

DNA 序列测定分析仪器进展

胡松年 林伟 蒋琰 曾长青

(北京华大基因研究中心 北京 101300)

摘要 本文简述 DNA 序列分析仪器的原理和发展过程,着重介绍在测序领域中常用的两家厂商的仪器系统。

关键词 DNA 序列分析 垂直板自动测序仪 毛细管自动测序仪

自从 50 年代 DNA 双螺旋结构的发现,打开人们了解遗传的基本结构——基因的本质之门,分子生物学的最基本目标之一便是揭示包括人类自身在内的各种生命个体的基因结构即其 DNA 序列编码,从而从根本上了解生命现象的本质。经过分子生物学家、生化学家和工程师们几十年的不懈努力,解码 DNA 的技术从无到有,从慢到快,已经从繁杂的手工操作发展到全自动化的仪器分析。今天 1 台高通量的测序仪 24 个小时可以读出约 250 万的碱基排列,超出一个大肠杆菌全基因组序列的 1 半还多。本文将对 DNA 序列分析仪器的的发展过程做一简要综述。

DNA 测序技术主要有双脱氧链终止法(Sanger 法)和化学降解法(Maxam-Gilbert 法)¹。二者均建立于 70 年代中期,其发明人 Frederick Sanger 和 Walter Gilbert 于 1980 年共同分享一半的诺贝尔化学奖(另一半为 DNA 重组技术的奠基人 Paul Berg 所获)。双脱氧链终止法又称末端终止法或酶法,由于其操作简便,易于自动化和规模化,已经成为当今 DNA 测序的主导方法。其测序原理是对所测定的 DNA 序列进行循环聚合反应并使用 2',3'-双脱氧核苷三磷酸(ddNTP)作为聚合反应的链终止剂。ddNTP 与普通 DNA 的单体 dNTP 之间的差别在于其脱氧核糖的 3' 位置少一个羟基,因而虽可在 DNA 聚合酶作用下通过其 5' 三磷酸基团掺入到正在增长的 DNA 链中,但因缺乏 3' 羟基而不能同后续的 dNTP 形成磷酸二酯键。由此,当正在增长的 DNA 链末端碱基为 ddNTP 时,链延伸反应终止。这样,在合适条件下,所测 DNA 的聚合反应的产物便为一系列长度呈梯形分布的多核苷酸链,即其长度差别为一个核苷酸。通过凝胶电泳,各种长度的多核苷酸链被分离出来,通过对其长度和标记物的共同识别,便能依次读出所测 DNA 的每一个核苷酸的顺序。

DNA 测序的仪器设备发展至今,大致经过了三个阶段。第一阶段为非自动化的同位素标记(多用 P³²,也可用 P³³ 或 S³⁵)垂直板凝胶电泳测序。测定的主要策略是用放射性同位素掺入到所有不同长度的聚合片段,在四个聚合反应中分别加入四种碱基的双脱氧核苷三磷酸,即 ddATP, ddTTP, ddCTP 和 ddGTP。反应完成后将得到四组分别在 A, T, C, G 位置终止的不同长度片段。然后将四个反应的产物分别在四条凝胶电泳道上进行分离;经过同位素曝光后,便可对四条电泳道上的不同片段进行拼读,从而获得所测片段的碱基序列。垂直板凝胶电泳测序所需的设备只是电泳仪,从 70 年代开始使用以来,在单个基因和基因片段的测序、重组 DNA 的检测和突变分析等分子生物学的众多实验中起了极其重要的作用,并且完成一些小型基因组如某些病毒等的全部 DNA 测定,包括较为著名的 Φ X174², cytomegalovirus (CMV)³, 和 smallpox virus 以及线粒体和叶绿体 DNA 等。但这种测序的读长只有 200~300 个碱基,并且需要人工制备凝胶板和等待同位素曝光,比起日后发展起来的自动化测序仪,显然是费力、费时而不能适用于对大量的 DNA 片段的测定,特别是全基因组的序列分析,因而目前仍只用于单个实验室对少数基因或片段的分析。并且,随着自动化测序仪的快速发展,越来越多的实验室放弃这一用法而将所需测定的 DNA 样品送至拥有自动测序仪的专门的公司进行序列测定。

测序仪器设备发展的第二阶段的产物为自动荧光垂直板凝胶电泳测序仪。90 年代初,荧光标记和检测的技术开始被应用到 DNA 序列分析中,同时诞生了自动测序仪,使测序技术发生革命性的改变。其主要突破是采用四种颜色的荧光双脱氧核苷酸进行标记,从而用单泳道电泳便可将不同长度的标记产物分离并且检定。在荧光 ddNTP 标记体系中,美国应用生物系统公司 (Applied Biosystems)⁴ 的

