

江西典型红壤区土壤细菌基因组文库构建 及功能初步分析*

黄婷婷 崔中利 张璐 曹慧[†] 李顺鹏

(南京农业大学农业部农业环境微生物工程重点开放实验室, 南京 210095)

CONSTRUCTION OF TYPICAL RED SOIL-DERIVED METAGENOMIC LIBRARY AND FUNCTIONAL ANALYSIS IN JIANGXI PROVINCE

Huang Tingting Cui Zhongli Zhang Lu Cao Hui[†] Li Shunpeng

(Key Laboratory of Microbiological Engineering of Agricultural Environment, Ministry of Agriculture, Nanjing Agricultural University,
Nanjing 210095, China)

关键词 红壤; DNA 提取; 基因组文库; 序列分析

中图分类号 S154.3; Q938.1 文献标识码 A

地球上的微生物,仅原核生物细胞总数就大约在 $4 \times 10^{30} \sim 6 \times 10^{30}$, 包含的独立基因型在 $10^6 \sim 10^8$ 种之间^[1]。因此,微生物所具备的分类、功能、遗传和系统发育等多样性在维持生物圈生态平衡和为人类提供资源方面起着重要的作用^[2]。然而长期以来由于受到研究方法和手段的限制,土壤中绝大多数微生物的功能还不清楚。环境样品的宏基因组学 (metagenomics) 是对环境样品中微生物群体基因组进行的分析,该方法在众多用于获得未培养微生物的生理和遗传特性的方法中,逐渐显示出强大的优势^[3,4]。国外已有从不同环境样品的宏基因组文库中,通过功能性筛选获得新的抗生素、脂肪酶、丁质酶、膜蛋白、4-羟基丁酸脱氢酶等合成基因或相关基因,以及对群落结构组成和功能进行分析的报道^[5~9]。国内也逐步开始有相关研究^[10,11]。

红壤在中国南方地区分布广泛。本文以江西典型红壤为研究材料,用间接法提取土壤微生物总 DNA 后直接酶切,利用 TA 克隆方法建立土壤微生物宏基因组文库,随机挑取克隆进行单向测序并结

合生物学软件初步分析,预测其中包含的基因序列及相应蛋白功能。

1 材料与方法

1.1 研究区域和土壤

研究区域位于江西省鹰潭市余江县刘家站,该地区为中亚热带温暖湿润的季风气候,年均温 17.8°C , 年均降雨量 1785 mm , 主要集中在 4~6 月。地形为低丘岗地,海拔 $35.0 \sim 54.6\text{ m}$, 坡度为 $5^\circ \sim 8^\circ$ ^[12]。成土母质为第四纪红粘土,土地利用类型为荒草地。土壤 pH 为 4.41,全氮 1.32 g kg^{-1} ,全磷为 0.67 g kg^{-1} ,有机质含量为 20.99 g kg^{-1} 。五点混合法采集土壤样品,采样深度为 0~5 cm、5~15 cm、15~30 cm (分别标注为样品 a, 样品 b, 样品 c)。土壤采集后用布袋封装,以增加通气性。样品带回室内,除去根系、石块等杂物,再过 4 mm 筛后取样用于土壤的细菌、放线菌和真菌三大菌群计数,其余保存于 -70°C 冰箱用于土壤总 DNA 提取。表 1 为三大菌群的计数结果。

* 国家自然科学基金项目 (40371069)、中国科学院知识创新工程项目 (KZCX3-SW-427) 和中国博士后科学基金项目 (2003033495) 资助

[†] 通讯作者, E-mail: hcao@njau.edu.cn

作者简介:黄婷婷 (1982~), 女, 江苏泰兴人, 硕士研究生, 主要从事土壤微生物分子生态方面的研究

收稿日期: 2005-12-01; 收到修改稿日期: 2006-08-09

表1 红壤微生物三大菌群计数结果(单位:cfu g⁻¹干土)

样品编号	剖面深度(cm)	细菌	放线菌	真菌
a	0~5	8.2 ×10 ⁷	2.3 ×10 ⁷	1.34 ×10 ⁶
b	5~15	4.75 ×10 ⁷	9.9 ×10 ⁶	2.45 ×10 ⁵
c	15~30	1.55 ×10 ⁷	2.65 ×10 ⁶	2.5 ×10 ⁴

