

## 环境样品中 DNA 的分离纯化和文库构建\*

王啸波<sup>1</sup> 唐玉秋<sup>2</sup> 王金华<sup>3</sup> 黄仪秀<sup>1</sup> 陈润生<sup>3</sup> 黄 力<sup>2\* \* \*</sup>

(<sup>1</sup> 北京大学生命科学学院生物化学系 北京 100871)

(<sup>2</sup> 中国科学院微生物研究所微生物资源前期开发国家重点实验室 北京 100080)

(<sup>3</sup> 中国科学院生物物理研究所 北京 100101)

**摘 要:**采用研磨/冻融和 SDS/蛋白酶 K 热处理等理化方法,直接从性质不同的环境样品中提取和纯化混合基因组 DNA。所获得纯品 DNA 的产量为每克样品 2~16 $\mu$ g。对纯品 DNA 进行限制性内切酶处理后,构建了以 pUC18 为载体的 DNA 文库。建库效率为从每克环境样品获得约  $10^3 \sim 10^4$  个含 3~8kb 外源随机插入片段的克隆。通过 DNA 序列测定和基因注释,对从文库中随机选取的克隆进行了分析,发现外源插入片段均含序列未见报道的新基因。本文所做的尝试对于保存、研究和开发未培养微生物基因资源具有意义。

**关键词:** 环境样品, DNA 分离纯化, 混合基因组 DNA 文库, 未培养微生物

**中图分类号:** Q346 **文献标识码:** A **文章编号:** 0001-6209 (2001) 02-0133-08

人类对微生物的利用有悠久的历史,现代生物技术产业的发展在很大程度上得益于微生物学的进步。但据估计,目前为人所知的微生物还不到自然界中微生物总数的 1%<sup>[1]</sup>,绝大多数微生物尚无法在实验室条件下培养,更不用说被开发利用。因此,未培养微生物可能是地球上最大的尚未开发的自然资源。显然,对于微生物学家而言,认识和利用未培养微生物既是机遇也是挑战。

在传统微生物学中,微生物的分离培养是研究利用微生物的第一步。随着分子生物学技术的广泛使用、微生物基因组序列测定工作的大规模开展、功能基因组学和生物信息学的迅速发展,人们现在已经有绕过微生物菌种分离培养这一步骤,直接在基因水平上研究和开发未培养微生物资源。近年来,国外在这方面的工作已屡见报道<sup>[2~4]</sup>。

在基因水平上研究未培养微生物的技术关键之一是从各种环境样品中高效地获得可进行分子操作的混合基因组 DNA。环境样品含大量理化性质复杂的有机和无机颗粒,微生物仅为其中极小的一部分;加之,微生物细胞通常紧密吸附于颗粒表面。因此,从环境样品中高效地获得 DNA 具有相当难度<sup>[3]</sup>。另外,由于环境样品中通常含有大量影响对 DNA 进行分子操作(如:限制性内切酶酶切、转化、PCR 扩增等)的物质<sup>[5~7]</sup>,有效地纯化从环境样品中得到的 DNA 也较为困难。因此,建立环境样品 DNA 的提取纯化方法对于在分子水平上研究未培养微生物具有重要意义。在本研究中,我们以三种性质不同的环境样品为材料,建立了较为有效的 DNA 提取纯化方法,并尝试构建了相应于这些环境样品的、可用来进行基因筛选的 DNA 文库。

\* 中国科学院生物科学与技术研究特别支持费资助(STZ97-3-01)

\*\* 通讯作者

作者简介:王啸波(1974-),男,四川自贡市人,北京大学硕士研究生。

收稿日期:2000-05-23,修稿日期:2000-08-15

# 1 材料和方法

## 1.1 材料

1.1.1 土样:样品 A,采自中国科学院微生物研究所址附近土壤(常温,pH 8.0,干重 97.12%);样品 B,采自北京造纸七厂污染土壤(常温,pH 8.5,干重 92.88%);样品 C,采自云南省腾冲热泉( $\sim 80^{\circ}\text{C}$ ,pH 3.0,干重 71.25%)。

1.1.2 其他材料:限制性内切酶、T4 DNA 连接酶、溶菌酶和碱性磷酸酶购自 Promega,蛋白酶 K、聚乙烯吡咯烷酮(polyvinylpyrrolidone,PVPP)和溴代十六烷基三甲胺(hexadecylmethylammonium bromide,CTAB)购自 Sigma。