Big Dye Terminator 和安玛西亚公司 (Amersham Biosciences)⁵ 的 ET Dye Terminator 是这种技术的主要代表。图 1 显示的是一段所测序列在仪器终端所显示的碱基排序。这种技术不但使测序的灵敏度大为提高,同时因为单泳道电泳便可拼读标记产物,从而大大降低测序模板的使用量。测序仪还在检测上实现了自动化,与普通电泳测序相比,具有快速、简便、灵敏度高,且无同位素污染的特点。这一技术已使得对大量 DNA 片段如 EST 库和小型基因组(病毒、细菌等)进行测序成为可能,从 1991 年开始的国际人类基因组 EST 测序计划便是用这类仪器开始的。90 年代后期完成的几项重要病毒和模式生物的全基因组测序结果,如 *Haemophilus influenzae*⁶, *E. coli* K12⁷, yeast⁸, *C. elegans*⁹ 等,也是用此类仪器完成的。这一测序技术的主要仪器产品是美国应用生物系统公司研制的几种不同电泳泳道的垂直板自动测序仪。其 90 年代最初的产品是 370 型垂直板自动测序仪,可进行 16 道的电泳测序;随后又在此基础上相继推出了具有多种电泳泳道选择的 373 型和 377 型垂直板自动测序仪,使仪器测序能力发展到最高 96 个电泳道。377 型是该公司最后一代垂直板自动测序仪,与前几代相比,它的自动化程度、测序能力及精确度都很高,读长最高可达 700~800bp,日分析能力可高达 300 个样品,并可根据样品量选择电泳泳道数(表 1),是目前垂直板自动测序仪类型中全球应用最广的型号。

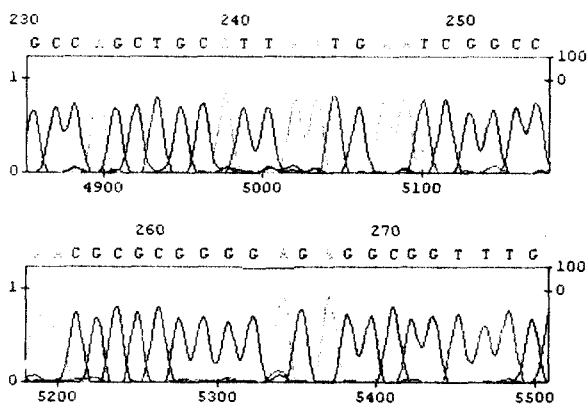


图 1 在自动测序仪上显示的 DNA 片段序列 A, G, T, C 四种碱基分别用四种颜色表示。峰图上面的灰绿色线段表明该峰值的数据可信度。M13 标准样品在 MegaBACE 1000 上的一段测序。

随着对大规模、高通量的 DNA 测序仪与日俱增的要求,第三代的自动化荧光毛细管凝胶电泳测序仪应运而生。90 年代中期,应用生物系统公司率先生产出单道毛细管电泳的 310 型测序仪。1998 年,

应用生物系统和安玛西亚公司相继推出 96 道的 ABI 3700 和 MegaBACE 1000 毛细管自动测序仪(表 1)。新型仪器用毛细管来代替传统的垂直电泳板,克服垂直板电泳需手工制胶和人工识别泳道的缺点,使测序技术完全走向低成本、高通量、自动化、规模化的道路。特别是,由于毛细管凝胶的单位面积的散热面积比板式凝胶的要大得多,不但提高测序的灵敏度和分辨能力,更可以使用比板式凝胶电泳高得多的电压,从而数倍地提高了电泳速度,使垂直板电泳每日 3 次的运行增加到每日 12 次。正是测序仪器的这一突飞猛进的发展,大大提高每台仪器的测序能力并降低测序成本,迅速推动大中型基因组的测序进程,使在短时间内获得大量基因组的序列信息成为可能。2001 年,参加国际人类基因组测序计划的 6 个国家的共 16 个中心共同宣布提前两年完成人类基因组 DNA 序列测定的工作框架图¹⁰;2002 年 5 月中国科学家率先公布水稻全基因组的工作框架图,并在《科学》杂志上以封面和 14 页的长文发表¹¹;这些举世瞩目的成就,主要是使用 ABI 3700 型或 MegaBACE 1000 型完成的。

表 1 当前常用 DNA 测序仪

Machines	Lanes	Runs /24 h	Maximum capacity /24 h	
			Samples	Bases
Applied Biosystems				
ABI PRISM 377	48/64/96	3	288	~173 Kb
ABI PRISM 3700	96	12	1152	~691 Kb
ABI PRISM 3100	16	12	192	~115 Kb
ABI PRISM 3730	48/96	24	2304	~1843 Kb
Amersham Biosciences				
MegaBACE 1000	96	12	1152	~634 Kb
MegaBACE 500	48	12	575	~316 Kb
MegaBACE 4000	384	12	4608	~2534 Kb

