

云南腾冲热泉土壤微生物基因组文库的构建与分析

蔡莹^{1,2} 陈秀珍¹ 杨克迁¹ 黄力¹ 董志扬^{1*}

(¹ 中国科学院微生物研究所 微生物资源前期开发重点实验室 北京 100080)

(² 中国科学院研究生院 北京 100049)

摘 要 采用冻融、蛋白酶 K、SDS-高盐-加热处理法联合的方法,直接从云南腾冲地区的一个弱碱性高温热泉沉积样品中提取和分离环境混合基因组 DNA,产量为每克样品 1~2 μ g DNA,用 Promega 试剂盒纯化后进行 *Pst* I 部分酶切处理,电泳回收 3~8kb 的片段后,构建了 pSK(+) 为载体的基因组文库,共获得 25000 个阳性克隆,平均插入片段长度为 4.6kb。通过随机 DNA 序列测定和基因注释,发现外源插入片段含有未见报道的序列。

关键词 免培养技术;未培养微生物;环境基因组文库

中图分类号:Q78,Q93 文献标识码:A 文章编号:1001-6209(2006)03-0427-05

高温热泉与地球早期环境比较接近,其中微生物生态系统相对简单、稳定和封闭,对其进行微生物研究不仅有助于认识高温环境中的微生物遗传多样性与功能多样性,而且对认识地球上生命的起源和进化具有十分深远的意义,同时,由于这些微生物具有特殊的生理机制和独特的基因,具有较高的应用价值,因此受到广泛的重视^[1]。

我国云南省腾冲县具有特殊的地热资源,是研究热泉微生物的重要地区^[2]。国内对腾冲热泉微生物的研究开展得比较早,在菌种分离鉴定^[3-6]、微生物多样性分析等方面取得了许多成果^[7,8]。而在通过构建环境基因组文库收集和发掘其微生物、特别是未培养微生物基因资源方面,尚处于起步阶段。

腾冲蛤蟆嘴热泉是一个弱碱性的高温热泉,温度达到 84 $^{\circ}$ C 以上,本研究采用改进的 DNA 提取方法,直接提取热泉底部土壤 DNA,成功构建了蛤蟆嘴热泉未培养微生物基因组文库,并对文库中的部分克隆进行了分析,为探索该热泉微生物新基因的研究和利用奠定基础。

1 材料和方法

1.1 材料

1.1.1 品采集 2004 年 3 月在云南腾冲县境内蛤蟆嘴热泉采集泉底土壤(灰色沙土),其采样点海拔 1450 米,北纬 24 $^{\circ}$ 57'00.5",东经 98 $^{\circ}$ 26'17.2",采样温度 84 $^{\circ}$ C,pH 8.0。样品采集后密封运回实验室于

-80 $^{\circ}$ C 冰箱保存。

1.1.2 菌株和载体的来源: *Escherichia coli* DH10B 和 pSK(+) 均为本实验室保存。

1.1.3 主要试剂: *Pst* I 等限制性内切酶购自 TaKaRa 公司;CIP(肠小牛碱性磷酸酶)等购自 NEB 公司;DNA 纯化试剂盒、T4 DNA 连接酶购自 Promega 公司;蛋白酶 K 购自 Merck 公司,其余均为国产分析纯试剂。

1.2 土壤总 DNA 的提取和纯化

1.2.1 土壤 DNA 的提取: 根据高温热泉样品特性,参考文献[9~12]方法并作了改进:称取 5g 土壤样品充分研磨后,置于 30mL 离心管中,加入 13.5mL DNA 抽提缓冲液(100mmol/L Tris·HCl,pH 8.0,100mmol/L EDTA,pH 8.0,10mmol/L 磷酸钠缓冲液,pH 8.0,1.5mol/L NaCl,1% CTAB),混匀后,置液氮中,然后取出在 65 $^{\circ}$ C 水浴中保温至融化,反复 3 次;冷却后加入 50 μ L 蛋白酶 K(20mg/mL)于摇床中水平振荡(37 $^{\circ}$ C,225r/min)约 30min 后,加入 2.0mL SDS(20%),混匀后,65 $^{\circ}$ C 保温 2~3h,每隔 15~20min 上下颠倒离心管混匀,7000 \times g 室温离心 10min,收集上清液,沉淀加入 5mL DNA 提取液,0.5mL 20%SDS,与土壤充分混匀,置 65 $^{\circ}$ C 20min,离心收集上清液,重复一次,合并 3 次上清液,加入等体积的氯仿:异戊醇(1:1)抽提两次,加入 0.6 倍体积的异丙醇,室温放置过夜后,12000 \times g,4 $^{\circ}$ C 离心 15min,沉淀用 70%乙醇洗后干燥,用 100 μ L TE(pH 8.0)溶解。-20 $^{\circ}$ C 保存备用。