## 1.2 环境样品中 DNA 的提取

参照文献[8]并作改动。称取 5 g(干重)环境样品,充分研磨后,置于离心管中,液氮冻融三次:加入 13.5 mL DNA 抽提缓冲液(100 mmol/L Tris-HCl,pH 8.0,100 mmol/L EDTA,100 mmol/L 磷酸钠缓冲液,pH 8.0,1.5 mol/L NaCl,1% CTAB),混匀后,置液氮中,然后取出在  $60^{\circ}\text{C}$  水浴中保温至融化。加入  $50\mu\text{L}$  蛋白酶 K(20 mg/mL)和 1.5 mL SDS(10%),混匀后, $60^{\circ}\text{C}$  保温 2~3 h,每隔 15~20 min 倒管混匀几次;7 000 g 室温离心 10 min,收集上清液,沉淀中加入 5 mL 水, $60^{\circ}\text{C}$  洗两次,每次 10 min,如前离心收集上清液;合并三次上清液,加入等体积的酚/氯仿(1:1)抽提两次;加入 1 g 酸洗 PVPP<sup>[9]</sup>, $37^{\circ}\text{C}$  保温 30 min,经孔径为 0.45 nm 的滤膜过滤除去 PVPP;滤液中加入 0.6 倍体积的异丙醇,室温放置 30 min 后,12 000 g、 $4^{\circ}\text{C}$  离心 15 min,70% 乙醇洗;沉淀用 1.5 mL TE(pH 8.0)溶解;加入 0.1 g 无水乙酸铵,8 000 g、 $4^{\circ}\text{C}$  离心 30 min,取上清液;加入 1 倍体积的异丙醇,冰浴 10 min,12 000 g、 $4^{\circ}\text{C}$  离心 30 min,70% 乙醇洗;沉淀用 0.5 mL TE(含  $20\mu\text{g/mL}$  RNase A)溶解。

## 1.3 粗制 DNA 的纯化

参照文献[10]所叙述的方法进行,采用电泳/电洗脱法或 CsCl 密度梯度离心法纯化粗制 DNA。

## 1.4 DNA 浓度测定

1.4.1 粗制 DNA 浓度测定:系列稀释已知浓度的 pUC18 DNA 的 *Eco*RI 酶切产物,与待测粗制 DNA 一同进行琼脂糖电泳,溴化乙锭染色后,比较 DNA 条带的荧光亮度,推测粗制 DNA 的浓度。

1.4.2 纯化 DNA 浓度测定:利用 DyNA Quant 200 荧光仪(Pharmacia)测定。

## 1.5 DNA 文库的构建

用 *Eco*RI 或 *Pst*I 对纯化 DNA 进行酶切,琼脂糖电泳分离酶切产物,用 DEAE 纤维素滤纸回收 3~8 kb 的酶切片段,将这些片段与经 *Eco*RI 或 *Pst*I 酶切和去磷酸化处理的 pUC18 载体连接,连接产物以氯化钙法转化大肠杆菌 DH5 $\alpha$ ,采用蓝白斑法检测含有重组质粒的大肠杆菌菌落<sup>[10]</sup>。将阳性菌落挑至含氨苄青霉素的 LB 固体培养基上(50 个阳性菌落/平皿),培养至菌落大小为 2~3 mm 后,用 0.5 mL 含氨苄青霉素的 LB 液体培养基洗下平皿上的菌落,加入 0.5 mL 含 30% DMSO 的 LB 液体培养基,混匀,室温放置 2 h,液氮速冻, $-70^{\circ}\text{C}$  保存。

1.6 DNA 文库的初步分析

随机挑选 20 个阳性克隆,对外源 DNA 插入片段的一端进行单向序列测定。对所获得序列进行去除载体部分处理和开放阅读框预测,然后用 BLAST 在包含所有已公布微生物核酸序列的微生物基因数据库中进行检索,根据同源性推测所获得开放阅读框可能编码的功能。

2 结果

2.1 环境样品中 DNA 的提取和纯化

从环境样品中高效提取可进行分子操作的基因组 DNA 是构建环境微生物 DNA 文库的前提。此过程包括:(1)从环境样品中高效获得粗制基因组 DNA 和(2)去除粗制 DNA 中的杂质。

