

# PCR 产物直接测序技术中影响因素的研究

徐祖元<sup>1,2</sup>,包其郁<sup>1</sup>,牛宇欣<sup>1</sup>

(1. 中国科学院遗传与发育生物学研究所人类基因组中心/北京华大基因研究中心,北京,100101;2. 湖北荆州师范学院,荆州,434104)

**摘要:**探讨了 PCR 产物直接测序技术中的影响因素,结果表明:PCR 产物特异性是影响其测序成败的关键因素,PCR 反应只有产生惟一扩增产物时,其产物才能被用来直接测序;PCR 反应体系残留混合物(dNTP、引物和盐离子等)对其测序质量有明显不利影响,PCR 产物纯化后其测序质量能明显提高;同时,PCR 产物大小不同,其测序反应的模板用量也不同,在一定长度范围内,最适模板用量随 PCR 产物长度增加而增加。

**关键词:**聚合酶链反应(PCR);PCR 产物直接测序;ABI-377 型 DNA 自动测序仪;Q20

中图分类号:Q78

文献标识码:A

文章编号:0253-9772(2002)05-0548-03

## Factors that Influence Direct Sequencing of PCR Products

XU Zu-yuan<sup>1,2</sup>,BAO Qi-yu<sup>1</sup>,NIU Yu-xin<sup>1</sup>

(1. Beijing Genomics Institute / Human Genome Center, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; 2. Jingzhou Teachers College, Jingzhou, Hubei 434104, China)

**Abstract:** Factors influenced direct sequencing of PCR (polymerase chain reaction) products were investigated in this paper. It showed that the specialization of PCR products played a key role in their sequencing reactions and only which could be sequenced directly. It also showed that the PCR reaction residues (including dNTP, primers, and metal ion) affected badly on the sequencing quality, so the purification of PCR products was necessary before sequencing. In addition, the optimum templates amount in sequencing reaction rose with the increasing of their DNA size in a certain range.

**Key words:** polymerase chain reaction(PCR); direct sequencing of PCR product; ABI 377-DNA sequencer; Q20

PCR 产物直接测序技术现已成为分子生物学和基因组学研究中一个重要技术,广泛用于基因突变检测、遗传性疾病诊断、单核苷酸多态性(SNP)研究、基因组重叠序列群(contig)之间的间隙填补(补洞)等方面<sup>[1~4]</sup>。与传统克隆测序技术相比较<sup>[5]</sup>,PCR 产物直接测序技术具有快速、简明、稳定的优点。本文结合作者近期科研“腾冲嗜热菌(*Thermoanaerobacter tengcongensis*)MB4<sup>T</sup>全基因组测序”工作,对 PCR 产物直接测序技术多种影响因素作一小结,旨在为实验研究提供一种参考。

## 1 材料与方 法

### 1.1 材 料

腾冲嗜热菌(*T. tengcongensis*)MB4<sup>T</sup>菌株由中国科学院微生物研究所提供。

### 1.2 方 法

1.2.1 腾冲嗜热菌基因组 DNA 提取按文献 6。

1.2.2 PCR 反应按常规方法进行,引物由北京奥科生物技术有限公司或北京赛百盛公司合成。体系 50ul,其中模板 20 ng、两条引物各 0.4 $\mu$ mol $\cdot$ L<sup>-1</sup>;热循环条件因引物和待扩增 DNA 片段长度等因素而异;长距离 PCR 反应按 Advantage Genomic PCR 试剂盒(Clontech 公司)要求进行;PCR 产物纯化采用 Wizard<sup>R</sup> PCR Preps DNA 纯化试剂盒(Promega 公司)或 MultiScreen 384 Filter Plates(Millipore 公司),电泳检测,产物 4 $^{\circ}$ C 保存备用。

收稿日期:2002-04-03;修回日期:2002-06-04

基金项目:中国科学院基金资助

作者简介:徐祖元(1967-),男(汉),湖北监利人,博士生,讲师,专业方向:基因组学。电话:010-80494199,E-mail:xuzy@genomics.org.cn

致 谢:中国科学院微生物研究所田宇清同志对本实验给予了大量帮助,特此致谢!