基金项目:国家“863 计划”(2004AA21480)

* 通讯作者。Tel/Fax:86-10-62551206;E-mail:dongzy@sun.im.ac.cn

作者简介:蔡莹(1981-)女,湖南人,硕士研究生,主要从事微生物学研究。E-mail:caiying81@163.com

收稿日期:2005-08-04;接受日期:2005-08-17;修回日期:2005-09-22

1.2.2 粗提 DNA 的纯化 按产品说明进行操作。

1.3 载体的制备

构建 *Pst* I 酶单酶切的 pSK(+) 质粒。

1.4 混合基因组 DNA 文库的构建和保存

首先用 *Pst* I 对纯化的环境微生物 DNA 进行不完全酶切,琼脂糖电泳分离酶切产物,用 Qiagen Quick Gel Extraction Kit 凝胶回收 3~8kb DNA 片段,然后与经 CIP 去磷酸化酶处理过的 pSK(+) 质粒 DNA 以合适的比例连接,连接产物用电转化方法导入 *E. coli* DH10B 涂布在含有 X-Gal (40 μ g/mL) IPTG (40 μ g/mL) Amp (50 μ g/mL) 的 LB 平板上。挑取白色菌落至含有 LB 培养基 (Amp 浓度同上) 的 96 孔培养板中,37 $^{\circ}$ C 恒温培养 24h 左右,加入甘油,保存于 -80 $^{\circ}$ C。

1.5 文库的质量分析

从 LB 平板上随机挑取 100 个白色菌落液体培养,用碱裂解法提取质粒,酶切、鉴定插入片段的大小。随机抽出 30 个阳性克隆进行测序,将序列递交到 GenBank,通过 BLASTX 软件进行同源性分析。

2 结果和分析

2.1 环境样品 DNA 的提取和纯化

采用改进的 DNA 提取方法,从每克土壤样品可获得 1~2 μ g DNA。电泳结果表明,提取的 DNA 完整性较好,大于 23kb,满足本研究中基因组文库构建的要求。经过电洗脱法、低熔点凝胶法以及试剂盒法等方法的比较,发现 Promega Wizard Clean Up 试

剂盒纯化方法效果比较理想,回收效率高,大约为 90%,降解程度低,回收后 DNA 纯度较高。

2.2 文库的构建

经纯化后的热泉环境微生物 DNA 能够被 *Pst* I 酶切,得到 1~20kb 范围的弥散 DNA 条带。用试剂盒凝胶回收其中 3~8kb 大小的酶切片段,将回收的 DNA 片段 (25~30ng) 与制备好的载体 (10ng) 进行连接,电转化宿主细胞 *E. coli* DH10B,将其铺在 Amp/IPTG/X-gal/LB 平板上进行培养 16h,获得带外源 DNA 片段的白色阳性克隆。

2.3 文库的评价

评价一个环境免培微生物基因组文库质量的好坏主要从以下 3 个方面考虑:文库的库容量、插入片段的大小、文库中基因的多样性。

在热泉环境微生物基因文库的构建中,每次连接反应,以 25~30ng 目的片段 DNA 连接后转化得到大约 25000 个白斑,转化效率为 10^6 个阳性克隆子/ μ g DNA。若将提取纯化的该热泉环境微生物 DNA 全部连接转化,可总共获得约 500000 个阳性克隆;随机从文库中挑取 100 个白斑,提取质粒酶切粒检测,其中 94 个均有插入片段,重组率为 94%,而其中绝大部分插入片段大于 3kb (图 1),大小不同,平均片段长度为 4.6kb。

