

# 宏基因组学研究进展

黄循柳<sup>1</sup> 黄仕杰<sup>1</sup> 郭丽琼<sup>1</sup> 林俊芳<sup>1,2\*</sup>

(1. 华南农业大学食品学院 广东 广州 510640)

(2. 华南农业大学 生物质能研究所 广东 广州 510640)

**摘要:** 不可培养微生物占据微生物总数的99%以上,这已成为微生物资源开发利用的一个限制性因素。宏基因组学是通过提取某一环境中的所有微生物基因组DNA、构建基因组文库及对文库进行筛选寻找和发现新的功能基因及活性代谢产物的一种方法。它避开了微生物分离培养的过程,极大地扩展了微生物资源的利用空间,是现代基因工程一个新的发展方向和研究热点。本文主要对宏基因组的DNA提取方法、文库的构建、筛选策略的选择及近年来宏基因组学在各领域中的应用研究现状进行了综述。

**关键词:** 宏基因组, 未培养微生物, 文库构建, 文库筛选

## Advances of Metagenomics

HUANG Xun-Liu<sup>1</sup> HUANG Shi-Jie<sup>1</sup> GUO Li-Qiong<sup>1</sup> LIN Jun-Fang<sup>1,2\*</sup>

(1. College of Food Science, South China Agricultural University, Guangzhou, Guangdong 510640, China)

(2. Institute of Biomass Research, South China Agricultural University, Guangzhou, Guangdong 510640, China)

**Abstract:** It has evolved as a main limiting factor for the exploitation and application of microbial resources that more than 99% of microorganisms in natural environments cannot be cultivated so far by current isolation and culture methods. Metagenomics is an advanced methodology by means of extracting all microbial genomic DNAs in certain environmental habitat, constructing and screening metagenomic libraries to seek novel functional genes and/or biologically active compounds. It can overcome the advantages of isolation and cultivation procedures in traditional microbial method, and thus greatly broaden the space of microbial resource utilization. So it has become one of the powerful research tools for microbiology, biotechnology, soil and environmental sciences, and a new field of genetic engineering. This paper mainly reviewed the metagenomic methodology including genomic DNA isolation, library construction and screening strategies. And this review also outlined the up-to-date applications of metagenomics in various fields.

**Keywords:** Metagenomics, Uncultured microorganism, Library construction, Library screening

环境微生物是自然界中分布最广、种类最多、数量最大的生物类群。自1663年,列文虎克利用自制显微镜首次揭开微生物神秘的面纱之后近300多年的时间里,人们对于微生物的研究主要是建立在

纯培养基基础上,后来人们发现通过纯培养方法估计的环境微生物多样性只占总量的0.1%~1%<sup>[1]</sup>,这使得微生物的多样性资源难以得到全面的开发和利用,如何充分开拓利用环境微生物新资源是目前环境微

生物研究的重要课题之一。近年来发展起来的宏基因组学技术避开传统的微生物分离培养方法直接从环境样品中提取总DNA, 通过构建和筛选宏基因组文库来获得新的功能基因和生物活性物质, 宏基因组文库既包括了可培养的, 又包括了未培养的微生物遗传信息, 因此增加了获得新生物活性物质的机会。本文主要对宏基因组研究策略及这一研究领域的最新进展进行简要综述。

## 1 宏基因组学研究策略

1998年 Handelman 等首次提出宏基因组的概念<sup>[2]</sup>, 并将其定义为: 一种以环境样品中的微生物群体基因组为研究对象, 以功能基因筛选和测序分析为研究手段, 以微生物多样性、种群结构、进化关系、功能活性、相互协作关系、及与环境之间的关系为研究目的的新的微生物研究方法, 并第一次使用了Metagenomics这个名词。其基本策略流程为: 样品和基因(组)的富集; 提取特定环境中的基因组DNA; 构建宏基因组DNA文库; 筛选目的基因; 目的基因活性产物表达(图1)。

### 1.1 环境样品及目的基因组富集

在宏基因组文库筛选过程中目的基因仅占总

DNA的一小部分, 对环境样品的预富集可以大大提高目的基因的检出几率。目前预富集技术主要分为细胞水平富集和基因组水平富集。其中细胞水平富集主要是通过利用选择培养基对目的微生物进行富集培养, 最常用的方法是底物选择<sup>[3]</sup>, 此外还包括营养选择<sup>[4]</sup>及物理化学指标选择<sup>[5]</sup>。但是由于富集培养选择性地富集了具有快速生长特性的菌群, 因此导致大部分物种多样性信息丢失。目前可以通过先在严格胁迫条件下短期处理, 然后改为较温和的处理条件, 可以从一定程度上克服这种方法的局限性。基因组水平富集常用的技术为稳定同位素探针技术(SIP), 原理是采用稳定同位素标记底物, 其中的“重”原子掺入到具有代谢活性的微生物核酸中, 采用密度梯度离心的方法将“重”的DNA与“轻”组分分离, 被标记的“重”核酸可以作为PCR的模板, 用来构建宏基因组文库<sup>[6]</sup>。例如采用C标记的甲醇研究森林土壤宏基因组DNA, 结果鉴定出变形细菌 $\alpha$ -亚纲中所有已知的嗜甲基菌, 并在一种嗜酸菌中发现了新的甲醇还原酶基因<sup>[7]</sup>。此外, 还有报道利用抑制性消减杂交技术<sup>[8]</sup>、差异显示技术<sup>[9]</sup>、噬菌体展示技术<sup>[10]</sup>、亲和捕获技术<sup>[11]</sup>及DNA微阵列技术<sup>[12]</sup>等技术来富集目的基因。