1.2.3 测序反应参照 ABI—BigDye™ 测序试剂盒(ABI 公司)说明进行,体系 10ul,其中模板用量按参数  $3 \sim 5 \times$  模板长度  $\div 100\text{ng}$  加入, BigDye Mix 1.5  $\mu\text{l}$ 、测序引物 0.5  $\mu\text{mol} \cdot \text{L}^{-1}$ ;热循环条件为 96℃ 变性 2min、96℃ 变性 10s、50℃ 退火 10s、60℃ 延伸 4min、第 2~4 步进行 35 个循环,4℃ 保存;测序反应产物纯化用 70% 乙醇沉淀法,电泳用 ABI—377 型 DNA 测序仪。

1.2.4 碱基测序质量分析用美国华盛顿大学 Phil Green 实验室开发的 Phred 软件<sup>[7-8]</sup>,高质量碱基的标准设定为大于 Q20(即每一个碱基测序正确率大于 99.0%)<sup>[3,7,8]</sup>,统计每一测序反应序列(read)读长以及测序反应成功率(该 read 中连续高质量碱基的数目大于 50 时即为成功测序反应)。

## 2 结果与讨论

### 2.1 PCR 产物特异性对测序质量的影响

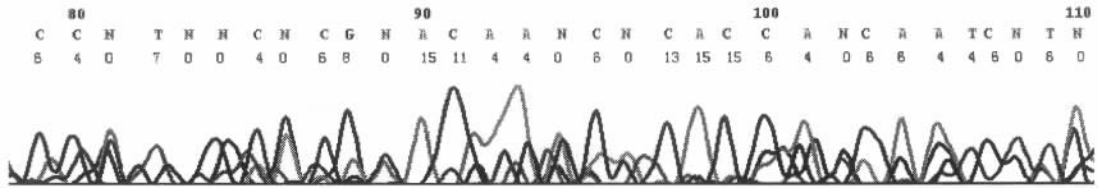


图 1 非特异性 PCR 产物直接测序反应的曲线峰图

Fig. 1 The trace of direct sequencing of non-specific PCR product

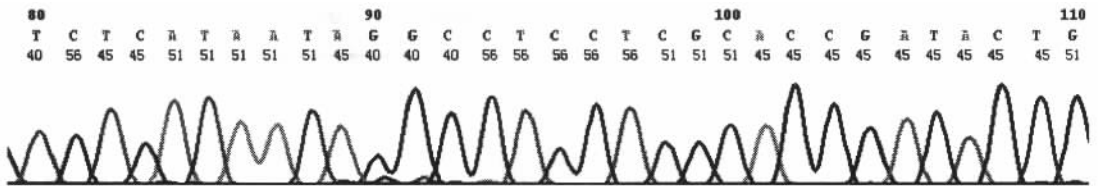


图 2 特异性 PCR 产物直接测序反应的曲线峰图

Fig. 2 The trace of direct sequencing of specific PCR product

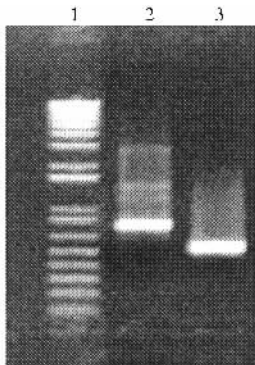


图 3 测序模板(PCR 产物)质量示意图

1:1kb ladder;2:低质量模板;3:高质量模板;电泳方向:从上向下。

Fig. 3 Comparison of quality of sequencing templates

1:1kb ladder;2:low quality template;3:high quality template.