由于构建的文库的库容足够大,大约包含 2.5GB 的外源 DNA,又因为这些外源 DNA 片段来自于高温热泉微生物,很有可能获得新的耐热蛋白质或酶等。

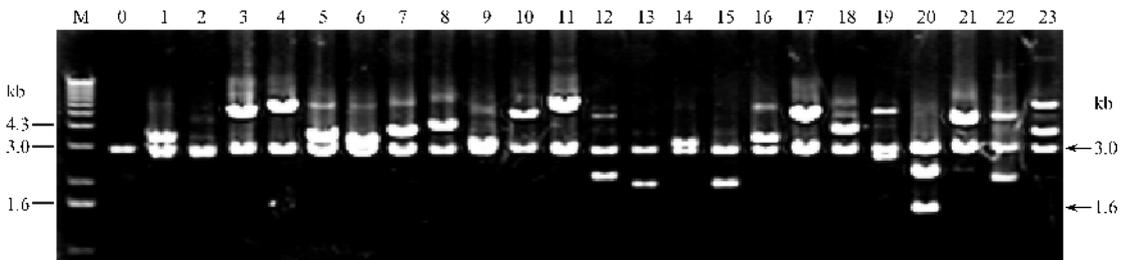


图 1 环境基因组文库的克隆检测

Fig.1 Analysis of metagenomic plasmid library. 1kb DNA ladder; 0. Liner pSK(+); 1~23. Recombinant plasmids digested with *Pst*I.

2.4 热泉未培养微生物基因文库的部分序列分析比较

从该基因文库中随机挑选 30 个质粒进行 DNA 序列分析,发现在 30 个测序样品中,测序没有出现重复的序列,发现片段中含有未见报道的序列。

经过与 BLASTX 比较后 (表 1),发现外源片段的大部分序列与已知嗜热微生物的基因有一定的同源性。这些嗜热微生物包括细菌和古菌。例如:序列

1、10、11、13、19 和 23 与 *Euryarchaeota* 的一些已知微生物的基因序列有相似性,由于蛤蟆嘴热泉环境温度较高,而且泉底土壤环境是一个厌氧的环境,*Euryarchaeota* 可能大量存在于这种环境之中;序列 7、17、24、25、26、29 与 *Proteobacteria* 类群的菌基因序列有相似性,*Proteobacteria* 是细菌中最大的和生理特性多样性最多的一个类群,它广泛地存在于土壤、水体等自然环境中;序列 4、9、14、15、22、30 与

Firmicutes 类群的菌基因序列有相似性,如 9、15 号和 22、30 号序列分别与 *Thermoanaerobacter tengcongensis* MB4、*Moorella thermoacetica* ATCC 39073 的基因部分序列有一定同源性,同源性分别为 42% ~ 55% 和 46% ~ 65%。另一方面发现 3、5、16、18 这 4 条序列与 *Chloroflexi* 类群中的 *Chloroflexus aurantiacus* 的基因部分序列有一定同源性,同源性 40% ~ 60%。2、20 号样品与 *Actinobacteria* 类群中的 *Thermobifida fusca* 的基因部分序列具有一定同源性,同源性为 27% ~ 52%。6、8、28 与 *Deinococcus-Thermus* 类群中基因部分序列同源性为 39% ~ 61% ;还有少数菌群如 *Chlorobi*(12)、*Planctomycetes*(21) 这些类群的菌等存在。从这些序列可以推测出在此高温热泉环境中,可能存在着这些类群的微生物或者与这些微生物亲

缘关系较近的菌种,而且多样性比较高,而这些信息又初步说明了所构建的环境文库中的序列来源多样性和可信性程度高,以及文库的代表性较高。

另外,在这些序列中,大部分都含有可预测功能的基因或基因的一部分,例如样品 22 号可能是 Threonine synthase 苏氨酸合成酶基因的一部分,该酶的这一部分与 *Moorella thermoacetica* ATCC 39073 已知的微生物的相应部分在氨基酸序列水平上的同源性为 65% ;又如样品 25 号含编码 acetyl-CoA carboxylase, carboxyl transferase α 亚基序列的一部分,该酶的这一部分与 *Geobacter sulfurreducens* PCA 种已知的微生物的酶相应部分的序列同源性为 59%。从序列比较结果可以推测出,预期可能获得新的酶基因。

表 1 随机抽取克隆中外源 DNA 部分序列假定的基因功能

Table 1 Putative gene function of partial DNA sequences derived from randomly selected positive clones