注:a,b,c为土壤样品。编号样品a为0~5 cm土层,样品b表示5~15 cm土层,样品c表示15~30 cm土层

1.2 主要菌株和试剂

pUC 18 质粒(Amp⁺)、大肠杆菌 DH5 均由本实验室保存; *Taq* 酶、限制性内切酶、TA 克隆载体 pMD18-T Vector 购自 Takara 公司, 氨苄青霉素(Amp)、溶菌酶、CTAB(十六烷基三甲基溴化铵)、SDS(十二烷基硫酸钠)、Tris(三羟甲基氨基甲烷)、EDTA 二钠盐购自南京赛吉公司; 蛋白酶 K 为德国 Merk 公司产品; 焦磷酸钠购自上海化学试剂厂。

1.3 土壤微生物总 DNA 的提取和质量检测

土壤微生物总 DNA 提取采用间接法 Blending method^[13]。

采用紫外分光光度法测波长 260 nm 的吸光度值对土壤总 DNA 进行定量,同时测定波长 230 nm 和 280 nm 的吸光度值以计算 DNA 的纯度。限制性内切酶酶切、细菌的 16S rDNA 通用引物扩增土壤总 DNA 以此检测 DNA 的提取质量。方法为常规分子生物学操作,参照文献[14]。

1.4 基因组文库的构建

1.4.1 基因组文库构建流程图 见图 1。

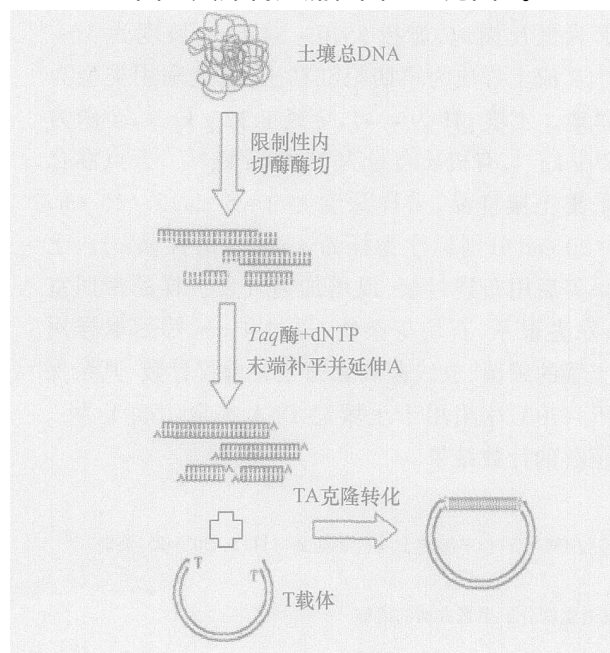


图1 建库流程

1.4.2 外源插入片段的制备 用限制性内切酶 *Bam*HI 直接酶切适量的土壤基因组 DNA,控制酶切时间和酶浓度使酶切后的片段大小相对集中在合适大小范围内。DNA 回收试剂盒切胶回收 2~6 kb 大小的片段。对回收后的 DNA 片段进行 45 min 的扩增延伸反应,利用 *Taq* 酶能使扩增片段末端加脱氧腺苷的特性使每个片段的粘性末端都被补平并带 dA。反应体系如下:10 × *Taq* 酶反应缓冲液 5 μl, Mg²⁺ (25 mmol L⁻¹) 4 μl, dNTP (25 mmol L⁻¹) 5 μl, DNA 回收片段 10 μl, *Taq* 酶 (5 U μl⁻¹) 0.5 μl, 加灭菌超纯水至 50 μl。反应条件:94 变性 3 min;72 延伸 45 min;10 10 min。

1.4.3 文库构建、保存 Inoue 法制备大肠杆菌 DH5 的高效感受态细胞^[14];外源片段与 T 载体 (pMD-18, Takara) 16 酶连过夜后,转化 DH5 高效感受态细胞,37 培养 12~14 h 后蓝白斑法筛选含有重组质粒的阳性克隆;挑取白色的转化子转接至含氨苄青霉素的 LB 固体培养基上,过夜培养后用无菌水洗下平板上的菌落,等体积加 30% 甘油混匀后于 -70 保存。

1.5 DNA 文库的初步分析

从文库中随机挑取转化子少量提取质粒验证插入片段大小和插入效率;挑取部分转化子,从插入片段的一端进行单向测序一个反应,然后对所得序列在微生物基因组数据库中用 BLAST 进行序列检索和同源性分析。结合 BioEdit、ClustalX 等生物软件分析预测可能包含的开放阅读框 (ORF) 及其推测蛋白质的功能。

