

毛细管阵列电泳与规模化 DNA 测序

甄志成, 姚志建

(国家人类基因组北方研究中心, 北京 100176)

摘要 根据 10 080 份基因组 DNA 测序的结果, 讨论了毛细管阵列电泳测序方法的技术特点, 并对影响测序结果的一些因素进行了分析。在此基础上与平板凝胶电泳方法进行比较, 显示了毛细管阵列电泳的优点。同时也对大规模测序技术环节之间的协调进行了探讨。

关键词 毛细管阵列电泳; DNA 测序; 基因组

中图分类号: O658 文献标识码: A 文章编号: 1000-8713(2001)04-0361-04

1 前言

1977 年 Sanger 等^[1]提出了用双脱氧链终止法进行 DNA 测序, 即利用双脱氧核苷三磷酸的渗入阻断 DNA 聚合酶的作用, 使合成链在双脱氧核苷酸处中止, 并采用聚丙烯酰胺凝胶电泳分离长度各差一个核苷酸的 DNA 单链, 以读出序列。双脱氧链终止法早期采用同位素标记(³²P, ³⁵S)的方法, 但同位素的散射与衰变影响测序结果读取的准确性, 且操作步骤繁琐, 不能满足大规模测序的需求。20 世纪 80 年代末, 荧光标记的测序技术逐渐取代同位素标记, 进而四色荧光标记的应用又使测序反应物的分离能在一个泳道完成, 从而降低了泳道间迁移率的差异对精度的影响, 为一块平板胶做 96 个样品的高密度电泳的发展奠定了基础。凝胶平板电泳 DNA 自动测序仪就是在此技术背景下发展起来的。

20 世纪 90 年代, 基因组学的出现和发展对测序数量的需要产生了数量级的跳跃, 毛细管阵列电泳 DNA 测序仪就应运而生了^[2]。本文从规模化测序的实际操作出发, 探讨毛细管阵列电泳测序的运行方式、操作中所遇到的问题、以及影响测序结果的因素。并对凝胶平板电泳测序仪和毛细管阵列电泳测序仪的操作过程作了比较。

2 材料与方法

2.1 测序标本

测序标本来源于人 3 号染色体短臂末端 (3pter-D3S3610) 的 BAC 克隆 (含 100 kbp ~ 200 kbp 的片段)。首先对 BAC 进行亚克隆操作, 待测序的基因组片段插入在 PUC 18 质粒中, 形成可供测序的亚克隆菌落, 插入片段长约 2 kbp。将亚克隆后的菌液在 LB 平皿上划线培养过夜。于次日随机挑取平板上的单个菌落接种于 96 孔方腔板中, 在 2 ×

YT 培养基中于 37 °C 下培养, 以 330 r/min 的速率摇菌过夜。采用碱裂解法提取质粒^[3], 并经过 Multiscreen 滤板 (Millipore 公司) 过滤。异丙醇沉淀质粒 DNA, 乙醇洗沉淀, 最后溶于去离子水中。此时质粒 DNA 含有约 2 kbp 长的待测片段。然后进行琼脂糖电泳操作, 将模板定量为 100 mg/L。

2.2 测序反应

采用末端终止法测序: 96 孔板中, 每孔反应体系含模板 200 ng, 测序混合物 (DNA 聚合酶, 脱氧核苷酸 dNTP, 荧光标记双脱氧核苷酸 ddNTP), 引物 (M13 正向或反向), 去离子水 (电阻大于 18.2 MΩ, 由 Milli-Q 纯水仪制得), 缓冲液, 总体积为 10 μL。

将 96 孔板放入 PCR 仪, 采用循环测序法。热循环式反应程序为 96 °C 保持 2 min, 96 °C 保持 10 s, 50 °C 保持 5 s, 60 °C 保持 4 min。步骤 2 至 4 之间循环 35 次, 最后在 4 °C 下保存。

反应产物用异丙醇沉淀, 乙醇洗后, 每孔质粒 DNA 溶于 20 μL 去离子水中, 等待上样。

2.3 电泳条件及数据处理

采用 PE 公司生产的 3700 毛细管阵列电泳仪, 测序胶为 POP6 型。电泳电压 5 250 V, 电泳温度 50 °C。电泳时间根据所拟的测序列长度而定, 大约每 15 min 电泳相当于读取 100 bp 的长度。上样时间为 20 s, 电压 1 500 V。

