文章编号:1000-5404(2006)06-0539-04

论著

难测序噬菌体基因组PaP1的大片段克隆及测序

李 明,申晓冬,丛延广,谭银玲,饶贤才,胡福泉 (第三军医大学基础医学部微生物学教研室,重庆市微生物工程重点实验室,重庆 400038)

提 要:目的 采用大片段克隆策略对难测序铜绿假单胞菌噬菌体PaP1基因组进行克隆及测序。**方法** 用限制性内切酶*Aat* Ⅱ将噬菌体PaP1基因组DNA切成几个大片段,分别进行末端修饰后与线性化的大片段克隆载体pYLTAC7连接,将连接产物电转化感受态细胞,筛选阳性克隆进行测序。**结果** *Aat* Ⅱ将PaP1基因组切成8个片段,通过大片段克隆共获得3个阳性克隆,测出插入片段大小分别为9 805、8 335 bp和3 465 bp,使该难测序基因组的测出量达到47 757 bp。**结论** 大片段克隆策略能够克服鸟枪法测序的不足,将有望获得噬菌体PaP1的全基因组序列。

关键词: 大片段克隆;噬菌体;基因组测序
中图法分类号: Q75;Q785;Q939.48 **文献标识码:** A

Application of large DNA fragment cloning and sequencing approach in the genomic sequencing of phage PaP1

LI Ming, SHEN Xiao-dong, CONG Yan-guang, TAN Yin-ling, RAO Xian-cai, HU Fu-quan (Department of Microbiology, College of Medicine, Third Military Medical University, Chongqing 400038, China)

Abstract: Objective To explore the genomic sequence of *Pseudomonas aerugonosa* phage PaP1 by large DNA fragment cloning and sequencing approach. **Methods** Genomic DNA was digested by restriction endonuclease *Aat* Ⅱ, were ligated with sequencing vector pYLTAC7 after modification. The ligation products were transformed into competent cells by electrotransformation, then the recombinant clones were screened and sequenced. **Results** Genomic DNA was cut into eight restriction fragments with *Aat* Ⅱ digestion, three of which were cloned successfully into vector pYLTAC7. The size of the three successfully inserted fragments was determined as 9 805, 8 335 and 3 465 bp respectively. **Conclusion** Large DNA fragment cloning approach can overcome the shortage of shotgun sequence assembly approach effectively and is helpful to complete the whole genomic sequence of phage PaP1.

Key words: large DNA fragment cloning and sequencing approach; phage; genome sequencing

随着基因组研究纵深与横向的不断发展,人们对病毒、细菌、蠕虫、果蝇乃至小鼠全基因组测序的重要性有了更新的认识。这些模式生物的基因组测序不仅为人类基因组计划提供了必要的经验,大量病原微生物的基因组测序加深了人们对这些病原微生物的理解,也为相关疾病防治提供了重要的分子生物学背景资料[1,2]。鉴于此,国际上纷纷展开对病毒、细菌等微生物的全基因组测序工作。

噬菌体作为细菌的病毒,在自然界中广泛存在。对噬菌体展开广泛深入的研究对于了解噬菌体与宿主的相互作用、揭示生物体的多样性以及治疗耐药菌感染、治理环境污染等方面都具有重要的意义[3]。国际上对噬菌体及其基因组学的研究一直是微生物学研究领域的热点之一。目前已报道的从形态学上鉴定的噬菌体有5 000多种,已完成全基因组测序的有302个。本室在前期工作中分离鉴定了3株铜绿假单胞菌噬菌体,分别命名为PaP1、PaP2和PaP3[4]。其中PaP2和PaP3的全基因组测序已经完成,并在GenBank上登录(PaP2登录号为AY575774,PaP3为AY078382)。然而,噬菌体PaP1的全基因组测序却遇到很大困难,建

基金项目:国家自然科学基金资助项目(30270065)
Supported by the National Natural Science Foundation of China(30270065)
作者简介:李 明(1979-),男,江西省武宁县人,博士研究生,助教,主要从事噬菌体基因组学方面的研究。电话:(023)68752239,E-mail:sotx7080@163.com
通讯作者:胡福泉,电话:(023)68752240,E-mail:hufuquan@yahoo.com.cn
收稿日期:2005-12-14;修回日期:2006-01-16

立鸟枪文库后，测序得到43个大小不等的重叠群序列（最大的26 152 bp），之后无法拼接，测序工作陷入困境。本研究旨在通过大片段克隆测序的方法获得噬菌体PaP1的未获得序列。难以测序这一特征预示着这一基因组具有特殊的结构特征，获得该基因组的全序列有益于揭示这一难测序基因组之谜。