NO.	Possible function	Nearest related organism	E Value	Identities
1	Predicted protein	<i>Methanosarcina acetivorans</i> str. C2A	3e-13	37/87 (42%)
2	COG1660 : Predicted P-loop-containing kinase	<i>Thermobifida fusca</i>	1e-55	108/207 (52%)
3	COG0778 : Nitroreductase	<i>Chloroflexus aurantiacus</i>	7e-15	55/119 (46%)
4	Glutamate-tRNA ligase	<i>Bacillus stearothermophilus</i>	3e-62	127/285 (44%)
5	COG0463 : Glycosyltransferases involved in cell wall biogenesis	<i>Chloroflexus aurantiacus</i>	2e-26	76/152 (50%)
6	Chorismate mutase	<i>Thermus thermophilus</i> HB27	2e-32	75/121 (61%)
7	3-methyl-2-oxobutanoate dehydrogenase (lipoamide)	<i>Anaeromyxobacter dehalogenans</i> 2CP-C	2e-69	126/228 (55%)
8	Cytochrome C-type biogenesis protein cmE	<i>Thermus thermophilus</i> HB27	3e-09	32/66 (48%)
9	Hypothetical protein TTE0112	<i>Thermoanaerobacter tengcongensis</i> MB4	1e-21	66/154 (42%)
10	Putative dehydrogenase	<i>Pyrococcus furiosus</i> DSM 3638	9e-25	74/241 (30%)
11	2'-5' RNA ligase	<i>Methanopyrus kandleri</i> AV19	6e-19	53/134 (39%)
12	Similar to ATPase (AAA + superfamily)	<i>Chlorobium phaeobacteroides</i> DSM 266	5e-05	34/78 (43%)
13	Hypothetical protein (multi-domain)	<i>Methanosarcina acetivorans</i> str. C2A	1.5	25/76 (32%)
14	Protein-tyrosine kinase	<i>Clostridium thermocellum</i> ATCC 27405	1e-07	54/179 (30%)
15	Predicted helicases	<i>Thermoanaerobacter tengcongensis</i> MB4	1e-46	95/170 (55%)
16	COG0182 : Predicted translation initiation factor 2B subunit , eIF-2B alpha/beta/delta family	<i>Chloroflexus aurantiacus</i>	8e-56	128/213 (60%)
17	DNA polymerase III alpha subunit	<i>Magnetococcus</i> sp. MC-1	6e-27	63/80 (78%)
18	COG2170 : Uncharacterized conserved protein	<i>Chloroflexus aurantiacus</i>	9e-26	58/101 (57%)
19	Dipeptide ABC transporter , dipeptide-binding protein (dppA)	<i>Archaeoglobus fulgidus</i> DSM 4304	2e-07	44/135 (32%)
20	COG3554 : Uncharacterized protein conserved in bacteria	<i>Thermobifida fusca</i>	1.7	22/79 (27%)
21	Probable squalene-hopene cyclase	<i>Rhodospirella baltica</i> SH 1	4e-21	55/152 (36%)
22	Threonine synthase	<i>Moorella thermoacetica</i> ATCC 39073	2e-86	176/269 (65%)
23	Glucosamine-fructose-6-phosphate aminotransferase	<i>Haloarcula marismortui</i> ATCC 43049	3e-30	68/147 (46%)
24	Inositol-1(or 4)-monophosphatase	<i>Geobacter metallireducens</i> GS-15	6e-35	74/157 (47%)
25	Acetyl-CoA carboxylase , carboxyl transferase , alpha subunit	<i>Geobacter sulfurreducens</i> PCA	3e-56	120/201 (59%)
26	Inner-membrane translocator	<i>Kineococcus radiotolerans</i> SRS30216	4e-28	91/201 (45%)
27	Hypothetical protein TTHA0285	<i>Thermus thermophilus</i> HB8	0.010	51/206 (24%)
28	Major facilitator superfamily	<i>Deinococcus geothermalis</i> DSM 11300	2e-09	63/160 (39%)
29	Ketopantoate hydroxymethyltransferase	<i>Desulfuromonas acetoxidans</i> DSM 684	6e-22	55/94 (58%)
30	DNA polymerase , beta-like region	<i>Moorella thermoacetica</i> ATCC 39073	1e-10	32/69 (46%)