使用 Sequencing Analysis 3.6 软件 (PE 公司) 对原始数据进行初步分析, 使用 Phred 软件 (华盛顿大学测序中心提供, 经中科院生物物理所改进) 进行质量控制。

3 结果与讨论

3.1 测序结果的统计分析

统计分析了得自 105 次电泳运行的 10 080 个

样品的测序结果。方法是将每次电泳运行得到的符合质量要求的结果累加,得出每一次电泳的样品总读长,以此数值作为衡量电泳运行质量高低的参数。将上述 105 次电泳样品的每板总读长划分为几个档位进行统计分析,则测序样品每板总读长在 50 kbp 以上的高质量样品约占 88%,其中又有超过半数达到 60 kbp 以上,而每板总读长低于 40 kbp 的只占 4%。由此可知,毛细管阵列电泳的测序结果是高质量和稳定的。而产生部分低质量测序结果的原因在于质粒提取及规模化操作中的一些失误,例如模板的纯度不高,模板量不足或过量等,并非毛细管电泳过程中的问题。

由于测序操作每批测 96 个样品,即 96 个样品/板,所以当每板累积总读长达到 60 kbp 以上时,平均每个成功样品的读长在 600 bp 以上。Phred 软件对色谱峰形要求比较高,对峰高亦有一定的要求。如果以目测观察是否出现 N(无法读出的碱基)来作为标准,则成功样品的读长可达到 700 bp。图 1 是 BAC 编号为 813n23 的一个次级克隆片段测序结果。由图 1 可见在 700 bp 左右,色谱峰的峰形依然较为尖锐,峰与峰的间距并没有随着电泳的进行而拉开;且由峰高可看出荧光信号在电泳接近尾声时仍保持很好的检测强度。良好的色谱峰峰形避免了测序软件的误读,从而得到较长的读长。



图 1 测序电泳图尾部图例

Fig.1 The tail portion of sequencing electrophoretogram

Electrophoresis voltage :5 250 V ; electrophoresis temperature :50 °C ; electrophoresis time :105 min ; loading time 20 s ; loading voltage :1 500 V .

3.2 DNA 测序模板的质量与浓度

模板的质量对测序有极重要的影响,因此测序模板的纯度要求很高。首先,不能有过多杂蛋白,因为毛细管壁上的硅醇基团易通过疏水作用与杂蛋白结合,从而降低毛细管的使用寿命,这一点在模板 DNA 提取过程中应格外注意。其次,模板中不能含有过多的盐,因为盐的存在会产生“歧视效应”(在电动上样过程中,盐会消耗大部分电功),从而造成上样量减少,进而影响信号强度。另外,在模板的规模化提取中有时会出现模板间的交叉污染,从而导致

色谱峰信噪比下降,读长缩短。如果交叉污染是在挑取次级克隆时发生的(由于克隆密集,所挑并非单克隆),就会在测序反应体系中存在两种同样浓度的模板,最终得到如图 2 所示的色谱峰。如果在挑取克隆之后发生污染,在反应体系中将出现两种不同浓度的模板,最终得到如图 3 所示的色谱峰,其污染峰高根据污染模板在反应体系中所占的比例决定^[4]。

模板提取后的琼脂糖电泳定量操作有利于提早确定模板的污染。图 4 中箭头所示两条亮度相近条带,即表示提纯的模板有交叉污染。

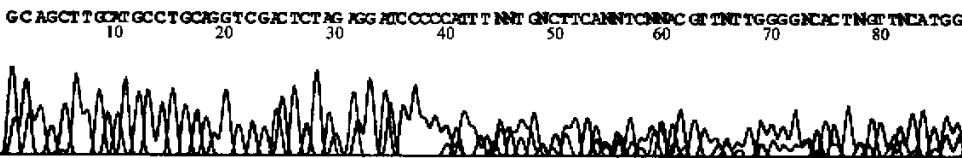


图 2 克隆挑取时引入的模板交叉污染

Fig.2 Cross-contamination induced by inoculation

Conditions as in Fig. 1.