本研究结合采用研磨/冻融和 SDS/蛋白酶 K 热处理等理化方法提取环境样品中的基因组 DNA。在土壤样品中加入已知量的 pUC18 DNA 作为对照,80% 的 DNA 在经过上述方法处理后得以回收。实验中发现,在将 SDS/蛋白酶 K 加入样品前对样品进行研磨及冻融处理可以提高 DNA 的产量。另外,如果用超声波破碎和液氮研磨作为前处理,前者在有效破碎细胞时,对基因组 DNA 造成较大损伤,大大降低了所获得 DNA 片的平均长度,不利于构建含 >4 kb 插入片段的 DNA 文库;后者则不宜用来处理含水量较大的样品(如样品 C)。采用本文的方法,从三种理化性质不同的环境样品中都成功地获得了粗制 DNA,提取效率与文献[8]报道的相近,能够满足酶切建库的要求(表 1)。从样品 C 中得到的 DNA 较少,这可能与该样品中生物量较低有关。

环境样品富含抑制限制性内切酶酶切、PCR、转化等过程的抑制剂(主要是腐殖酸)。一般采用 CTAB 和 PVPP 去除这些抑制性物质<sup>[9]</sup>。用  $A_{260}/A_{280}$  和  $A_{260}/A_{230}$  分别检测样品中蛋白质和腐殖酸杂质含量<sup>[11,12]</sup>,发现样品 DNA 经过 CTAB 和 PVPP 处理后纯度有明显改善,但仍远不如纯品 DNA(表 2)。这样的 DNA 样品仍难以被限制性内切酶酶切。为了进一步纯化 DNA 样品,我们对粗制 DNA 进行了电泳/电洗脱处理。从样品 A 及样品 C 中提取的 DNA 经处理后达到了进行分子操作所需的纯度,但样品 B 含杂质较多,从中提取的 DNA,我们在电泳/电洗脱前增加了 CsCl 密度梯度离心步骤,这样得到的 DNA 可进行酶切等操作。三种粗制 DNA 样品纯化后的回收率见表 1。其中,样品 B DNA 经 CsCl 密度梯度离心纯化后的回收率约为 34%,经电泳/电洗脱纯化后的回收率约为 65%。

表 1 从三种环境样品中得到的粗制 DNA 和纯化 DNA 的产量及纯化回收率

Table 1 Yields of crude and purified DNA and recoveries of DNA purification for three environmental samples

Environmental sample	Crude DNA yield/( $\mu\text{g/g}$ )	Purified DNA yield/( $\mu\text{g/g}$ )	Crude DNA recovered /%
Sample A	25.6	15.5	61
Sample B	21.3	4.7	22
Sample C	3.2	1.9	59

表 2 DNA 样品经 CTAB 和 PVPP 处理后的纯度变化

Table 2 Change in the purity of DNA following the treatment with CTAB and PVPP

DNA sample	Sample B		Sample C		Pure salmon testes DNA	
	$A_{260}/A_{280}$	$A_{260}/A_{230}$	$A_{260}/A_{280}$	$A_{260}/A_{230}$	$A_{260}/A_{280}$	$A_{260}/A_{230}$
No CTAB/No PVPP	1.12	0.70	1.18	0.72	1.91	1.43
CTAB/PVPP	1.35	0.88	1.45	0.91		

用四种限制性内切酶(*Pst*I, *Eco*RI, *Bam*HI 和 *Hind*III)对三种纯化后的 DNA 样品进行酶切试验发现,这些样品均能被酶切,但对酶切的敏感程度有区别。*Pst*I 的酶切效果最好;*Eco*RI 和 *Bam*HI 次之;*Hind*III 最差(*Eco*RI 和 *Pst*I 的酶切结果见图 1)。