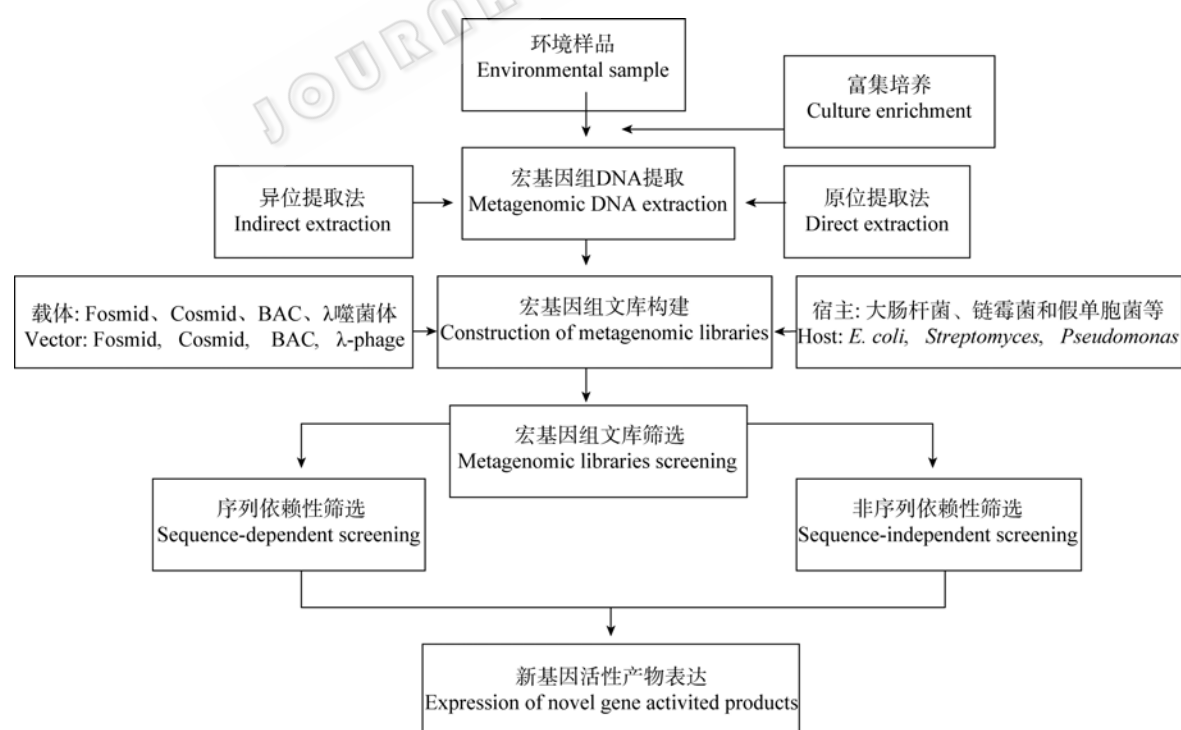


图1 环境微生物宏基因组学研究策略

Fig. 1 General process of metagenomic strategie

## 1.2 宏基因组 DNA 的提取

获得高浓度、大片段、无偏好的环境样品总DNA是宏基因组文库构建的难点之一。近年已有多种宏基因组DNA的提取纯化方法陆续建立起来,大体可分为两类:一类是直接提取法,又称原位提取法。这类方法不经过样品中微生物的培养和分离,通过化学法、酶解法或物理法直接破碎环境中的微生物细胞而使DNA得以释放,并对DNA进行纯化。其中以Zhou法<sup>[13]</sup>和Tsal<sup>[14]</sup>法最为常用。该法操作简便、省时、成本低,所获得DNA具有较好的完整性,并能够代表某一生境的微生物群落多样性。但该发放常会出现细胞裂解不完全或DNA与土壤杂质成分产生共沉淀而无法有效地去除等问题,所以一般需要进一步的DNA纯化处理,同时所提取获得的DNA片段较小(1 kb~50 kb),主要适用于小片段文库的构建;另一类是间接提取法,即异位提取法,是利用离心介质或者梯度离心等方法先把微生物从环境样品中分离出来,再按处理纯培养细胞的方法裂解微生物细胞提取DNA<sup>[15-17]</sup>。该法获得的宏基因组DNA受到胞外杂质污染干扰较少,纯度较高、DNA完整性好(20 kb~500 kb),适合构建大片段的宏基因组文库。但该法操作较繁琐、费时、成本高、偏差大、DNA得率较低,其产率只是直接裂解法的1%~10%<sup>[18]</sup>,且获得的DNA往往不能完全代表样品所在生境的生态学多样性。本实验室对温泉淤泥、纸厂污泥、红树林淤泥、沼泽淤泥等12个环境样品的总DNA提取方法进行了摸索,研究工作表明,直接提取法提取效率较高,不容易丢失物种信息,能够获得23 kb左右的DNA片段。并首次加入漆酶来有效去处腐殖酸类物质,比聚乙烯吡咯烷酮(PVPP)处理效果更好,DNA得率更高。而间接提取法DNA提取效率较低,容易丢失物种信息。因此,建议采用直接提取法获得环境样品的总DNA,以便分析环境样品微生物群落多样性信息。

## 1.3 宏基因组文库的构建

**1.3.1 载体的选择:**载体选择在宏基因组技术中占有十分重要的地位。载体系统的选择主要依赖于DNA提取质量、插入片段、质粒拷贝数、宿主菌及筛选方法。根据克隆载体的不同宏基因组文库可分为质粒文库、细菌/酵母人工染色体文库、噬菌体类载体文库。早期宏基因组文库主要以质粒载体和大肠杆菌宿主细胞构建而成的小片段(<15 kb)文