1.6 克隆序列的 GenBank 登录号

将经同源性比对后确认的新序列,登录 GeneBank 并提交。编号为 H3, H5, H10, H2-2, H2-3, H2-7, X6, Q41 的克隆插入序列,序列号 (Accession number) 分别为: DQ185049, DQ185050, DQ185051, DQ185052, DQ185053, DQ195083, DQ195084, DQ195085。

2 结果与分析

2.1 红壤中细菌总 DNA 的提取和纯度检测

从图 2 可看到获得土壤细菌总 DNA 片段大小超过 23 kb。分别用凝胶电泳法和紫外扫描法进行核酸定量,最终确定从每克干土中得到的 DNA 为 6.85 ± 0.24 μg,在片段大小和浓度上符合后续操作的要求。从土壤中提取的微生物基因组 DNA 中蕴含着腐殖酸、多糖等杂质会影响酶切、PCR 等分子生

物学操作,因此能否进行后续的可作为 DNA 纯度的评价标准^[15,16]。此外 A_{260}/A_{230} 和 A_{260}/A_{280} 的比值也可作为总 DNA 中腐殖酸和蛋白质含量的评判标准^[17]。尽管本研究采用的方法所得总 DNA 计算出的比值不高(表 2),但直接或经过适当稀释后即可用于酶切和 16S rDNA 扩增,因此可以认为提取 DNA 的纯度已足够用于后续操作。考虑到土壤中微生物的含量及外界环境因素可能造成对 0~5 cm 土层的干扰,我们选择从 5~15 cm 土层中提取的总 DNA 进行宏基因组文库的构建。

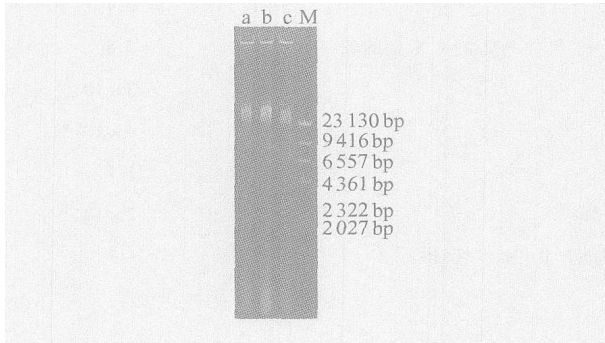


图 2 间接法提取的微生物总 DNA

a, b, c: 三种土壤样品提取的 DNA; M 为 DNA/ *Hind* Marker

表 2 DNA 样品的产量及纯度参数

样品编号	A_{260}/A_{230}	A_{260}/A_{280}	DNA $\mu\text{g g}^{-1}$ 干土
a	1.065	1.001	7.56
b	1.062	1.433	6.85
c	1.015	1.374	1.22

土壤总 DNA 是混合的基因组 DNA, 从不同生境中得到 DNA 的 G+C% 等都存在差异, 因此不同的限制性内切酶对总 DNA 的酶切敏感程度是不同的。用四种限制性内切酶 (*Pst*, *EcoR*, *BamH* 和 *Hind*) 对红壤总 DNA 进行酶切试验后发现, *BamH* 的酶切效果最好(图 3)。

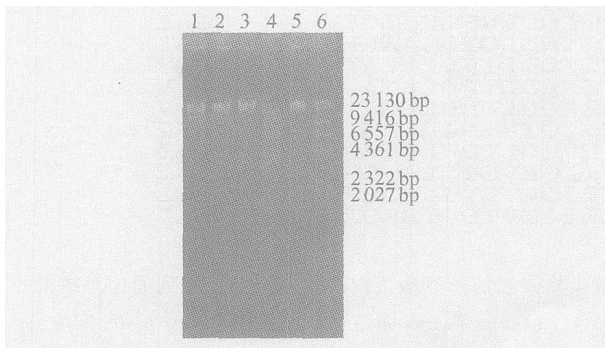


图 3 四种限制性内切酶酶切总 DNA

1: *Pst* 酶切; 2: *EcoR* 酶切; 3: *Hind* 酶切; 4: *BamH* 酶切; 5: 未酶切的 DNA; 6: DNA/ *Hind* Marker