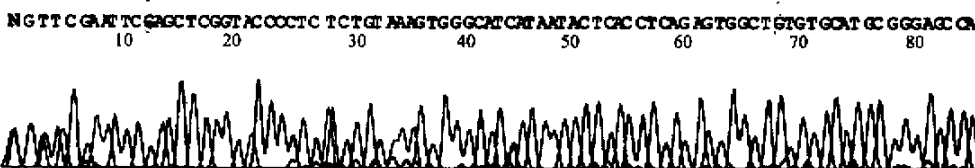


图 3 克隆挑取后引入的模板交叉污染

Fig.3 Cross-contamination induced after inoculation

Conditions as in Fig. 1.

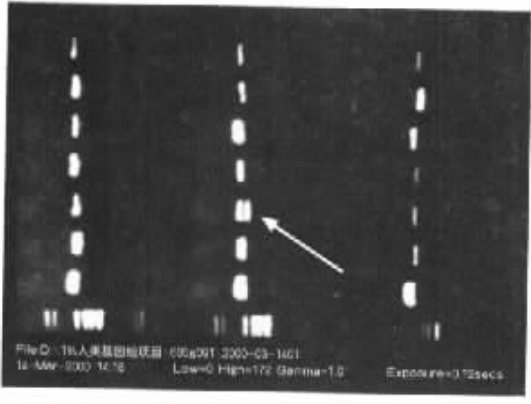


图 4 琼脂糖电泳检定克隆挑取时引入的模板交叉污染

Fig.4 Evaluating the purity of template by batch slab agarose electrophoresis

The cross-contaminated template is pointed by an arrow, see Fig.2 for its chromatogram.

Electrophoresis conditions :0.8% agarose , sample volume :2 μ L , loading volume 8 μ L , electrophoresis voltage :150 V .

除了模板之间的交叉性污染以外,还可能在模板提取中引入外源性污染。外源性污染和交叉污染可表现出不同的色谱图,因为交叉污染时虽然有两种模板,但它们的载体区相同,所以色谱图的前一部分显示相同的谱形,而外源性污染则会表现出自始至终的高噪声,如图 5 所示。

测序反应体系中,模板 DNA 的含量必须控制在一定范围内,10 μ L 反应体系所用的模板量约为 200 ng。若模板过量,而体系中测序混合物和其他成分所占的比例不变,则会在循环早期过多消耗 dNTP 和 ddNTP,使后反应体系中两者浓度降低,不能合成长链,从而出现色谱峰到某一区域迅速下降的现象。若模板 DNA 含量偏低,则表现为色谱峰高降低、信噪比下降。

如果模板中 GC 重复序列偏长,则会在测序反应时影响聚合反应,使之难以进行下去,从而导致电泳图中色谱峰突然丢失的现象。此问题是目前测序技术的主要难题之一。

3.3 进样量

毛细管阵列电泳采用电迁移进样,进样量对测序结果的影响明显,可以通过改变上样电压和时间

来调整进样量。当进样量不足时,增大进样电压和延长进样时间一方面会使进样量增加,另一方面随着进样电压的升高,柱效将下降,且下降趋势随进样时间的增加而加剧。因此,只要保证检测器能接受到足够强的信号,进样区带越小越好,即进样电压和时间的调整必须限制在一定范围内。同时,上样时间过长还会引起样品的区带扩散,导致分辨率下降。故而倾向于在上样时间较短的前提下,适当提高上样电压,借以找到最适值。目前采用的上样电压为 1 500 V,时间为 20 s。

3.4 毛细管阵列电泳仪与平板凝胶电泳仪的比较

如果以达到读长 700 bp 作为目标,将平板凝胶电泳测序仪与毛细管阵列电泳测序仪的运行情况进行比较,结果见表 1。

表 1 毛细管阵列电泳仪与凝胶平板电泳仪的比较

Table 1 Comparison between capillary array electrophoresis and gel slab electrophoresis

Item	Capillary array	Gel slab
Running time	4 h	7 h
Pre-steps before running	no	washing plate , making gel
Sample treatment before loading	no	denaturing
Loading procedure	automatical	manual
Analysis of chromatogram	no tracking step	tracking before analysis

由表 1 可知,采用平板凝胶电泳测序仪需要 7 h 的电泳时间,若将电泳前的洗板、铺胶、进样等辅助工作所需时间计算在内,则总共约需 10 h。而采用毛细管电泳测序仪耗时较少,仅需 4 h,其原因在于此仪器的高度自动化,使用时只要将样品板放在操作平台上,测序仪就会自动清洗毛细管、灌胶、加样,直至结束电泳。电泳结束后,平板凝胶电泳测序仪要通过人工进行泳道校正以减少误差,确定泳道,一般每板样品约耗时 30 min ~ 60 min,而且这一步人工操作极易出现差错,造成样品序列之间混淆。由于毛细管阵列电泳使每个样品都独享一个泳道,因此也就避免了繁琐的泳道校正,降低了出错几率。