1 材料与方法

1.1 *Aat* Ⅱ酶切片段的分离与处理

限制性内切酶*Aat* Ⅱ酶切PaP1基因组后，先用T_4 DNA聚合酶对酶切产生的片段进行平端化，通过脉冲场凝胶电泳分离各片段，然后用琼脂水解酶β-agarase Ⅰ回收纯化。再将纯化过后的各目的片段连接接头，完毕后再通过凝胶电泳和β-agarase Ⅰ的方法回收纯化各片段，最后用T_4多聚核苷酸激酶（T_4 Polynucleotide Kinase）对连接有HSB接头的各片段进行磷酸化修饰，酚氯仿抽提乙醇沉淀干燥后，溶于适量ddH_2O中，定量备用。

1.2 pYLTAC7载体的线性化与去磷酸化

大片段克隆载体pYLTAC7由中科院上海植物生理生态研究所董志诚博士惠赠。用限制性内切酶*Hind*Ⅲ酶切载体后，将酶切产物进行低熔点琼脂糖凝胶电泳，用β-agarase Ⅰ回收纯化。将纯化后的线性化载体进行去磷酸化，酚氯仿抽提乙醇沉淀干燥后，溶于适量ddH_2O中，定量备用。

1.3 *Aat* Ⅱ各酶切片段与pYLTAC7载体的连接

取修饰后的*Aat* Ⅱ各酶切片段和pYLTAC7载体按约3∶1（摩尔比）混合，加入适量的T_4 DNA连接酶及其缓冲液，于16 ℃连接过夜。

1.4 连接产物的沉淀与电转化

反应结束后，将连接产物用酚氯仿抽提后于上清中加入2 μl糖原（Glycogen），乙醇沉淀干燥后溶于适量ddH_2O中，取2 μl产物于18 kV/cm电压条件下电转化DH10B感受态细菌，电转结束后，将菌液涂布于含卡那霉素（25 μg/ml）和5%蔗糖的LB平板，37 ℃孵箱培养过夜。

1.5 重组质粒的筛选与鉴定

次日挑取平板上的菌落，扩大培养后用碱裂解法抽提质粒DNA，再溶于适量ddH_2O中（含RNase A）。取2 μl进行0.6%琼脂糖凝胶电泳，设立空质粒对照，挑选出明显滞后于空质粒条带的质粒DNA，再分别用*Bam*HⅠ和*Hind* Ⅲ酶切，将酶切产物进行0.6%琼脂糖凝胶电泳，凝胶成像仪扫描并保存图像。根据PaP1/*Aat* Ⅱ酶切图谱，初步判断插入片段的条带。

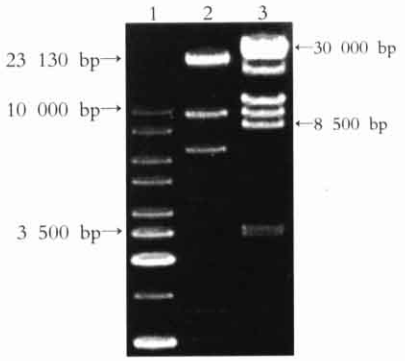
1.6 阳性质粒的测序与拼接

将筛选到的阳性质粒送人类基因组计划南方中心测序，测序结果参与拼接。

2 结果

2.1 *Aat* Ⅱ酶切PaP1基因组DNA的电泳结果

PaP1基因组DNA被内切酶*Aat* Ⅱ酶切后，酶切产物经脉冲场电泳后的结果见图1。



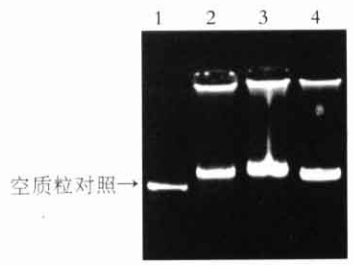
1：1 bp标准；2：λ/*Hind*Ⅲ标准；3：PaP1基因组/*Aat* Ⅱ酶切

图1 PaP1基因组/*Aat* Ⅱ酶切电泳图

从图1中可见，*Aat* Ⅱ能将PaP1基因组切成8条带。经Quantity One软件分析得出各条带大小，按条带大小分别命名为*Aat* Ⅱ①～⑧。将各条带的分子量相加，得知该基因组约有90 kb左右，是1个比较大的噬菌体基因组。通过电子酶切分析知道，在26 152 bp序列中没有*Aat* Ⅱ的酶切位点，所以最大的那条约30 kb的条带很可能包含此26 152 bp序列的片段，其余7条带就应该包含尚未拿到的序列。分别克隆这些大片段，然后进行测序，最后进行拼接，则有望拿到PaP1的未尽序列。