库<sup>[19-21]</sup>。该文库操作简便、拷贝数高、对DNA质量要求不高,适合纯度低或剪切严重的DNA模版。但插入片段小,筛选量大,活性比率低,对大基因簇无能为力,主要适用于分析新的基因序列信息及筛选编码代谢相关的单一基因或小操纵子<sup>[22]</sup>。然而,很多微生物活性物质是其次生代谢产物,代谢途径由多基因簇调控,因此尽量插入大片段DNA以获得完整的代谢途径多基因簇是很有必要的,目前已有报道能容纳15 kb~40 kb外源DNA的Fosmid文库<sup>[23-25]</sup>和Cosmid文库<sup>[26-28]</sup>,同时已经成功构建了容量高达200 kb~350 kb外源DNA的细菌人工染色体文库和酵母人工染色体文库<sup>[29-31]</sup>。该文库不仅拥有巨大的插入片段载量,而且稳定性好、筛选效率高、克隆表达能力强,可获得基因簇表达活性物质。主要用于分析某一生境中微生物群落多样性及筛选由基因簇或多个基因控制表达的活性物质。但其操作较复杂繁琐、对DNA纯度要求较高,且拷贝数低,扩增较困难。此外, $\lambda$ -噬菌体和其他病毒也可以作为构建宏基因组克隆文库的载体<sup>[32-34]</sup>。噬菌体展示库是一种十分有效的高通量筛选技术,该技术通过对表面展示表达产物的亲和选择分离相应的DNA序列,能够从宏基因组中富集稀有DNA序列<sup>[10]</sup>,但其局限性在于只能表达分子量小于50 kD的蛋白。

**1.3.2 宿主细胞的选择:**不同的微生物种类所产生的活性物质类型有明显差异。宿主菌株的选择主要考虑转化效率、重组载体在宿主细胞中的稳定性、宏基因的表达、目标性状(如抗菌)缺陷型等因素。其中*E. coli*是最为常用的宿主细胞<sup>[35-38]</sup>,其优点是操作简单、繁殖迅速、培养代谢易于控制。但是由于环境样品的总DNA有很大一部分为真核基因组DNA,其在细菌宿主中往往不能表达,大大限制了这部分基因的筛选。最近已有报道利用链霉菌、假单胞菌等真菌作为受体来鉴定有关新抗生素生物合成的基因<sup>[39,40]</sup>。但是最近的生物信息学研究表明,对于一个插入子(<10 kb)的文库,为寻找一个目的基因需要筛选 $10^5 \sim 10^6$ 个克隆<sup>[41]</sup>,这表明从复杂的宏基因组中筛选目的基因在技术上还存在着其特殊的困难。因此,宿主系统的进一步探索是更有效的开发宏基因组资源的关键技术之一。

## 1.4 宏基因组文库筛选

由于宏基因组文库容量较大,目标克隆子的筛

选一直是宏基因组学技术中的瓶颈。近年来,各种宏基因组文库筛选方法相继建立起来,根据筛选原理大体分为两大类:序列依赖性筛选法和非序列依赖性筛选法。

1) 序列依赖性筛选法,是根据文库中的已知序列或保守序列来设计杂交探针或PCR引物,从宏基因组文库中筛选目标序列。该法克服了功能筛选法中必须依赖表达后的活性产物进行检测,效率较高。其中已知功能基因PCR扩增技术是最为常用的序列依赖性筛选法<sup>[42-44]</sup>。此外,DNA微阵技术和整合子系统技术也可以作为序列依赖性筛选法。例如Wagner等应用DNA微阵列技术对活性污泥中微生物的群落结构进行了研究,获得了大量新的信息<sup>[45]</sup>,Stokes等利用整合子系统技术成功筛选到与DNA糖苷酶、磷酸转移酶和甲基转移酶同源的新基因<sup>[46]</sup>。但是该筛选法存在2个主要的缺陷:(1)引物的设计依赖于已有的序列信息,共同进化的2个功能相似的基因难于用“族特异性”引物区分开来,因此很难发现新的功能基因;(2)通过PCR扩增功能基因,一般只能获得结构基因的一个片段,而不能获得完整的功能基因。

2) 非序列依赖性筛选法,其不依赖于任何已知序列信息,仅根据文库克隆子产生的活性物质进行筛选。其中以功能筛选法最为常用,原理是直接对目的克隆表达的性状在选择培养基上进行筛选,已有报道成功筛选出产生木聚糖酶、纤维素酶、脂肪酶及抗菌活性的克隆子<sup>[47-50]</sup>。该法筛选能够发现全新的活性物质或全长基因,但工作量大、效率低,生物转化的产物仅在少数情况下具有可见性状,酶活性的检测受到研究方法的限制。此外,由于受到酶活检测方法的限制,大多数宝贵的酶资源不能通过上述基于活性的方法发掘出来,近期基因陷阱技术筛选法倍受研究学者们的关注。其中,(1)利用目的基因缺失的突变体宿主菌株,转化后在选择性培养基上进行功能互补的生长特征来筛选。例如Majernik等通过缺陷型大肠杆菌为宿主菌构建土壤宏基因组文库,筛选出了与 $\text{Na}^+/\text{H}^+$ 反向转运通道相关新基因<sup>[51]</sup>;(2)将宏基因组DNA随机克隆到无启动子的绿色荧光蛋白基因(GFP)前面,然后通过荧光激活细胞分离(FACS)技术,在添加特定底物的条件下对表达库进行富集,就能够选择出对该底物具有代谢活性的克隆。如Uchiyama等利用底物诱导基