3 讨论

在自然环境中, 99% 以上的微生物在现有实验条件下无法直接培养获得^[13]。利用免培养技术直接从环境中获得微生物 DNA 构建环境基因组文库, 对环境微生物特别是一些特殊环境微生物研究具有重要应用价值。目前该技术已受到国内外学者高度重视, 并有大量相关研究报道^[14-16], 利用该技术已经获得了许多新基因^[17, 18]。但是环境基因组文库的建立有一定难度, 如何获得高纯度和高分子量的环境微生物 DNA 是构建环境基因组文库的关键步骤。土壤环境成分的复杂性给提取 DNA 造成了困难。而且, 不同的环境样品有不同的特性, 没有一种 DNA 提取方法适用于所有样品。在高温热泉样品中, 微生物的生物量非常少, 通过现有的 DNA 提取方法得到大量的、纯度较高的总 DNA 是非常困难的。

本研究采用冻融法、蛋白酶 K 法、SDS-高盐-加热等联合处理方法, 充分地裂解土壤中的微生物 DNA, 使提取的 DNA 最大程度地反映了土壤中微生物的真实情况, 为构建的文库更具代表性奠定了基础。采用试剂盒纯化环境 DNA 减少了纯化过程中所造成的 DNA 损失, 提高了 DNA 的质量, 方便了后期建库工作。

从部分测序结果分析来看, 蛤蟆嘴热泉环境中的微生物多样性较高。而在某些自然环境中, 常常有某种优势菌群存在, 如本实验室在对西藏一个温度为 80℃ 中性古堆热泉土壤的研究中, 用同样的方法构建文库, 从中测序分析了 35 个序列, 发现 50% ~ 60% 的序列都能与 *Thermus thermophilus* HB27 菌的部分基因序列有同源性, 而且大部分序列的同源性较高 (> 60%)。 *Thermus thermophilus* HB27 菌是 2004 年全基因组测序完成的一株菌, 是从日本一中性热泉分离出来, 生长温度为 80℃ 以上, 存在环境和古堆热泉相似, 推测古堆热泉环境中可能含有此种菌群, 而且为优势菌群, 文献上有报道这种情况发生, 如 2004 年 Tyson 等^[19]在 Nature 上发表他们对酸性环境 (Acid biofilm) 的环境基因组文库大量随机测序, 也出现此特殊环境中的微生物的低生物多样性 (Low biodiversity) 的情况, *Leptospirillum* group 这个菌群在其中占了 85%^[20]。

另外值得一提的是橙色绿曲挠丝状菌 (*Chloroflexus aurantiacus*) 是一种在厌氧和低光照条件下形成淡绿色丝状体, 在其它情况下都为橘黄色,

在本研究提取土壤总 DNA 的过程中, 在细胞破壁后, DNA 提取液常常呈现绿色, 且溶于有机相, 可被随后的氯仿抽提, 推测为绿色色素, 因为采样来自泉底土壤, 低光照和厌氧的条件, 而且所得到的序列中有部分序列 (4 条序列) 与 *Chloroflexus aurantiacus* 有一定的同源性, 推测此热泉土壤可能含有这种微生物。

腾冲蛤蟆嘴未培养微生物基因组文库的成功构建, 可以有效地克服制约对该热泉微生物研究的技术瓶颈, 为保存、研究热泉环境微生物基因资源、开发具有重要应用价值的基因提供了有利的技术平台。后期要做的工作是通过高通量筛选的方法对文库进行活性筛选, 以及更多克隆序列测序及利用生物信息学手段分析数据, 两方面的结合有助于发现及开发新的活性物质以及更详细地阐释蛤蟆嘴热泉环境微生物的特性。