在规模化测序过程中,为了尽可能发挥仪器的作用,一般都使仪器 24 h 连续工作。一台凝胶平板电泳至少需要 2 人轮换,才能每天运行 3 次,而毛细管阵列电泳只需 1 人工作 6 h,即可达到上述工作量

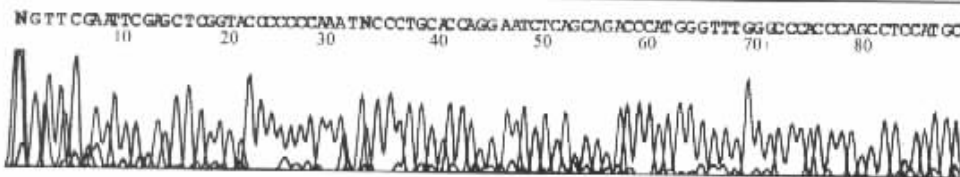


图 5 测序中的外源性污染

Fig.5 The contamination induced from factors outside the DNA sequencing reaction

Conditions as in Fig. 1.

的 2 倍,完成测序反应、反应产物纯化、仪器操作到结果分析等工作。通过分析 10 080 份样品的测序结果,证明毛细管阵列电泳测序仪与较早时期广泛应用的平板凝胶电泳仪相比有明显的优点,适于承担大量测序任务及研究基因组学的实验室使用^[5]。

大规模测序目前已发展到工业化阶段。由于测序量增大及必然伴随而来的高额成本,更加突出了对测序质量及成功率的要求。测序质量的提高除仪器因素以外,主要取决于样品准备,电泳参数设置等质控和管理的因素。希望上述内容对规模化 DNA 测序操作者有所助益。

致谢 感谢中国医学科学院的沈岩教授在实验过程中的指导和贡金英同志的协助。

参考文献:

- [1] Sanger F , Nicklen S , Coulson A R. Proc Natl Acad Sci USA , 1977 , 74 : 5 463-5 467
- [2] Adams M D , Celniker S E , Holt R A , et al. Science , 2000 , 287 : 2 185-2 195
- [3] Sambrook J , Fritsch E E , Maniatis T. Molecular cloning. 2nd Ed. Beijing : Science Press , 1992. 19-22
- [4] XIAO Cui-ying ZHANG Si-zhong , WU Hui , et al. Chinese Journal of Medical Genetics , 1998 , 15 (4) : 238-241
肖翠英 张思仲 , 武 辉 , 等. 中国医学遗传学杂志 , 1998 , 15 (4) : 238-241
- [5] CHEN Zhu ZHANG Si-zhong. Chinese Journal of Medical Genetics , 1998 , 15 (4) : 195-197
陈 竺 张思仲. 中国医学遗传学杂志 , 1998 , 15 (4) : 195-197

Capillary Array Electrophoresis and High-Throughput DNA Sequencing

ZHEN Zhi-cheng , YAO Zhi-jian

(Chinese National Human Genome Center , Beijing 100176 , China)

Abstract: Depending on the outputs of 10 080 DNA sequencing samples from the human genome project , technical features of capillary array electrophoresis are discussed. The results show that 88% of the total readlength were higher than 50 kbp/run. It means that most of the readlength could reach 500 bp or more with reasonable quality. The results were much better than expected when the technology of capillary array appeared about one year ago.

A key parameter affecting the quality of DNA sequencing is the purity of sequencing template. Therefore , a batch agarose slab electrophoresis was applied to check the purity before loading. Other conditions , such as concentration of template and loading amount , are discussed as well. Finally , comparisons based on the electrophoresis timing , the pre-treatment of samples , the loading procedures and the results analysis between capillary array electrophoresis and slab gel electrophoresis have been made. From the point of view of high-throughput DNA sequencing , especially in the field of genome research , the capillary array electrophoresis should be the better choice as a more efficient technical platform.

Key words : capillary array electrophoresis ; DNA sequencing ; genome