认为有界(kingdom)特异性的蛋白域实际上不止在一个界存在，可能一些代系(lineages)更古老，或者它们拥有一些可以水平转移的基因序列。并且发现了比预期更多的新的病毒起源的序列，包括适应大洋强紫外线进行修复的 UV 修复酶等。提示了海洋微生物群落在海洋生态系统物质循环和能量循环中的重要作用。

### 2.5 比较元基因组学

Tringe 等<sup>[27]</sup>发现通过元基因组学的方法获得的大量复杂的环境样品数据很难装配成完整的基因组，因为在来自土壤文库的将近 150 000 个读码框中只有不到 1%和独立克隆之间有重叠域。该研究小组将每一个环境样品独立的小片段 DNA 定义为一个基因标签(EGT)，比较了来自土壤的 3 个 whale fall 样品，

来自酸矿污水及马尾藻海的 3 个独立样品的装配和未装配的开放读码框。对 90% 的 EGTs 进行了预测, 其中超过 1/3 的 EGTs 包含了两个或更多的预测的开放读码框, 该小组通过环境中基因和功能模块的比较分析为预测能量来源及污染等级提供了大量信息。基于环境基因组 EGT 的比较基因组学研究, 对于认识微生物群落以及他们所处的环境起着十分重要的作用。

### 3 元基因组的研究展望

元基因组的分析揭示了表面上一致的种群包含了真实的微观不均一性<sup>[23]</sup>。这些研究暗示了对混合种群进行基因组研究的重要性, 继“人类基因组计划”之后, “人类元基因组计划”在 2006 年开始启动, 计划发现百万个新基因, 有助于阐明许多疾病的发生机理, 并为新药及控制药物毒性等研究发挥作用。事实上, 元基因组也能用于可培养微生物的基因组多样性的研究, 如已经将微生物的元基因组技术用于沼气发酵系统。

充分发挥元基因组的潜能, 关键在于文库构建和分析方法上的改进。基于序列分析的方法, 关键是加快测序速度和降低成本, 通过探测同源体序列(更依赖于预测蛋白质的三级结构而不是简单的初级序列)使大文库的管理和分析变得容易(包括利用生物信息学工具来分析数据中的大序列, 快速装配多种基因和提供文库构型的基因芯片<sup>[28]</sup>), 筛选得到未显现表型功能的阳性克隆。而功能驱动的分析主要依赖方法上的创新, 包括改善异源基因表达的策略以及有效筛选大文库的方法。应用 DGGE/TGGE 技术监测复杂环境中特异的功能基因及表达, 为元基因组学研究提供有利信息和基因筛选方案<sup>[29,30]</sup>, 相信元基因组会继续扩大并丰富人们对微生物的认识。

元基因组改变了微生物学家解决问题的方法, 尤其对于深海环境中大量未培养微生物的研究, 重新定义了基因组并且加速了新基因发现的速度。元基因组在生物工业上也具有很广阔的前景, 新基因的发现可以提供微生物群落结构的信息和功能, 用来解决医药、农业以及工业问题。

参考文献:

[1] Lane D J, Pace B, Olsen G J, *et al.* Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses [J]. *Proc Natl Acad Sci USA*, 1985, **82**(20): 6 955-6 959.

- [2] Pace N R, Stahl D A, Lane D J, *et al.* The analysis of natural microbial populations by ribosomal RNA sequences [J]. *Adv Microb Ecol*, 1986, **9**: 1-55.
- [3] Schmidt T M, DeLong E F, Pace N R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing [J]. *J Bacteriol*, 1991, **173**(14): 4 371-4 378.
- [4] Healy F G, Ray R M, Aldrich H C, *et al.* Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose [J]. *Appl Microbiol Biotechnol*, 1995, **43**(4): 667-674.
- [5] Stein J L, Marsh T L, Wu K Y, *et al.* Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon [J]. *J Bacteriol*, 1996, **178**(3): 591-599.
- [6] 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 [J]. *Chem Biol*, 1998, **5**(10): 245-249.
- [7] Handelsman J. Metagenomics: Application of genomics to uncultured microorganisms [J]. *Microbiol Mol Biol Rev*, 2004, **68**(4): 669-685.
- [8] Sharma R, Ranjan R, Kapardar R K, *et al.* 'Unculturable' bacterial diversity: An untapped resource [J]. *Curr Sci*, 2005, **89**(1): 72-77.
- [9] Béjà O, Spudich E N, Spudich J L, *et al.* Proteorhodopsin phototrophy in the ocean [J]. *Nature*, 2001, **411**: 786-789.
- [10] Diaz-Torres M L, McNab R, Spratt D A, *et al.* Novel tetracycline resistance determinant from the oral metagenome [J]. *Antimicrob Agents Chemother*, 2003, **47**(4): 1 430-1 432.
- [11] Henne A, Schmitz R A, Bömeke M, *et al.* Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli* [J]. *Appl Environ Microbiol*, 2000, **66**(7): 3 113-3 116.
- [12] Majernik A, Gottschalk G, Daniel R. Screening of environmental DNA libraries for the presence of genes conferring Na<sup>+</sup>(Li<sup>+</sup>)/H<sup>+</sup> antiporter activity on *Escherichia coli*: Characterization of the recovered genes and the corresponding gene products [J]. *J Bacteriol*, 2001, **183**(22): 6 645-6 653.
- [13] Abbot P, Withgott J H, Moran N A. Genetic conflict and conditional altruism in social aphid colonies [J]. *Proc Natl Acad Sci USA*, 2001, **98**(21): 12 068-12 071.
- [14] Hughes D S, Felbeck H, Stein J L. A histidine protein kinase homolog from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila* [J]. *Appl Environ Microbiol*, 1997, **63**(9): 3 494-3 498.
- [15] Seow K T, Meurer G, Gerlitz M, *et al.* A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: A means to access and use genes from uncultured microorganisms [J]. *J Bacteriol*, 1997, **179**(23): 7 360-7 368.
- [16] Courtois S, Cappellano C M, Ball M, *et al.* Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products [J]. *Appl Environ Microbiol*, 2003, **69**(1): 49-55.

(下转第 88 页)