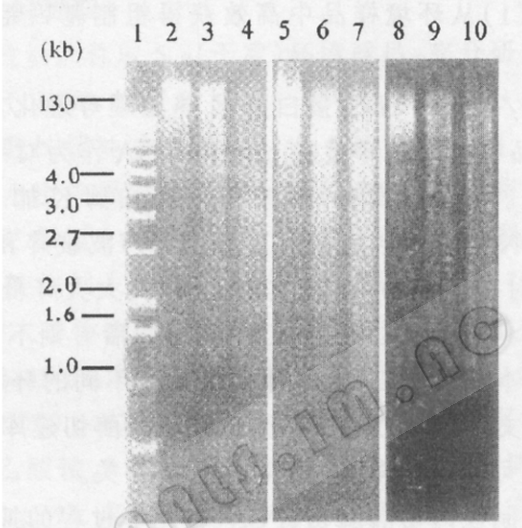


图 1 三种环境样品 DNA 纯化后的酶切结果

Fig.1 Restriction digestion of the three purified environmental DNA samples

1. 1 kb DNA ladder (0.5  $\mu$ g); 2, 5, 8. Purified DNAs from sample A,B and C, respectively;
- 3, 6, 9. Purified DNAs from sample A, B and C, respectively, digested with *Eco*RI;
- 4, 7, 10. DNAs from sample A, B and C, respectively, digested with *Pst* I .

2.2 环境样品 DNA 文库建构

考虑到 *Eco*RI 和 *Pst*I 对从温泉土样和污染土样提取纯化的 DNA 均有良好的酶切效果,且前者的酶切位点富含 A/T,后者的酶切位点富含 G/C,因此,采用这两种酶对样品 DNA 进行了酶切建库的尝试。首先,对样品 DNA 进行酶切,回收 3~8kb 的酶切片段。然后,将所得的酶切片段与相应的、分别经 *Eco*RI 和 *Pst*I 酶切的 pUC18 去磷酸载体连接。再用连接产物转化大肠杆菌 DH5 $\alpha$  菌株。据估计,采用上述方法,从 1g 样品 C 中获得了大约 5 000~7 000 个含 *Eco*RI 酶切片段的阳性克隆或 2 500~5 000 个含 *Pst*I 酶切片段的阳性克隆。而从 1g 样品 B 中,获得了 6 000~12 000 个含 *Eco*RI 酶切片段的阳性克隆或 12 000~15 000 个含 *Pst*I 酶切片段的阳性克隆。通过对阳性克隆所含质粒的酶切分析发现,90% 以上的插入片段大小在 3~8kb 之间(图略)。

2.3 环境样品 DNA 文库的初步评估

从样品 C 基因组 DNA 文库中随机挑出 20 个克隆,对插入片段的一端进行单向测

序。然后,用 BLAST 在微生物基因组数据库中对所得到的插入片段序列进行检索和同源性分析,预测这些序列所含的开放阅读框或部分开放阅读框可能编码的功能(表 3)。

表 3 随机抽取克隆中外源 DNA 部分序列的基因功能注释

Table 3 Gene annotation of partial DNA sequences derived from inserts in randomly selected clones

Sample code	Possible function	COG <sup>[13]</sup>	COG function code <sup>[13]</sup>	S value <sup>[14]</sup>	E value <sup>[14]</sup>
9912	Periplasmic protease	COG0793	M	41	0.001
9914	Ribonuclease HII	COG0164	L	29	5.1
9915	Branched-chain amino acid ABC-type transport system, permease components	COG0559	E	27	9.6
9916	Molecular chaperones, DnaJ family (contain C-terminal Zn finger domain)	COG0484	O	42	6e-04
9918	Protein-disulfide isomerases DsbC/DsbG	COG1651	O	64	8e-11
9919	Periplasmic protease	COG0793	M	39	0.003
9920	*				
9921	*				
e10	Dipeptidyl aminopeptidases/acylaminoacylpeptidases	COG1506	E	31	0.99
e3	Uncharacterized ACR	COG1469	S	31	1.3
e5	Uncharacterized proteins, BmrU family	COG1597	S	31	1.00
e6	Thioredoxin reductase/alkyl hydroperoxide reductase	COG0492	O	48	7e-06
e8	Anaerobic dehydrogenases, typically selenocysteine-containing	COG0243	C	365	e-101
e9	Ribosomal protein S1	COG0539	J	66	2e-11
p1	*				
p10	Periplasmic protease	COG0793	M	29	3.8
p6	*				
p7	Anaerobic dehydrogenases, typically selenocysteine-containing	COG0243	C	352	e-110
p8	*				
p9	*				