为进一步适应对不同规模和通量的测序要求,应用生物系统公司又在 2000 年生产出 16 电泳泳道的 3100 型毛细管自动测序仪,2002 年推出 48 和 96 两种泳道的 3730 型毛细管自动测序仪。除了对样品量的不同选择,在检测技术上又有很大改进。从 3700 型的单光束单侧激光激发,发展到 3100 型和 3730 型的双光束双侧激光激发、光栅分光装置和后置超薄 CCD 检测成像系统,使测序的准确度和灵敏度都有较大的提高,测序成本也显著降低。特别是 3730 型采用新型的 Pop-7 凝胶系统及长度更短的毛细管,不但使电泳的速度增加一倍,测序的平均读长也大为增加,使日分析能力比 3700 型(从样品数量到每个样品)的读长都大为提高(表 1)。

同样,安玛西亚公司继 MegaBACE 1000 型毛细

管自动测序仪之后,也于2000年和2001年相继推出48道的MegaBACE 500型和384道的MegaBACE 4000型。尤其是为适应更快速,更高效的测序需求而设计开发的MegaBACE 4000型,可以在2个小时内分析384个样品。与MegaBACE 1000型相比,测序量提高4倍(表1),而消耗的时间和工作量是相同的,成为目前基因组大规模测序中最高通量的设备。

从荧光垂直板凝胶电泳测序仪的使用到高通量的毛细管自动测序仪,全球的大型基因组研究中心均主要采用ABI或MegaBACE的体系,使得应用生物系统公司和安玛西亚公司已经成为全球主要的测序仪研究与制造公司。此外,其它公司也始终在积极研究开发各具特色的测序仪并且不断有新型号出现。例如,LI-COR公司(LI-COR Biosciences)的NEN Global IR²测序仪¹²,其96泳道垂直板电泳胶可以重复使用3次。特别是这一体系用一种新的名为A-cycloTerminatorsTM的非核苷酸末端终止试剂取代经典的ddNTP,并且采用红外的双色荧光染料进行标记,使之在99%的精确度下可以读到1000个碱基,因而非常适用于如gap拼接等需要很高读长的测序工作。MJ公司(MJ Research, Inc.)的BaseStation¹³,由于采用超薄(75 μm)光敏聚合凝胶、8道机械臂上样、水平电泳等技术,使其在测序质量、分析能力、自动化程度等方面都较垂直板测序仪有所提高,可以说是介于垂直板自动测序仪和毛细管自动测序仪之间的一类测序仪。此外还有贝克曼公司的8泳道毛细管自动测序仪CEQ2000和CEQ8000¹⁴以及日本岛津公司的384道毛细管自动测序仪RISA-384等¹⁵。

值得注意的是,除DNA测序功能外,各厂家的测序仪还都向着综合化、系统化的方向发展。目前上述公司的大部分机型大都兼具Genotyping、基因扫描、引物延伸等功能。例如,应用生物系统公司和安玛西亚公司分别提供专门用于Genotyping的SnapShot和SNuPe kit,分别用在ABI 3700/3730和MegaBACE 1000/4000型测序仪上进行引物延伸反应,对样品片段的单核苷酸多态位点进行Genotyping。

DNA测序技术从70年代发展到今天,高通量、高速度的毛细管自动测序仪已占据主导地位。但对于从事分子生物学和基因组学的研究者来说,应根据自己的研究情况选择通量和价格合适的测序仪。一般说来,进行大规模的基因组测序研究者,应用生物系统公司的3730型毛细管自动测序仪、安玛西亚公司的MegaBACE 4000型毛细管自动测序仪是目前市场的主流产品;而对于只需中小规模测序

的研究者来说,应用生物系统的377型垂直板自动测序仪和3100型毛细管自动测序仪、安玛西亚公司的MegaBACE 500型毛细管自动测序仪、MJ公司的BaseStation水平板自动测序仪、贝克曼公司的CEQ系列毛细管自动测序仪,则在性能和价格上比较合适。