图 1 为一个低质量 PCR 产物直接测序反应的曲线峰图,其背景噪音(噪峰)杂,信噪比(信号/噪音)低;大多数小峰对应碱基不明(“N”较多),所有碱基的测序质量低于 Q20,即没有一个碱基的测序正确率在 99.0% 以上。与图 1 相反,图 2 为一个高质量 PCR 产物直接测序反应的曲线峰图,其背景无噪音(噪峰),每一小峰对应特定碱基,所有碱基的测序质量在 Q40 以上(测序正确率大于 99.999%),大大高于 Q20 标准(碱基下面的数字表示该碱基的测序质量参数)。

通过分析,造成上述测序结果的差异是由于测序反应的模板质量高低即 PCR 产物特异性与否所致。PCR 产物电泳检测发现,图 1 测序反应的模板不纯,质量差(见图 3),表现为除主带之外还有 3 条弱带,这 3 条弱带显然也可作为模板参与测序反应,导致测序失败;而图 2 测序反应的模板质量高,仅一条清晰带。由此表明,PCR 反应只有产生惟一扩增产物时,其产物才能被用来直接进行测序,否则测序反应不能成功。

### 2.2 PCR 反应体系残留混合物对测序质量的影响

PCR 产物未经纯化而直接测序的结果表明,当模板长度大于 1500 bp 时,不容易得到成功的测序反应,当模板长度小于 1500 bp 时,尽管得到了成功的反应,但测序质量并不高。PCR 产物纯化前后测序质量对比见表 1,PCR 产物纯化后的测序平均读长、最大读长、反应成功率分别为 341.4 nt、613 nt 和 85.6%,比纯化前明显提高。

PCR 反应完成后,其反应体系混合物中往往剩有过量 dNTP、引物及盐离子等,它们对测序反应会产生不利影响。其中,dNTP 进入测序循环反应后,会使测序试剂盒中的 dNTP 和 ddNTP 之间的比值发生改变,从而影响链终止反应;过剩引物或引物二聚体可能干扰测序引物与模板之间的适当退火进程,并消耗 dNTP 和 ddNTP;残留盐离子可能对测序酶及其缓冲液有不良影响,得不到理想测序结果。

表 1 PCR 产物纯化前后测序质量对比

Table 1 Comparison of direct sequencing quality between purified and non-purified PCR products

	模板用量/ng amount of template	模板长度/bp length of template	平均读长/nt average length of reads	最大读长/nt maximum length among reads	反应成功率(%) rate of successful sequencing
纯化前*	2~4×模板长度÷100	<1500	89.7	202	41.3
纯化后**	3~5×模板长度÷100	250~6000	341.4	613	85.6

\* 统计 52 个测序反应, \*\* 统计 180 个测序反应。

### 2.3 测序反应模板用量对测序质量的影响

根据图 4 结果,在 250~6000 bp 长度范围内的 10 个被调查模板长度中,对于 250 bp 长度模板,当用量为 10ng 左右时,其测序序列(read)读长最长;对于 6000 bp 长度模板,当用量为 200ng 左右时,其 read 读长最长;最适模板用量的总趋势表现为:当模板长度小于 2000 bp 时,其用量参数为 4~5×模板长度÷100ng 左右;当模板长度在 2000 bp 至 6000 bp 之间时,其用量参数为 3~4×模板长度÷100ng 左右。从测序反应的曲线峰图上看(图略),当模板用量过高

时,测序反应峰图曲线起始处信号太强,然后迅速减弱,可判读序列短(又叫头重脚轻现象);反之,当模板用量过低时,峰图曲线上信号自始至终太弱,无法判读碱基;而当模板量适当时,起始信号强度适当,信号衰减幅度小,整个序列曲线峰图平缓,可判读序列长(参见图 2)。

上述结果表明,每一特定长度 PCR 产物在进行测序反应时都有最适的模板用量,当高于或低于这个用量时,其测序质量都会受到影响。在 250~6000bp 长度范围内,最适测序反应模板用量随 PCR 产物长度增加而增加。