因表达法从环境宏基因组文库中筛选到新的生物代谢相关基因<sup>[52]</sup>;Williamson等利用底物诱导基因表达法从泛洪区土壤宏基因组文库中筛选到新的生物小分子物质<sup>[53]</sup>。该法优点在于它为高通量筛选提供了保障,而且不需要对底物进行修饰,尤其适合于工业生产上的应用。但也存在一些问题:(1)无法利用不能进入细胞质的底物;(2)荧光激活细胞分离仪(FACS)对进样设备的要求也比较高;(3)建立高度耐受“超相容性”表达系统,以表达任意来源的编码蛋白和酶具有一定难度。

## 2 宏基因组学应用研究进展

### 2.1 宏基因组学在微生物生态学上的应用

随着分子生物学技术的发展,基于16S rRNA基因分析的变性梯度凝胶电泳(DGGE)、温度梯度凝胶电泳(TGGE)、单链构象多态性(SSCP)、限制性片段长度多态性(RFLP)等新技术的应用,宏基因组技术在分析环境微生物复杂群落结构领域得到广泛应用。如Martin等构建地中海深层水体微生物宏基因组文库,通过序列分析和16S rRNA系统发育比对,发现该水体的微生物种群与太平洋阿罗哈水域中层水体的微生物种群具有一定的相似性,并提出在无光的条件下,温度是影响微生物种群在水体中分布的主要因素<sup>[54]</sup>。Zhang等构建红树林淤泥宏基因组文库,通过PCR扩增及变性梯度凝胶电泳(DGGE)对该区域固氮菌的多样性进行分析,结果揭示红树林地区固氮菌的生物多样性特征,其结果表明多数为变形菌,也含少数的固氮菌属、除硫单胞菌属、德克斯氏菌属和根瘤菌等<sup>[55]</sup>。张薇等采用宏基因组技术对西北黄土高原柠条种植区土壤微生物多样性进行分析,发现变形杆菌纲是根表土壤区系中的有优势微生物菌群(70.3%),尤其存在大量能够诱导植物形成根瘤的根瘤菌和对植物有促生长作用的 $\gamma$ -Proteobacteria类微生物,说明了植物根系和土壤环境微生物菌群具有相互选择性<sup>[56]</sup>;2008年肖凯等以宏基因组DNA为模板,采用不同的PCR引物对温泉的高温水底沉积物微生物多样性进行分析,发现了一株新的菌株JS-X2与在美国黄石公园温泉发现的未培养细菌有95%的相似性,并且与嗜热蓝细菌聚球藻有89%的相似性<sup>[57]</sup>。近年来宏基因组技术在环境病毒及噬菌体的多样性分析方面也被广泛应用。例如Fierer等通过构建牧场、沙漠、雨林土壤宏

基因组文库对环境中的细菌、古生菌、真菌及病毒多样性进行了研究,并揭示了土壤环境中包含着大量不可培养的新病毒种类,其基因型特征与常规培养获得的病毒具有很大的差异性<sup>[58]</sup>;Kim等通过构建稻田土壤宏基因组文库,利用多重置换扩增(MDA)技术对土壤样品中病毒基因多样性进行了研究,结果表明扩增得到的病毒基因序列与目前报道的病毒序列具有很大的差异性,进一步说明了土壤环境中包含着大量不可培养的病毒种类<sup>[59]</sup>;Sabet等通过构建宏基因组文库对美国莫诺湖水体中噬菌体的多样性进行了研究,研究发现不可培养的噬菌体才是该特殊生境中的优势群体,揭示了海洋是一个巨大的未知RNA病毒库<sup>[60]</sup>;Breitbart等通过构建海水及海底沉积物宏基因组文库对该地区不可培养病毒的多样性进行分析,结果发现扩增得到的病毒基因型中65%为新的基因型,其中包含一类海藻病毒,多数病毒具有新的基因型,与节肢动物和高等植物病毒存在很大的序列差异性<sup>[61]</sup>;2003年Breitbart等首次通过构建宏基因组文库对人体排泄物中的未培养病毒多样性进行研究,经过扩增及鸟枪测序鉴定,结果表明获得的病毒大约有1200种基因型,其基因序列与先前报道的病毒序列具有很大的差异性,多为新的基因型病毒,并揭示存在人体中的新病毒与人类疾病可能具有一定的相互性<sup>[62]</sup>;2005年Cann等通过构建宏基因组文库对马排泄物中的未培养病毒多样性进行研究,经过测序鉴定,获得233种不同基因型的病毒,其中52%为长尾噬菌体科,26%为未分类的噬菌体,17%为肌病毒科;4%为短尾病毒科;2%为脊椎动物正痘病毒<sup>[63]</sup>。总的来说,宏基因组技术为微生物多样性研究提供了巨大的平台。

## 2.2 宏基因组学在生物酶制剂开发中的应用

宏基因组学技术最引人注目的贡献主要集中在新型生物酶制剂的探索 and 开发领域。近年来研究者们已成功构建了土壤、海底淤泥、温泉淤泥、油厂污泥、动物瘤胃内容物、动物粪便等宏基因组文库,并筛选到脂肪酶、蛋白酶、淀粉酶、乙醇氧化酶、木聚糖酶、纤维素酶及脱羧酶等酶制剂,并且在此基础上获得新酶的许多特征信息(表1)。所采用的载体种类十分广泛,包括 Fosmid、Cosmid、BAC、 $\lambda$ 噬菌体以及各种穿梭载体,所采用的宿主系统为常用的大肠杆菌、链霉菌和假单胞菌等。