2.2 红壤细菌基因组文库的构建

常规建库过程是将酶切片段与相应酶切、脱磷酸化的载体连接后转化大肠杆菌。但脱磷酸化后载体连接效率降低, 在提取得到总 DNA 浓度较低的前提下, 转化子数量会很少。而 TA 克隆的连接效率要高于载体脱磷后再连接。此外, 由于土壤性质、提取方法等因素的影响, 对该土壤提取的总 DNA 中包含有小片段的 DNA(图 2 的电泳泳道中 DNA 拖尾现象), 对这样的 DNA 片段酶切后难以确保其末端为相应的粘性末端, 从而与载体准确连接, 这必然会降低酶连效率, 造成 DNA 片段的损失。而 TA 克隆的方法构建文库可有效的避免这一个问题^[18]。根据预酶切的实验结果, 选择 *BamH* 对总 DNA 进行部分酶切后, 回收 2~6 kb 大小的酶切片段, 利用 *Taq* 酶能使 PCR 产物末端带一个 dA 碱基的特性, 对酶切片段进行末端补平并加 dA, 然后与 T 载体相连从而避开载体脱磷造成的效率降低的影响。采用该方法, 初步估计从 1g 土壤中能得到约 15 000 个克隆。通过随机挑取阳性克隆少量提取质粒检测, 发现 80% 以上克隆都含有插入片段(图 4)。与酶切片段与相应酶切后脱磷酸化的载体连接后转化大肠杆菌相比, 效率大大提高。

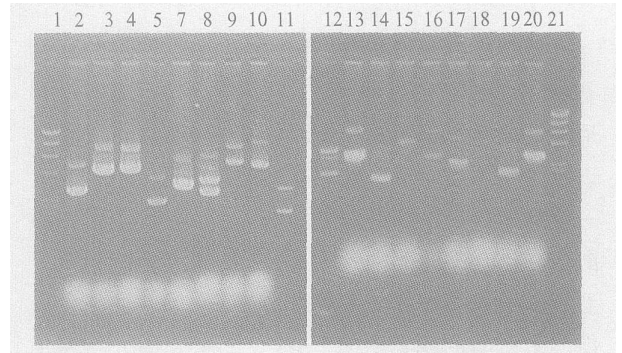


图 4 随机挑取阳性克隆提取质粒电泳

1, 21: DNA/ *Hind* Marker; 2~10, 13~20: 随机挑选的 16 个转化子; 11, 12: 对照线性 pUC18 质粒

2.3 克隆基因功能的初步分析

从文库中随机挑取 10 个克隆, 对插入片段的一端进行单向测序, 并在微生物基因组数据库中用 BLAST 程序对所得测序结果进行核酸序列检索和同源性分析。在这 10 个序列中, 除 2 个为载体 DNA 的部分序列外, 其余 8 个序列均不存在于现有的基因组数据库中, 应为新序列。

通过生物学软件 BioEdit 和 ClustalX 分析预测序列中可能包含的开放阅读框 (Opening reading frame,

ORF),根据氨基酸序列预测其中所包含的开放阅读框 或部分阅读框可能编码蛋白质产物及其功能(表 3)。

表 3 随机挑取克隆中插入片段部分序列中 ORF 预测及可能的基因功能

样品编号	登录序列号	推测可能的 ORF ¹⁾	预测功能	S 分值 ²⁾	E 值 ³⁾
H2-2	DQ185052	Frame 1, 239aa	未命名的蛋白质产物	33.1	8.4
		Frame-1, 98aa	蛋白激酶	172	2 ×10 ⁻⁴²
		Frame-2, 140aa	LcmF,细菌中保守存在的一种未知蛋白	120	1 ×10 ⁻²⁶
H2-3	DQ185053	Frame 1, 117aa	丝氨酸蛋白酶	106	2 ×10 ⁻²²
H2-7	DQ195083	Frame 1, 169aa	含有 WD40 重复蛋白的蛋白激酶	86.7	3 ×10 ⁻¹⁶
		Frame-1, 74aa	转录激活蛋白	32.7	3.2
		Frame-2, 129aa	推测的 DNA 聚合酶 I	33.1	2.4
H3	DQ185049	Frame 1, 87aa	类胰岛素的丝氨酸蛋白酶,有一个包含 PDZ 的 C 末端的域	33.1	2.8
		Frame-1, 87aa	依赖于 ATP 的 DNA 连接酶	79.7	3 ×10 ⁻¹⁴
H5	DQ185050	Frame 1, 74aa	LacOPZ- 肽	59.3	4 ×10 ⁻⁸
H10	DQ185051	Frame 1, 131aa	推测的与溶血相关的膜蛋白	34.3	1.1
		Frame-1, 131aa	未培养细菌中推测的 -内酰胺酶	67	2 ×10 ⁻¹⁰
		Frame 3, 68 aa	与 ATP 结合 cassette 类似,sub-family B (MDR/ TAP)	32.3	4.1
X6	DQ195084		**		
Q41	DQ195085		**		