\* *E. coli* or vetor sequence

在 20 个测序样品中,6 个为大肠杆菌或载体 DNA 序列,其余 14 个序列均不存在于现有序列数据库中,因而为新序列。在这些新的序列中,12 个含可预测功能的基因或基因的一部分。例如,样品 e6 含编码硫氧还蛋白还原酶基因的一部分。该酶的这一部分与 27 种已知的微生物的还原型辅酶 II 结合硫氧还还原酶(NTR:NADPH-binding thioredoxin reductase)的相应部分在氨基酸序列水平上的同源性为 26%~42%,该同源序列含有与

FAD 结合相关的区域(图 2,参考文献[15]);又如,样品 9918 含编码蛋白二硫键异构酶基因的一部分,该酶的这一部分与 4 种已知的微生物蛋白二硫键异构酶相应部分的序列同源性为 29%~36%(图 3),该保守区域内的 C-Xa-Xb-C 可能是此类蛋白的活性中心<sup>[16]</sup>。

FAD	CONSENSUS	T...G...AAGD
MP592	PQ[REDACTED]IT[REDACTED]EQ-----METS[REDACTED]KGLFAAGDCRSKHF[REDACTED]GT[REDACTED]	(266-300)
MG102	QN[REDACTED]LG[REDACTED]EN-----MQTN[REDACTED]KGFVAGDCRSKSF[REDACTED]AT[REDACTED]	(266-300)
TrxB	N-[REDACTED]KVQSGIHGNATQTS[REDACTED]PG[REDACTED]AAGD[REDACTED]MDHIY[REDACTED]A[REDACTED]TS	(261-299)
YDR353w	EA[REDACTED]KTVPG-----SSLTS[REDACTED]PGF[REDACTED]AAGD[REDACTED]QDSKY[REDACTED]A[REDACTED]TS	(265-300)
BS-ahpF	RM[REDACTED]E[REDACTED]IV[REDACTED]KH-----GATS[REDACTED]PG[REDACTED]AAGDCTDSAYN[REDACTED]IS	(457-491)
HI1158	N-[REDACTED]VVKSGLDGNATATS[REDACTED]EG[REDACTED]AAGD[REDACTED]MDHNY[REDACTED]A[REDACTED]TS	(260-298)
jhp0764	EY[REDACTED]S[REDACTED]VV[REDACTED]FS-----MKTN[REDACTED]QG[REDACTED]AAGD[REDACTED]RIFAP[REDACTED]C[REDACTED]	(259-293)
HP0825	EY[REDACTED]S[REDACTED]VV[REDACTED]FS-----MKTN[REDACTED]QG[REDACTED]AAGD[REDACTED]RIFAP[REDACTED]C[REDACTED]	(259-293)
MJ1536	KK[REDACTED]KT[REDACTED]EN-----CRTN[REDACTED]DG[REDACTED]AVGD[REDACTED]RG-GVM[REDACTED]AK[REDACTED]	(250-283)
MTH708	KG[REDACTED]IT[REDACTED]KF-----QRTN[REDACTED]PL[REDACTED]AAGD[REDACTED]TG-GLN[REDACTED]T[REDACTED]	(252-285)
PH1426	EY[REDACTED]PV[REDACTED]MY-----MRTK[REDACTED]EG[REDACTED]AAGD[REDACTED]TN-VF[REDACTED]AV[REDACTED]	(279-312)
TM0869	PY[REDACTED]IT[REDACTED]EN-----METS[REDACTED]KG[REDACTED]AVGD[REDACTED]RKKNL[REDACTED]T[REDACTED]	(267-301)
e6	GL[REDACTED]N[REDACTED]IV[REDACTED]SQ-----QRTP[REDACTED]SG[REDACTED]AIGD[REDACTED]VN-EIN[REDACTED]AV[REDACTED]	(125-158)

图 2 样品 e6 编码的一个蛋白质的部分序列与 12 种微生物还原型辅酶 II 结合硫氧还原酶的相应序列比较(T...G...AAGD 为 FAD 结合区)