通过宏基因文库筛选得到的生物分子大多数与

已知的基因产物相似性差或者完全是新的分子,这些新的生物分子主要来源于环境未培养微生物的基因和其多样的代谢物。环境样品DNA的克隆和筛选只是所有环境遗传信息多样性的一小部分,环境微生物和宏基因组的多样性仍旧是发现新的天然活性产物的丰富广阔资源,为研究者们探索开发新的生物催化剂提供了巨大的资源空间。

## 2.3 宏基因组学在医学中的应用

宏基因组技术的出现为新药物的探索和发现提供了可能的技术支持,并扩大了微生物代谢产物及分子活性物质筛选平台。例如早在2000年,Wang等构建土壤宏基因组文库,通过文库筛选获得TerragineA及其相关成分,目前已广泛应用于医学治疗领域<sup>[64]</sup>,证明了自然环境中的丰富微生物代谢产物可以通过宏基因组技术为人们所利用;同年Brady等从土壤宏基因组文库中筛选发现一种长链N-酰氨基酸抗生素物质<sup>[50]</sup>,并在2004年构建凤梨科植物树茎流出液宏基因组文库,筛选鉴定获得了抗菌物质PalmitoylPutrescine<sup>[65]</sup>。2001年Macneil等构建了土壤宏基因组BAC文库,通过序列分析筛选获得5个能产生抗菌小分子物质靛玉红并对其相关成分进行研究<sup>[66]</sup>;2002年Gillespie等构建土壤宏基因组文库筛选获得两种抗菌物质Turbomycin A和B,并且发现Turbomycin A和B对革兰氏阴性和革兰氏阳性菌具有广谱抗菌活性<sup>[67]</sup>;2003年DiazTorres等通过构建人唾液宏基因组文库,筛选获得一种新的四环素抗性基因Tet(37),该活性物质对四环素具有很好的抗性<sup>[68]</sup>;2008年Mori等通过活性污泥宏基因组文库筛选获得两种不同的博来霉素抗性基因,经过比对发现于来源放射菌类基因差异较大,可能为新的博来霉素抗性基因<sup>[69]</sup>。国内在利用宏基因组技术获取新型药物的研究较少,尚处于萌芽阶段,赵晶等从南极中山站排污口采集污泥,构建宏基因组文库,并通过差异性DNA修复实验(DDRT)筛选得到具有抗肿瘤效应的物质<sup>[70]</sup>。同时利用宏基因组技术研究探讨人类肠道中不可培养微生物多样性也有了很大进展。如Kurokawa等利用宏基因组技术对13个处于不同年龄层的健康人体粪便微生物种群进行了研究,结果分析表明未断奶婴儿的肠道微生物种群在系统发育和基因组成上具有较大个体差异性,而在成人及已断奶儿童则呈现出高度的功能一致性,并对成人及婴儿肠道内编码该生境微生物主要功能的基因家族的特性进行分析,发

表 1 已构建的宏基因组文库及其筛选到的活性克隆  
Table 1 Examples of recently identified biocatalysts from metagenome libraries

环境样品 Environmental sample	目的基因 Target gene	载体/宿主 Vector/Host	筛选 Screening	文献 References
深海淤泥 Deep sea sludge	碱性脂肪酶	Fosmid/ <i>E. coli</i> DH5α	活性	[23]
温泉 Hot spring sludge	嗜热脂肪酶	pET(+)/ <i>E. coli</i> TOP10	活性/测序	[49]
海底淤泥 Sea-bottom sludge	酯酶	pIndigoBAC/ <i>E. coli</i> EPI300	活性	[29]
水体富集物 Water enrichment	酯酶、纤维素酶	λ-phage/ <i>E. coli</i>	活性	[32]
礁湖污泥 Reef lake sludge	嗜冷木聚糖酶	pET-22b/ <i>E. coli</i> DE3	活性	[19]
土壤 Soil	木聚糖酶	pHBM803/ <i>E. coli</i> DE3	活性	[47]
海岸淤泥 Coastal sludge	纤维素酶	Fosmid/ <i>E. coli</i> EPI300	活性/测序	[24]
纸浆沉淀物 Paper mill effluent	纤维素酶	pWEB::TNC/ <i>E. coli</i> EPI100	序列/活性	[48]
牛瘤胃内容物 Bovine rumen	淀粉酶、纤维素酶	pCC1BAC/ <i>E. coli</i> EPI300	活性/序列	[30]
土壤 Soil	淀粉酶	BAC/ <i>E. coli</i> DH10B	活性/测序	[31]
马排泄物 Horse excremen	多功能糖基水解酶	λ-phage/ <i>E. coli</i> EPI100	活性	[33]
沿海海水 Coastal seawater	几丁质酶	λ-phage/ <i>E. coli</i> EPI100	活性	[34]
土壤 Soil	乳酸解聚酶	pUC18/ <i>E. coli</i> DH10B	活性	[20]
泰山土壤 Taishan Mountain Soil	β-半乳糖苷酶	pET30a/ <i>E. coli</i> DE3	序列/活性	[21]
牛瘤胃内容物 Bovine rumen	β-葡萄糖苷酶	pWE::TNC/ <i>E. coli</i> EPI100	序列	[27]
温泉土壤 Hot spring soil	耐热蛋白酶	pSK(+)/ <i>E. coli</i> DH10B	序列	[43]
土壤 Soil	乙醇氧化酶	pSK+/ <i>E. coli</i> DH5α	活性/测序	[18]
温泉土壤 Hot spring soil	铁过氧化物歧化酶	pUC19/ <i>E. coli</i> DH5α	活性/测序	[44]
耕作淤泥 Arable soil	聚酮体合成酶	Cosmid/变铅青链霉菌	活性	[39]
甘油厂污泥 Glycerol mill mud	甘油脱水酶	pSE380/ <i>E. coli</i> JM109	序列	[42]
碱性污泥 Akaline sludge	脱羧酶	pETB+/ <i>E. coli</i> DH5α	活性	[35]
油料污泥 Ptroleum sludge	萘类分解酶	pKS13S/ <i>E. coli</i> XL1B-MR	活性	[38]
土壤 Soil	5-烯醇式丙酮莽草酰-3-磷酸合酶	pACYC184/AKM4188	活性	[25]
活性污泥 Activated sludge	聚-3-羟基丁酸代谢酶	pRK7813/ <i>E. coli</i> HB101	活性	[37]
地热沉积物 Geothermal sediments	亚铁血红素合成酶、磷酸二脂酶	TOPO2XL/ <i>E. coli</i> TOPO10	活性	[36]
深海污泥 Deep-sea sludge	TEM 型-内酰胺酶	pCC1FOS/ <i>E. coli</i> EPI300	活性	[28]
深海污泥 Deep-sea sludge	烷烃降解酶	PWEB:TNC/ <i>E. coli</i> EPI100	活性	[26]