1) 可能的 ORF 由软件分析预测,其中 frame 1, 2 分别表示正向序列推测的 ORF,而 frame-1, - 2 分别表示反向序列预测得到的可能的 ORF。
 ** 表示软件分析包含可能的 ORF 但通过氨基酸序列比对未能找到相应功能的蛋白质。2) S 分值(bits score)用于描述不同的比对得分系统的统计学特征以反映比较序列间的相似程度,分值越高,相似程度越高。3) E 值(E value)代表被比对的两个序列不相关的可能性,E 值越低序列的相似性越大;表格中为比对结果中 S 分值最高和 E 值最低的对象

从表 3 中可以看到,通过对序列正向和反向的氨基酸序列比对得到的开放阅读框中,很多与已报道的诸如激酶、蛋白酶、DNA 连接酶、溶血相关蛋白以及一些功能尚不明确的蛋白的部分氨基酸序列有一定同源性。例如,在样品 H2 - 3 中包含编码丝氨酸蛋白酶基因的部分序列,并且从比对结果中可以看到其编码产物的氨基酸序列与其他微生物中的丝氨酸蛋白酶的同源性在 29 % ~ 55 % 之间。与亲缘

关系不同的 9 种微生物的丝氨酸蛋白酶序列比对(图 5) 可看出某几个氨基酸位点的保守性是很高的。由于所得到的序列长度有限,在这段氨基酸序列中,我们并没有能找到丝氨酸蛋白酶的活性中心保守组分,即由丝氨酸(S)、组氨酸(H)和天冬氨酸(D)三者构成一个电荷中继网,但若库容足够大,或者通过构建大片的文库都将会增加获得完整基因的可能性。

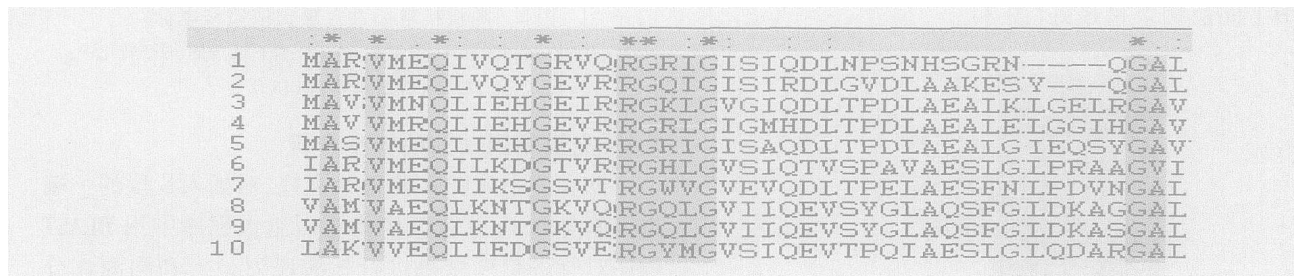


图 5 H2-3 样品可能编码的一个蛋白的部分氨基酸序列与 9 种微生物来源的丝氨酸蛋白酶相应氨基酸序列的比对结果
 1: H2-3; 2: *Bradyrhizobium japonicum* USDA 110; 3: *Magnetospirillum magnetotacticum*; 4: *Paracoccus denitrificans*; 5: *Rhodospirillum rubrum*; 6: *Rhizobium* sp.; 7: *Neisseria meningitidis* Z2491; 8: *Neisseria gonorrhoeae* FA 1090; 9: *Thiobacillus denitrificans* ATCC 25259; 10: *Mesorhizobium* sp. BNC1

从样品 H2-2 反向序列中预测的一个 ORF 所编码的氨基酸序列为一种蛋白激酶的一部分。我们将

该氨基酸序列分别与几种革兰氏阴性细菌(如副球菌属、产硫杆菌)、革兰氏阳性细菌(芽胞杆菌属、

酸乳杆菌属、链霉菌属、诺卡丁氏菌属)中该蛋白酶的相应部分氨基酸序列进行分析比对(图6)。可以看到,在同一属中该蛋白的相应部分氨基酸序列的保守性是相当高的,亲缘关系不同的微生物间也存在着多个高度保守的氨基酸位点。