参 考 文 献

- [1] 和致中, 彭谦, 陈俊英. 高温菌生物学. 北京: 科学出版社.
- [2] 中国科学院青藏高原综合科学考察队. 腾冲地热. 北京: 科学出版社, 1989.
- [3] 和致中, 彭谦, 马俊, 等. 云南温泉高温菌的研究 VII. 腾冲酸性高温温泉中的极端嗜热性芽孢杆菌. 微生物学报, 1989, 29(30): 161-165.
- [4] Xue Y, Xu Y, Liu Y, et al. *Thermoanaerobacter tengcongensis* sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. *Int J Syst Evol Microbiol*, 2001, 51: 1335-1341.
- [5] Chen C, Lin L, Peng Q, et al. *Meiothermus rosaceus* sp. nov. isolated from Tengchong Rehai in Yunnan, China. *FEMS Microbiol Lett*, 2002, 216: 263-268.
- [6] Lin L, Chen C, Peng Q, et al. *Thermus rehai* sp. nov. isolated from Tengchong Rehai in Yunnan, China. *J Basic Microbiol*, 2002, 42: 337-344.
- [7] 王涛, 柴丽红, 崔晓龙, 等. 免培养法对一热泉细菌多样性的初步研究. 微生物学报, 2003, 5: 541-542.
- [8] 王涛, 柴丽红, 崔晓龙, 等. 从腾冲热泉免培养获得的 5 个 16S rDNA 克隆分析. 云南大学学报, 2003, 25(1): 73-76.
- [9] 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, 4: 133-140.
- [10] 蒋承建, 隆文杰, 梁璇, 等. 碱性土壤基因组 DNA 的分离纯化和基因文库的构建. 广西农业生物科学, 2004, 23(1): 63-66.
- [11] Zhou J, Bruns MA, Tiedje JM. DNA recovery from soils of diverse composition. *Appl Environ Microbiol*, 1996, 62: 316-322.
- [12] 王涛, 柴丽红, 郭春雷, 等. 几种直接从高温热泉沉积物中提取 DNA 方法之比较. 生物技术, 2003, 13(1): 17-18.

- [13] Torsvik V , Goksoyr J , Daae FL. High diversity in DNA of soil bacteria. *Appl Environ Microbiol* , 1990 , **56** :782 – 787.
- [14] Falkowski PG , de Vargas C. Genomics and evolution. Shotgun sequencing in the sea : a blast from the past ? *Science* , 2004 , **304** (5667) :58 – 60.
- [15] Riesenfeld C , Schloss PD , Handelsman J. Metagenomics : genomic analysis of microbial communities , *Annual Review of Genetics* , 2004 , **38** :525 – 552.
- [16] Handelsman J , Liles M , Mann D , *et al.* Cloning the metagenome : culture-independent access to the diversity and functions of the uncultivated microbial world. *Methods in Microbiology-Functional Microbial Genomics* , 2002 , **22** :241 – 255.
- [17] Handelsman J , Rondon MR , Brady SF , *et al.* Molecular biological access to the chemistry of unknown soil microbes : a new frontier for natural products. *Chem Biol* , 1998 , **5** :245 – 249.
- [18] Henne A , Daniel R , Schmitz RA , *et al.* Construction of environmental DNA libraries in *Escherichia coli* ad screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol* , 1999 , **65** :3901 – 3907.
- [19] Henne A , Bruggemann H , Raasch C , *et al.* The genome sequence of the extreme thermophile *Thermus thermophilus* . *Nat Biotechnol* , 2004 , **22** (5) :547 – 553.
- [20] Tyson GW , Chapman J , Hugenholtz P , *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* , 2004 , **428** :37 – 43.

Construction and analysis of a metagenomic library from Tengchong hot spring soil in Yunnan Province

CAI Ying^{1, 2} , CHEN Xiu-zhen¹ , YANG Ke-qian¹ , HUANG Li¹ , DONG Zhi-yang^{1*}

(¹ Institute of Microbiology , Chinese Academy of Science , Beijing 100080 , China)

(² Graduate School of Chinese Academy of Science , Beijing 100049 , China)

Abstract: By a combination of freezing/thawing/ proteinase K-based method and SDS/high-salt/heating treatment , the mixed environmental genomic DNA was isolated directly from a hot spring soil in Tengchong , Yunnan , China. With this method , The DNA yield was up to 1 ~ 2 μ g/g soil. After purification with the Wizard DNA clean up system(Promega ,Madison ,Wis) , the mixed genomic DNA was partially digested with restriction enzyme Pst I . Digested DNA fragments of 3 ~ 8kb were recovered from agrose gel and ligated to the pSK(+) vector. The ligation mixture was transformed into DH10B strain , resulting in the construction of a metagenomic library with about 2.5 \times 10⁴ clones. Restriction enzyme analysis revealed that the average insert is about 4.6kb. Some novel sequences were identified via sequencing and gene annotation analysis of 30 clones randomly chosen from this library.

Keywords : Culture-independent method ; Uncultured microorganism ; Metagenomic library

Foundation item : Chinese National Programs for High Technology Research and Development(2004AA21480)

* Corresponding author. Tel/Fax :86-10-62551206 ; E-mail :dongzy@sun.im.ac.cn

Received : 8 August 2005/Accepted : 17 August 2005/Revised : 22 September 2005