现了一个新的人类肠道微生物基因家族和一个共扼转座子<sup>[71]</sup>。2003 年Breitbart等通过构建宏基因组文库对人体排放物中的未培养病毒多样性进行研究,经过鸟枪测序法鉴定,获得的病毒大约有 1200 种基因型,结果比对表明其基因序列与先前报道的病毒具有很大的差异性,大多数为新病毒,并证明这些病毒极有可能与人类的疾病有着密切的关系<sup>[62]</sup>。2008 年Finkbeiner等通过构建 12 个腹泻小孩肠道内容物宏基因组文库,来观察病人肠道中病毒的生物多样性,发现扩增得到的病毒序列与GenBank病毒库中的已知序列同源性很低,并推断这些病毒极有

可能与人类的腹泻疾病有着密切的关系<sup>[72]</sup>。随着宏基因组技术的成熟,必将加快宏基因组技术在医学中的应用。在人体微生物抗药性的研究,人体与不可培养病原菌的相互关系的探索等方面将做出重大贡献。

3 结语与展望

到目前为止,研究者们已通过宏基因组克隆技术筛选获得全新的功能基因和活性物质,大大促进了药学和生物技术工业的快速发展,增长了人类对微生物的生态学与生理学的认知,也为人类探索自

自然界广袤的微生物资源提供了无限的遐想。但由于宏基因组技术作为一种新的技术刚刚起步,在构建宏基因组文库的实施过程中还存在着很多待以解决的问题。一方面需要探索更加适合构建宏基因组文库的载体和真核表达宿主细胞。就目前报道文库构建一般以质粒、细菌人工染色体、酵母人工染色体等为载体,以细菌、酵母、链霉菌等为受体。这些质粒的克隆能力相对较小,不能满足大片段目的基因的插入,而BAC载体虽然能插入较大的基因片段,但其拷贝数较少,不适合次级代谢产物的生产。而表达宿主细胞多以细菌为主,而真菌表达体系较少,但由于很多真核生物基因不能在细菌中表达出具有生物活性的功能蛋白,因此研究和探索多种真菌作为宿主来表达外源基因将是一项重要任务;另一方面需要完善宏基因组学与生物信息学的结合。获得宏基因组复杂的序列信息是一项巨大的工程,利用传统的遗传学方法对其进行分析不仅效率低、成本高、且需花大量时间,而生物信息学的出现为宏基因组研究人员带来极大的便利。芯片技术已经用于检测土壤微生物群落的基因表达,序列技术和生物信息学的发展将会有更好的方法来分析宏基因组文库中的巨大数据。利用蛋白质矩阵和蛋白质学检测活性微生物将会在未来发挥重要作用。随着生物学的不断发展,文库构建和筛选策略的不断提高,异源基因表达能力和产量的增强,宏基因组学技术将成为研究环境微生物多样性,筛选新的功能基因和生物活性物质的重要手段之一。

## 参 考 文 献

[1] Torsvik V, Ovreas L. Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol*, 2002, **5**(3): 240–245.

[2] Handelsman J, Rondon MR, Brady SF, *et al.* Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, 1998, **5**(10): 245–249.

[3] Rees HC, Grant WD, Jones BE, *et al.* Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. *Extremophiles*, 2004, **8**(1): 63–71.

[4] Venter JC, Remington K, Heidelberg JF, *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 2004, **304**(5667): 66–74.

[5] Schloss PD, Handelsman J. Biotechnological prospects from metagenomics. *Curr Opin Biotechnol*, 2003, **14**(3): 303–310.

[6] Jeon CO, Park W, Padmanabhan P, *et al.* Discovery of a bacterium with distinctive dioxygenase that is responsible for in situ biodegradation in contaminated sediment. *Proc Natl Acad Sci USA*, 2003, **100**(23): 13591–13596.

[7] Radajewski S, Webster G, Reay DS, *et al.* Identification of active methylotroph populations in an acidic forest soil by Stable-isotope probing. *Microbiology*, 2002, **148**(8): 2331–2342.

[8] Galbraith EA, Antonopoulos DA, White BA. Suppressive subtractive hybridisation as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model. *Environ Microbiol*, 2004, **6**(9): 928–937.

[9] Brzostowicz PC, Walters DM, Thomas SM, *et al.* mRNA differential display in a microbial enrichment culture: Simultaneous identification of three cyclohexanone monooxygenases from three species. *Appl Environ Microbiol*, 2003, **69**(1): 334–342.

[10] Cramer R, Suter M. Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for the selection of functional gene products linked to the genetic information responsible for their production. *Gene*, 1993, **137**(1): 69–75.