为获得高质量测序模板,以下

几点是值得考虑的。(1)PCR 产物在电泳鉴定时,要求条带清晰、无杂带、无拖尾;不过,受电泳分辨率限制,有些非特异扩增条带由于具有相同或相近的 DNA 分子量而不能得到分辨,容易被误认为是同一条带,导致测序反应失败。(2)要得到理想的测序结果,必须在测序反应前对 PCR 产物进行纯化,以清除残留的 dNTP、引物和盐离子等;纯化方法包括经典的酚-氯仿-乙醇法和琼脂糖凝胶电泳切胶回收法,以及现在商品化的微柱纯化法、

384/96 孔膜过滤纯化法;目前的纯化方法仍须改进,表现为纯化过程费时,回收率不稳定,回收配套设备多,操作有繁琐之嫌。(3)测序模板纯化后,要进行定量,这可通过电泳根据定量 DNA Marker 进行而不必用其他方法。(4)一般情况下,不必考虑测序模板结构特征,但有些模板结构特殊,如富含 GCs 模板、同聚物模板(如 poly(A)模板),对于这种情况,改变测序反应条件(如测序酶)及循环条件可能有一定帮助。

### 参考文献(References):

[1] 李华川,陆士新. PCR 直接测序方法及其在肿瘤研究中的应用[J]. 生物化学与生物物理进展, 1994, 21(2): 167~190.  
[2] 刘奇迹,龚瑶琴,陈丙玺,等. Smith-Fineman-Myers 综合征与 GRIA3 基因的连锁和突变分析[J]. 遗传学报, 2001, 28

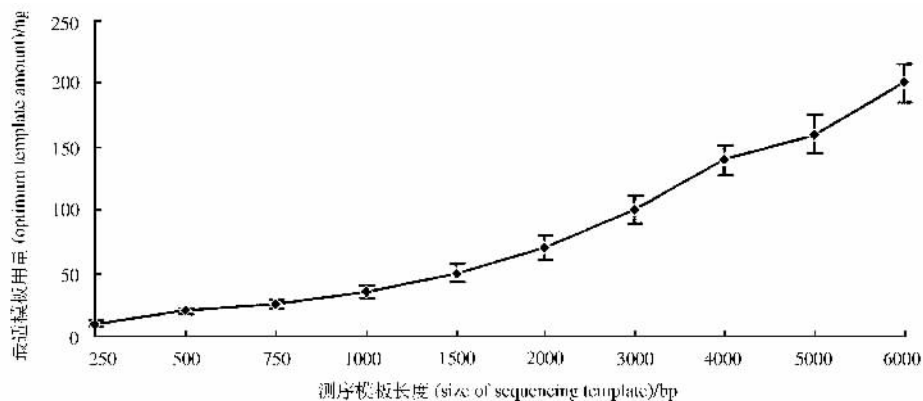


图 4 测序模板长度与最适模板用量之间的关系

Fig. 4 The relationship between the size of sequencing template and its optimum template amount

(11):985~990.

[3] Deckert G, Warren P V, Gassterland T, et al. The complete genome of the hyperthermophilic bacterium aquifex aeolicus [J]. Nature, 1998, 392: 353~358.  
[4] Bao Q, Tian Y, Li W, Xu Z, et al. A Complete Sequence of T. tengcongensis Genome [J]. Genome Res, 2002, 12(5): 689~700.  
[5] 袁汉英,李纹,李育阳. 一种简便的双链 DNA 测序法[J]. 遗传, 1996, 18(6): 25~26.  
[6] Marmur J. A procedure for the isolation of deoxyribonucleic acid from micro-organisms [J]. J Mol Biol, 1961, 3: 208~218.  
[7] Ewing B, Hillier L, Wendl M C, Green P. Base-calling of automated sequencer traces using Phred I. Accuracy assessment [J]. Genome Res, 1998, 8(3): 175~185.  
[8] Wing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities [J]. Genome Res, 1998, 8(3): 186~194.