Fig.2 Comparison of partial amino acid sequences from a protein encoded by a cloned insert in sample e6 and 12 NADPH-binding thioredoxin reductases of microbial origin (T...G...AAGD;the FAD-binding region)  
MP592(AAB96239), *Mycoplasma pneumoniae*; MG1022(AAC71320), *Mycoplasma genitalium*; trxB(AAC73974), *Escherichia coli*; YDR353w(AAB64789), *Saccharomyces cerevisiae*; BS.ahpF(CAB16047), *Bacillus subtilis*; HI1158(AAC22813), *Haemophilus influenzae*; jhp0764(AAD06343), *Helicobacter pylori* J99; HP0825(AAD07875), *Helicobacter pylori*; MJ1536(AAB99556), *Methanococcus jannaschii*; MTH708(AAB85213), *Methanobacterium thermoautotrophicum*; PH1426(BAA30532), *Pyrococcus horikoshii*; TM0869(AAD35951), *Thermotoga maritima*.

Cpn0228	HFITPYNTV EE SA E T E L H EISFTL P C RG (41-98)
Ct177	YFITPYNTV EE SA EE S E L H EASLT L P C RG (91-148)
RP025	NDM KKSNI FS PH Y HQT E K -----N A IR (95-145)
BS-yvgV	GQ KDD T GDYK PS KV N D K Q D K D -----FS NV (48-100)
9918	ODM SAD T AST VH A HET EQ TN -----G R FR (34-84)
ACTIVE SITE (C-Xa-Xb-C)	

图 3 样品 9918 编码的一个蛋白质与 4 种蛋白二硫键异构酶 DsbC/DsbG 的部分序列比较(C-Xa-Xb-C 与催化活性有关)

Fig.3 Comparison of partial amino acid sequences from a protein encoded by a cloned insert in sample 9918 and 4 microbial protein disulfide bond isomerases (The C-Xa-Xb-C region is probably involved in catalysis)  
Cpn0228(AE001608), *Chlamydia pneumoniae*; Ct177(AE001291), *Chlamydia trachomatis*; RP025(CAA14496), *Rickettsia prowazekii*; BS.yvgV(CAB15353), *Bacillus subtilis*.

从随机样品的分析结果可以看出,本研究建立的基因组 DNA 文库中所含非污染的外源 DNA 序列基本上均为新序列。由于这些外源 DNA 片段来自于高温热泉样品,可以预期,由此很可能获得新的、具有耐热性的蛋白质或酶。

### 3 讨论

近年来,从环境样品中提取纯化基因组 DNA 已经有许多报道<sup>[5,7,8]</sup>。在这些工作中,纯化 DNA 多被用作 PCR 反应模板,用来进行 16S rDNA 序列分析、样品中的基因检测等。最近,有人通过设计 PCR 引物直接从环境样品 DNA 中扩增得到编码已知蛋白类似物的基因<sup>[4,17]</sup>。此外,一些生物技术公司(如 Diversa)采用分子克隆的方法从环境样品中直接获得了很多新酶<sup>[2]</sup>,但这方面的研究工作尚未见报道。

文献中报道的从环境样品中提取纯化 DNA 的方法一般包括对样品的物理和化学处理。为了提高 DNA 的提取效率,常采用剧烈的物理方法(例如,机械珠打、超声破碎等)。但由此得到的 DNA 片段往往较小且多为平末端,不利于建库。在本研究中,我们以获得对分子操作敏感、可用来高效建库的基因组 DNA 为主要目标,尝试了多种物理化学处理的组合,建立了一套较为有效的方法。采用这一套方法得到的 DNA 片段较长,酶切、连接以及插入载体后转化的效率较好,产量也令人满意。在本工作中,我们从 1g 样品(采自云南腾冲热泉或北京造纸七厂污染土壤(得到了  $10^3 \sim 10^4$  个含 3~8kb 插入片段的克隆。由于微生物基因组中非编码部分较少,因此,从这 1g 样品制备得到的 DNA 文库中的外源基因数目可多达  $10^4$  以上。根据本文对混合基因组 DNA 文库的初步分析,这些外源基因的序列很可能是新的。

混合基因组文库的构建为保存、研究和利用未培养微生物基因资源创造了条件。随着 DNA 测序成本的不断降低,现在已经有可能通过构建特定环境混合基因组文库、测定所有克隆片段(或其末端)来研究基因多样性、开发具有潜在应用价值的基因。若所得到的序列为开放阅读框的一部分,则可通过 PCR、杂交或采用基于杂交技术的功能基因组学方法从用来构建上述文库的混合基因组 DNA 中获得完整的基因。另外,有了基因组文库,还可以采用基于序列同源性或酶活性的筛选方法,寻找目的基因。