[11] Demidov VV, Bukanov NO, Frank K-Amenetskii MD. Duplex DNA capture. *Curr Issues Mol Biol*, 2000, **2**(1): 31–35.

[12] Wu L, Thompson DK, Li GS, *et al.* Development and evaluation of functional gene arrays for the election of selected genes in the environment. *Appl Environ Microbiol*, 2001, **67**: 5780–5790.

[13] Zhou J, Bruns MA, Tiedje JM. DNA recovery from soils of diverse composition. *App Microbiol*, 1996, **62**(2): 316–322.

[14] Tsai YL, Olson BH. Rapid method for direct extraction of DNA from soil and sediments. *Applied and Environmental Microbiology*, 1991, **57**(4): 1070–1074.

[15] Wechter P, Williamon J, Robertson A, *et al.* A rapid, cost-effective procedure for the extraction of microbial DNA from soil. *World J Microbiol Biotechnol*, 2003, **19**(1): 85–91.

[16] Faegri A, Torsvik VL, Goksoyr J, *et al.* Bacterial and fungal activities in soil: separation of bacteria and fungi by a rapid fractionated Centrifugation technique. *Soil Biol Biochem*, 1977, **9**: 105–112.

[17] Jiang CJ, Wu B. Molecular cloning and functional characterization of a novel decarboxylase from uncultured microorganisms. *Biochemical and Biophysical Research Communications*, 2007, **357**(2): 421–426.

[18] Daniel R. The Metagenomics of soil. *Nature Reviews Microbiology*, 2005, **3**(6): 470–478.

[19] Lee CC, Accinelli R EK, Wagschal K, *et al.* Cloning and characterization of a cold-active xylanase enzyme from an environmental DNA library. *Extremophiles*, 2006, **10**(4): 295–300.

- [20] Mayumi D, Akutsu SY, Uchiyama H, *et al.* Identification and characterization of novel Poly(DL-lactic acid) depolymerases from metagenome. *Appl Microbiol Biotechnol*, 2008, **79**(5): 743–750.
- [21] 魏 萍, 宋文刚, 郝岗平. 泰山土壤宏基因组 DNA 中耐热  $\beta$ -半乳糖苷酶基因的克隆、表达及性质. *微生物学通报*, 2008, **35**(11): 1686–1690.
- [22] Gabor EM, Vries EJ, Janssen DB. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environmental Microbiology*, 2004, **6**(9): 948–958.
- [23] Park HJ, Jeon JH, Kang SG, *et al.* Functional expression and refolding of new alkaline esterase EM2L8 from deep-sea sediment metagenome. *Protein Expression and Purification*, 2007, **52**(2): 340–347.
- [24] Lee DG, Jeon JH, Jang MK, *et al.* Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. *Biotechnol Lett*, 2007, **29**(3): 465–472.
- [25] Jin D, Lu W, Ping SZ, *et al.* Identification of a new gene encoding EPSPS with high glyphosate resistance from the metagenomic library. *Curr Microbiol*, 2007, **55**(4): 350–355.
- [26] Xu MX, Xiao X, Wang FP. Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment. *Extremophiles*, 2008, **12**(2): 255–262.
- [27] 郭 鸿, 封 毅, 莫新春, 等. 水牛瘤胃宏基因组的一个新的  $\beta$ -葡萄糖苷酶基因的克隆 umceL3G 表达及其表达产物的酶学特性. *生物工程学报*, 2008, **24**(2): 232–238.
- [28] Song JS, Jeon JH, Lee JH, *et al.* Molecular characterization of TEM-type  $\beta$ -lactamases identified in cold-deep sediments of Edison Seamount. *The Journal of Microbiology*, 2005, **43**(2): 172–178.
- [29] Chu XM, He HZ, Guo CQ, *et al.* Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl Microbiol Biotechnol*, 2008, **80**(4): 615–625.
- [30] 朱雅新, 王加启, 马润林, 等. 荷斯坦奶牛瘤胃微生物元基因 BAC 文库的构建与分析. *微生物学报*, 2007, **47**(2): 213–216.
- [31] Rondon MR, August PR, Bettermann AD, *et al.* Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol*, 2000, **66**(6): 2541–2547.
- [32] Rees HC, Grant S, Jones B, *et al.* Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles*, 2003, **7**(5): 415–421.
- [33] Palackal N, Lyon CS, Zaidi S, *et al.* A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Appl Microbiol Biotechnol*, 2007, **74**(1): 113–124.
- [34] Cottrell MT, Moore TA, Kirchman DL. Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol*, 1999, **65**(6): 2553–2557.
- [35] Knietzsch A, Waschkwitz T, Bowien S, *et al.* Construction and screening of metagenomic libraries derived from enrichment cultures: Generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol*, 2003, **69**(3): 1408–1416.
- [36] Wilkinson DE, Jeanicke T, Cowan DA. Efficient molecular cloning of environmental DNA from geothermal sediments. *Biotechnol Lett*, 2002, **24**(2): 155–161.
- [37] Wang C, Meek DJ, Panchal P, *et al.* Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl Environ Microbiol*, 2006, **72**(1): 384–391.
- [38] Ono A, Miyazaki R, Sota M, *et al.* Isolation and characterization of naphthalene catabolic genes and plasmids from oil-contaminated soil by using two cultivation independent approaches. *Appl Microbiol Biotechnol*, 2007, **74**(2): 501–510.
- [39] Courtois S, Cappellano CM, Ball M, *et al.* Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol*, 2003, **69**(1): 49–55.
- [40] Martinez A, Kolvek SJ, Yip LT, *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. *Appl Environ Microbiol*, 2004, **70**(4): 2452–2463.
- [41] Henne A, Daniel R, Schmitzet RA, *et al.* Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol*, 1999, **65**(9): 3901–3907.
- [42] 吴杰群, 周文广, 杨登峰, 等. 从甘油富集菌宏基因组中克隆甘油脱水酶基因. *微生物学通报*, 2006, **33**(6): 52–56.
- [43] 蔡 莹, 陈秀珍, 杨克迁. 云南腾冲热泉土壤微生物基因组文库的构建与分析. *微生物学报*, 2006, **46**(3): 427–431.
- [44] He YZ, Fan KQ, Jia CJ, *et al.* Characterization of a Hyperthermostable Fe-superoxide dismutase from hot spring. *Appl Microbiol Biotechnol*, 2007, **75**(2): 367–376.
- [45] Wagner M, Nielsen PH, Loy A, *et al.* Linking microbial community structure with function: fluorescence in situ hybridization micro-autoradiography and isotope arrays.