有趣的是,氨基酸序列检索结果显示它与酸杆菌类群(*Acidobacteria* group)中一分类地位尚未明确界定的细菌(*Solibacteres*)相应氨基酸序列的同源性

达到81%,而与其他微生物来源的序列仅有40%~50%的同源性。基于16S rDNA的方法在世界范围内的不同生境中均检测到酸杆菌类群的存在,对其生理学信息却知之甚少。目前越来越多的研究证实它们对维护红壤酸性环境具有重要的作用和生态学意义。可以预想,通过构建宏基因组文库的方法可以为我们认识土壤微生物的构成及其功能提供一个新的视角。

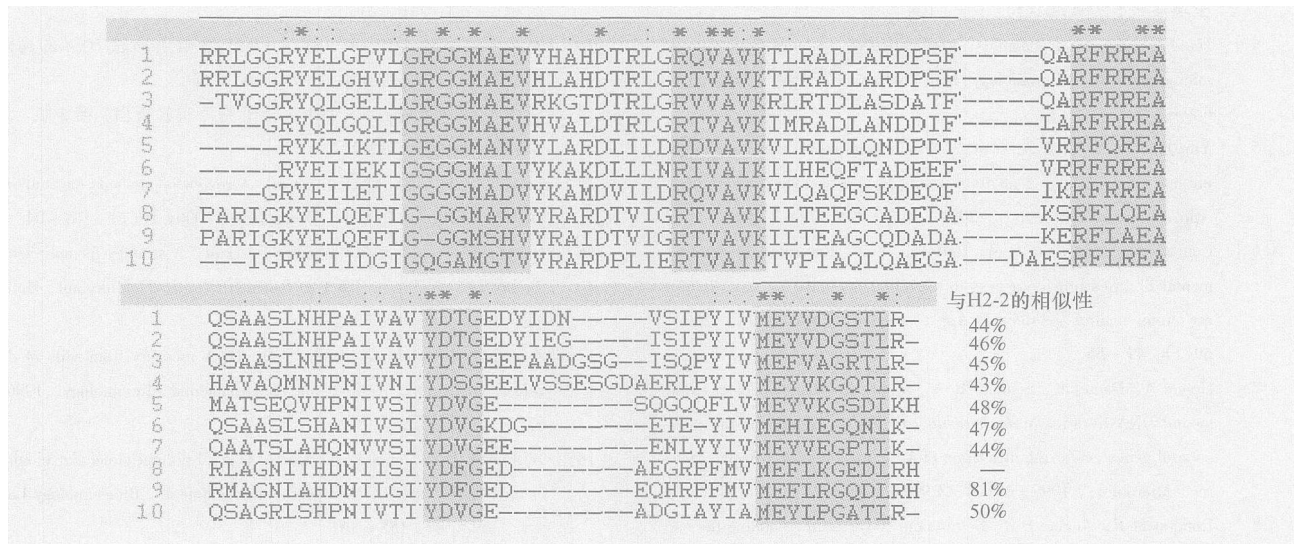


图6 样品H2-2反向序列预测的ORF编码蛋白与亲缘关系不同的蛋白激酶部分序列比对结果

1: *Streptomyces coelicolor* A3(2); 2: *Streptomyces avemitilis* MA-4680; 3: *Nocardioideis* sp. JS614; 4: *Bifidobacterium longum* DJO10A; 5: *Lactobacillus casei* ATCC 334; 6: *Desulfotobacterium hafniense* DCB-2; 7: *Bacillus halodurans* C-125; 8: H2-2; 9: *Solibacter usitatus* Elin6076; 10: *Thiobacillus denitrificans* ATCC 25259

3 讨论

从环境样品中提取高质量的基因组DNA是构建宏基因组文库的前提。提取方法可分为直接法(细胞原位裂解)和间接法(细胞回收)两种。目前的文献报道一般采用直接法提取土壤总DNA。直接法提取得到DNA的量较高,但其中包含的腐殖酸、粘粒和多糖等抑制物含量也相对较高,这些物质会干扰酶切、PCR扩增、杂交等后续操作;且直接法得到的总DNA包含了真核生物以及土壤中游离的DNA,它们不一定能在原核宿主中表达,却增加了为得到目的基因文库所需达到的库容^[13]。间接法得到主要是原核生物体的DNA,这有利于以大肠杆菌为宿主的文库构建。