### 参 考 文 献

- [1] Amann R I, Ludwig W, Schleifer K H. *Microbiol Rev*, 1995, **59**: 143~69.
- [2] <http://www.diversa.com>
- [3] Head I M, Sauder J R, Pickup R W. *Microb Ecol*, 1998, **35**: 1~21.
- [4] Cottrell M T, Moore J A, Kirchman D L. *Appl Environ Microbiol*, 1999, **65**: 2553~2557.
- [5] Porteous L A, Armstrong J L. *Curr Microbiol*, 1991, **22**: 345~348.
- [6] Tebbe C C, Vahjen W. *Appl Environ Microbiol*, 1993, **59**: 2657~2665.
- [7] Tsai Y L, Olson B H. *Appl Environ Microbiol*, 1992, **58**: 2292~2295.
- [8] Zhou J, Bruns M A, Tiedje J M. *Appl Environ Microbiol*, 1996, **62**: 316~322.
- [9] Holben W E, Jansson J K, Chelms BK, et al. *Appl Environ Microbiol*, 1988, **54**: 703~711.
- [10] Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning: A laboratory manual*, Second edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989.
- [11] Ogram A, Sayler G S, Barkay T. *J Microbiol Methods*, 1987, **7**: 57~66.

- [12] Steffen R J, Goksoyr J, Bej A K, *et al.* *Appl Environ Microbiol*, 1988, **54**: 2908~2915.
- [13] Roman L T, Michael Y G, Darren A N, *et al.* *Nucl Acids Res*, 2000, **28**: 33~36.
- [14] Stephen F A., Warren G, Webb M, *et al.* *J Mol Biol*, 1990, **215**: 403~410.
- [15] Ben-Menachem G, Himmelreich R, Herrmann R, *et al.* *Microbiology*, 1997, **143**: 1933~1940.
- [16] Raina S, Missiakas D. *Annu Rev Microbiol*, 1997, **51**: 179~202.
- [17] Seow K T, Meurer G, Gerlitz M, *et al.* *J Bacteriol*, 1997, **179**: 7360~7368.

## EXTRACTION OF DNA FROM ENVIRONMENTAL SAMPLES AND CONSTRUCTION OF MIXED GENOMIC DNA LIBRARIES\*

Wang Xiaobo<sup>1</sup> Tang Yuqiu<sup>2</sup> Wang Jinhua<sup>3</sup> Huang Yixiu<sup>1</sup>

Chen Runsheng<sup>3</sup> Huang Li\*\*

(<sup>1</sup>Department of Biochemistry & Molecular Biology, College of Life Sciences, Beijing University, Beijing 1000871, China)

(<sup>2</sup>State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Science, Beijing 100080, China)

(<sup>3</sup>Institute of Biophysics, Chinese Academy of Science, Beijing 100101, China)

**Abstract:** A method has been developed for extracting and purifying genomic DNA from environmental samples. In this method, an environmental sample is treated first by grinding and freezing/thawing and subsequently by SDS/proteinase K-based DNA extraction. The yields of purified DNA from three samples used in this study ranged from 2 to 16 $\mu$ g per gram of dry sample. Mixed genomic DNA libraries for two of the environmental samples were constructed by inserting restriction fragments (3~8 kb) of the purified DNAs into plasmid pUC18 and transforming *E. coli* DH5 $\alpha$  with the resultant plasmids. Approximately 10<sup>3</sup> to 10<sup>4</sup> insert-containing clones were obtained from 1g of each sample. Clone libraries were analyzed by DNA sequencing and gene annotation. Among 20 randomly-selected clones, 14 contained an insert whose sequence had not been reported while the rest had an insert of either *E. coli* or vector origin. A search of sequence databases using the end sequences of each of the foreign inserts showed that each sequence was part of a gene encoding, in most cases, a predictable function. Our results are of significance to the collection, investigation and exploitation of the genes of uncultured microorganisms.

**Key words:** Environmental samples, DNA extraction, Mixed genomic DNA libraries, Uncultured microorganisms

\* Project Supported by a Special grant from the Biological Science and Technology Program of the Chinese Academy of Sciences (ST297-3-01)

\*\* Corresponding author