- Current Opinion in Biotechnology*, 2006, **17**(1): 83–91.
- [46] Stokes HW, Holmes AJ, Nield BS, *et al.* Gene cassette PCR: sequence independent recovery of entire genes from environmental DNA. *Appl Environ Microbiol*, 2001, **67**(11): 5240–5246.
- [47] Hu Y, Zhang GM, Li AY, *et al.* Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. *Appl Microbiol Biotechnol*, 2008, **80**(5): 823–830.
- [48] 许跃强, 段承杰, 周权能, 等. 造纸废水纸浆沉淀物中未培养微生物纤维素酶基因的克隆和鉴定. *微生物学报*, 2006, **46**(5): 783–788.
- [49] Tirawongsaroj P, Sriprang R, Harnpicharnchai P, *et al.* Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring Metagenomic library. *Journal of Biotechnology*, 2008, **133**(1): 42–49.
- [50] Brady SF, Clardy J. Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J Am Chem Soc*, 2000, **122**(51): 12903–12904.
- [51] 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 Bacteriol*, 2001, **183**(22): 6645–6653.
- [52] Uchiyama T, Abe T, Ikemura T, *et al.* Substrate induced gene expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol*, 2005, **23**(1): 88–93.
- [53] Williamson LL, Borlee BR, Schloss PD, *et al.* Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl Environ Microbiol*, 2005, **71**(10): 6335–6344.
- [54] Martín-Cuadrado AB, López-García P, Alba JC, *et al.* Metagenomics of the deep mediterranean, a Warm Bathypelagic Habitat. *PLoS-ONE*, 2007, **2**(9): 9–14.
- [55] Zhang Y, Dong J, Yang Z, *et al.* Phylogenetic diversity of nitrogen-fixing bacteria in mangrove sediments assessed by PCR–denaturing gradient gelelectrophoresis. *Arch Microbiol*, 2008, **190**(1): 19–28.
- [56] 张 薇, 胡跃高, 黄国和, 等. 西北黄土高原柠条种植区土壤微生物多样性分析. *微生物学报*, 2007, **47**(5): 751–756.
- [57] 肖 凯, 曹理想, 陆勇军, 等. 广东金山温泉沉积物中原核与真核微生物多样性初步分析. *微生物学报*, 2008, **48**(6): 717–724.
- [58] Fierer N, Breitbart M, Nulton J, *et al.* Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol*, 2007, **73**(21): 7059–7066.
- [59] Kim KH, Chang HW, Nam YD, *et al.* Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Applied and Environmental Microbiology*, 2008, **74**(19): 5975–5985.
- [60] Sabet S, Chu WP, Jiang SC. Isolation and genetic analysis of haloalkaliphilic bacteriophages in a north American Soda lake. *Microbial ecology*, 2006, **51**(4): 543–554.
- [61] Breitbart M, Salamon P, Andresen B, *et al.* Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA*, 2002, **99**(22): 14250–14255.
- [62] Breitbart M, Hewson I, Felts B, *et al.* Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol*, 2003, **185**(20): 6220–6223.
- [63] Cann A, Fandrich S, Heaphy S. Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes. *Virus Genes*, 2005, **30**(2): 151–156.
- [64] Wang GY, Graziani E, Waters B, *et al.* Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett*, 2000, **2**(16): 2401–2404.
- [65] Brady SF, Clardy J. Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water. *J Nat Prod*, 2004, **67**(8): 1283–1286.
- [66] MacNeil IA, Tiong CL, Minor C, *et al.* Expression and isolation of anti microbial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol*, 2001, **3**(2): 301–308.
- [67] Gillespie DE, Brady SF, Bettermann AD, *et al.* Isolation of anti-biotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol*, 2002, **68**(9): 4301–4306.
- [68] DiazTorres ML, McNab R, Spratt DA, *et al.* Novel tetracycline resistance determinant from the oral metagenome. *Antimicrob Agents Chemother*, 2003, **47**(4): 1430–1432.
- [69] Mori T, Mizuta S, Suenaga H, *et al.* Metagenomic screening for bleomycin resistance genes. *Appl Environ Microbiol*, 2008, **74**(21): 6803–6805.
- [70] 赵 晶, 杨祥胜, 曾润颖. 南极土壤微生物宏基因组文库构建及其抗肿瘤活性初探. *自然科学进展*, 2007, **17**(2): 267–271.
- [71] Kurokawa K, Itoh T, Kuwahara T, *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Research*, 2007, **14**(4): 169–181.
- [72] Finkbeiner SR, Allred AF, Tarr PI. Metagenomic analysis of human diarrhea: Viral detection and discovery. *PLoS Pathogens*, 2008, **4**(2): 1–9.