在从土壤中提取总DNA的过程中,由于机械损伤、试剂等作用以及土壤本身性质的影响,大片的DNA分子不可避免的会受到损伤而断裂成10 kb以

下甚至更小的片段,且多为平末端,这显然不利于片段与脱磷载体相连,而经过PCR反应延伸加dA尾后可以方便的与末端为dT的载体相连,减少这部分核酸的损失。相对于脱磷载体而言,T载体的连接效率更高,这为高效构建文库提供了一个更为可行的方法。

南方红壤地区是我国热带、亚热带经济林果、经济作物和粮食生产的重要基地,但随着人口、资源、环境之间矛盾的加剧,以及红壤本身所具有的富铝化、酸化、铁质化及抗蚀性弱等物质循环特征与规律,造成该地区当前生态环境系统及土壤退化问题极为严重^[19]。另外,尽管红壤中的生物量不高,但并不意味其中的生物多样性程度低,中国第一个有自主知识产权的生物农药——井冈霉素即来源于红壤。因此,红壤中可能蕴含着新的、还不为人知的基因资源。通过构建红壤的宏基因组文库来研究其微生物群落结构,能揭示出微生物群落结构与红壤理化性质之间的内在联系,从而改善红壤当前的生

态环境,并有利于获得新的、有用的基因资源。

参考文献

- [1] Whitman W B, Coleman D C, Wiebe W J. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci.*, 1998, 95: 6 578 ~ 6 583
- [2] Torsvik V, Ovreas L. Microbial diversity and function in soil: From gene to ecosystems. *Current Opinion in Microbiology*, 2002, 5: 240 ~ 245
- [3] 马万里, Tibbits J, Adams M. 土壤微生物多样性研究的新方法. *土壤学报*, 2004, 41 (1): 103 ~ 108
- [4] Handelsman J. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Biology Reviews*, 2004, 68: 669 ~ 685
- [5] Yun J, Kang S, Park S, *et al.* Characterization of a novel amyolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl. Environ. Microbiol.*, 2004, 70 (12): 7 229 ~ 7 235
- [6] Courtois S, Cappellano C M, 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
- [7] Henne A, Daniel R, Schmitz R A, *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): 3 901 ~ 3 907
- [8] Rondon M R, August P R, Bettermann A D, *et al.* Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.*, 2000, 66 (6): 2 541 ~ 2 547
- [9] Tyson G W, Chapman J, Hugenholtz P, *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 2004, 428: 37 ~ 43
- [10] 王啸波, 唐玉秋, 王金华, 等. 环境样品中 DNA 的分离纯化和文库的构建. *微生物学报*, 2001, 41 (2): 133 ~ 140
- [11] 饶志明, 赵有玺, 李辉, 等. 太湖流域土壤微生物基因组总 DNA 分离纯化及其质粒文库的初步构建. *应用与环境生物学报*, 2004, 10 (6): 774 ~ 777
- [12] 雷志栋, 杨诗秀, 谢森传. *土壤水动力学*. 北京: 清华大学出版社, 1988. 231 ~ 236
- [13] 黄婷婷, 曹慧, 崔中利, 等. 一种土壤微生物总 DNA 的高效提取方法. *土壤*, 2004, 36 (6): 662 ~ 666
- [14] 萨姆布鲁克 J, 拉塞尔 D W. *分子克隆实验指南*. 第 3 版. 北京: 科学出版社, 2002
- [15] Liu B R, Jia G M, Chen J, *et al.* A review of methods for studying microbial diversity in soils. *Pedosphere*, 2006, 16 (1): 18 ~ 24
- [16] Burgmann H, Pesaro M, Widmer F, *et al.* A strategy for optimizing quality and quantity of DNA extracted from soil. *J. Microbiol. Methods*, 2001, 45 (1): 7 ~ 20
- [17] Zhou J Z, Bruns M A, Tiedje J M. DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, 1996, 62 (2): 316 ~ 322
- [18] Wilkinson D E, Jeanicke T, Cowan D A. Efficient molecular cloning of environmental DNA from geothermal sediments. *Biotechnology Letters*, 2002, 24: 155 ~ 161
- [19] 赵其国. *红壤物质循环及其调控*. 北京: 科学出版社, 2002. 6 ~ 10