2003年将是DNA双螺旋结构发现50周年的纪念。届时,国际人类基因组测序计划将在各个参加国同时向全球公布这一堪与人类登上月球想媲美的跨世纪壮举的提前完成。这在人类自然科学史上具有最大的公益性和最广泛的合作性的国际计划的启动和完成,不但使基因组学和生物信息学的研究成为新兴学科热点,也极大地推动了DNA序列分析技术和仪器的发展。继已经或接近全部完成的人类、水稻及果蝇基因组的项目,小鼠、大鼠、黑猩猩和鸡的基因组测序正在进行之中。可以预料,越来越多的模式生物和具有重要经济价值的动植物及微生物的全基因组规模的序列分析将更迅速地展开和完成。同时,在基因组水平上对基因的功能的研究,也向序列分析提出了更多和更高的要求。因此,对于未来的测序仪器的发展,一方面在其规模或通量以及降低成本方面将有更大的突破,使一个中、大型的全基因组能够在短期内用少数机器便可完成。另一方面,多功能、综合化,系统化亦将成为测序仪器的一个重要发展方向,使之能够更广泛地适应于分子生物学和基因组学研究的发展需要。

参考文献

- 1 Sambrook, J. and Russell, D.W. (2001) DNA sequencing. In *Molecular cloning, a laboratory manual*. 3rd ed. Cold Spring Harbor Laboratory Press. 3(12):1~12.90
- 2 Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocumbe, P.M. and Smith, M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977, 265 (5596):687~695
- 3 Rizzo, T.M. and Palukaitis, P. Nucleotide sequence and evolutionary relationships of cucumber mosaic virus (CMV) strains: CMV RNA 2. *J. Gen. 1988, Virol.* 69 (Pt 8):1777~1787
- 4 <http://www.appliedbiosystems.com/>
- 5 <http://www.amershambiosciences.com>
- 6 Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs,

- T., Hedblom, E., Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995, 269 (5223): 496-512
- 7 Blattner, F.R., Plunkett, G.III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. The complete genome sequence of *Escherichia coli* K-12. *Science* 1997, 277 (5331): 1453-1474
- 8 Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. Life with 6000 genes. *Science* 1996, 274 (5287): 546
- 9 *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 1998, 282 (5396): 2012-2018
- 10 The International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001, 409: 860-921
- 11 Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, ~Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 2002, 296(5565): 79-92
- 12 <http://www.licor.com/>
- 13 <http://www.basestation.com/>
- 14 www.beckman.com/
- 15 www.shimadzu-biotech.jp

Advances in DNA sequencing approach

Hu Songnian Lin Wei Jiang Yan Zeng Changqing

(Beijing Genomics Institute Beijing 101300)

Abstract This article briefly reviewed the development of DNA sequencers. It also introduced various models of two widely used sequencer systems.

Key words DNA sequence analysis Automated fluorescent sequencer Automated capillary sequencer

关注环境保护事业 再续国内外“双高”合作模式

安捷伦科技与中国科学院生态环境研究中心

续签生态中心—安捷伦亚太环境分析实验室合作协议

中国最具学术影响的环境研究结构中国科学院生态环境研究中心和全球领先的高科技跨国公司安捷伦科技(NYSAE:A)于2002年10月31日在北京续签了“生态中心—安捷伦亚太环境分析实验室”的合作协议,并继续在环境污染物分析和食品安全检测两大主要方面展开进一步的合作,共同推动中国环保事业的发展。

中国科学院生态环境研究中心—安捷伦亚太环境分析实验室成立于2000年,是安捷伦在亚太地区生命科学与化学分析领域的三大实验室之一。在建立后两年间,双方先后三次追加了对实验室的投资,使实验室由最初的规模发展为目前国内最高规格的应用实验室之一。目前实验室配备有国际最先进的气相色谱/质谱联用仪和电感耦合等离子质谱仪等分析仪器,具备了完成包括痕量环境污染物分析研究、食品安全检测方法等强大实力。

在合作协议续签仪式上,中科院生态环境研究中心副主任江桂斌研究员在致词中说:“生态中心—安捷伦亚太环境分析实验室是国内外“双高”合作模式的成功范例,即国内最高水平的环境科研机构 and 国外著名高科技公司之间的合作。这种合作模式有利于引进国外最先进的环境分析技术和分析方法,同时在合作实验室中将国外的技术和方法本地化,这将大大加速中国环境污染物分析及食品安全检测与国际接轨,更好的迎接WTO所带来的机遇与挑战。

安捷伦科技生命科学与化学分析事业部中国总经理牟一萍女士在发言中充分肯定了二年来合作所取得的成果,她说:“在过去两年中,实验室先后开展了低分辨率联机检测二恶英类方法、肉产品中瘦肉精检测方法、动物源性食品中氯霉素残留检测方法项目,为中科院生态环境研究中心和农业部等有关单位提供了大量农药残留检测分析的数据和水质分析数据,二年的合作硕果累累。

在合作协议续签仪式上,双方对共同组建亚太环境分析实验室的合作前景充满信心。

关于中国科学院生态环境研究中心,更多信息请访问 www.rcees.ac.cn

关于安捷伦科技,更多信息请访问 www.agilent.com.cn