2.2 重组质粒的筛选与鉴定

将*Aat* Ⅱ②～⑧各目的片段与pYLTAC7载体的连接产物进行电转后，发现各平板均有十几至几十个不等的菌落生长。分别挑取所有阳性克隆扩大培养后，用碱裂解法抽提质粒进行0.6%琼脂糖凝胶电泳，发现有3个质粒明显滞后于空质粒条带（图2），分别命名为重组质粒①～③。

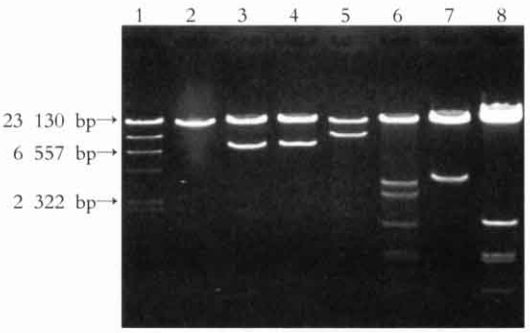


1：空质粒对照；2：重组质粒①；3：重组质粒②；4：重组质粒③

图2 重组质粒电泳图

将它们分别进行*Bam*HⅠ及*Hind*Ⅲ酶切，酶切产物的电泳图谱如图3所示。从图中可以看出，重组质粒①可以用*Bam*HⅠ及*Hind*Ⅲ切出1个约8 kb的片段（泳道3、4），该片段对应于PaP1/*Aat* Ⅱ酶切图谱中的片段5。从泳道5可见，用*Bam*HⅠ可以从重组质粒②中切出一10 kb左右的片段，该片段很可能是PaP1/*Aat* Ⅱ酶切图谱中的片段4，用*Hind*Ⅲ可将其切成至少6条小片段（泳道6）。从泳道7可见，重组质粒③可被*Bam*HⅠ切出1个约3.5 kb的片段，该片段很可能是PaP1/*Aat* Ⅱ酶切图

谱中的片段 8，该片段又可被 *Hind* Ⅲ切成 4 条小片段（泳道 8）。根据插入片段的大小所对应的 PaP1/*Aat* Ⅱ中的酶切条带位置，将这 3 个重组质粒分别命名为片段 4 质粒、片段 5 质粒和片段 8 质粒。



1：λ/*Hind* Ⅲ标准；2：pYLTAC7/*Hind* Ⅲ；3：质粒①/*Bam*H Ⅰ；4：质粒①/*Hind* Ⅲ；5：质粒②/*Bam*H Ⅰ；6：质粒②/*Hind* Ⅲ；7：质粒③/*Bam*H Ⅰ；8：质粒③/*Hind* Ⅲ

图 3　重组质粒①～③的酶切鉴定电泳图谱

2.3　重组质粒的测序结果

对含片段 4、5 及 8 的 3 个质粒进行测序，结果证实片段 4 的实际大小为 9 805 bp，与已测得的最大的 26 152 bp 重叠群中 7 144～16 948 bp 区段的序列完全一致；片段 8 的实际大小为 3 465 bp，与已测得的最大的 26 152 bp 重叠群中 4 606～8 070 bp 的序列完全相同。而片段 5 经测序后证实其实际大小为 8 335 bp，经过与 26 152 bp 重叠群进行比对后得知此序列是不同于 26 152 bp 的新序列。

3　讨论

自从 1977 年大肠杆菌噬菌体 ΦX174 的全基因组序列的公布以来，到目前为止完成全基因组测序并在 GenBank 上登录的噬菌体已经达到 302 个，且其数目还在不断增加。铜绿假单胞菌噬菌体 PaP1 我室自行分离鉴定的 1 株噬菌体，已经明确其核酸类型为双链 DNA。我们在前期工作中采用鸟枪法随机测序和重叠群组装策略对噬菌体 PaP1 基因组进行测序，遇到了相当大的困难。酶切图谱分析（图 1）表明该基因组大小约有 90 kb 左右，但实际测得的序列只能组装出 1 个最大的 26 152 bp 重叠群，剩下 42 个大小不一的重叠群无法再拼接。随后又测定了 600 多个序列（约 400 kb），用 Phred/Phrap/Consed 软件包进行组装时，全部又落入了前面 26 152 bp 的范围内，噬菌体 PaP1 的全基因组测序工作一度陷入困境。

实验证明鸟枪法能相对较快地测定基因组的序列，但它有一个很大的缺点，那就是当所测序基因组存在许多重复序列时，在拼接的过程中会面临困难。具体表现为：①一些本来存在于不同位置上的完全重复序列，可被计算机软件错误地放到同一位置，这样拼接的结果就是真实存在的重复序列被丢失了；②一些具有部分重叠的序列被错误地连接在一起。这样，使得众多小重叠群无法拼接到一起。因此，我们推测噬菌体 PaP1 基因组可能存在特殊的拓扑结构，使得传统的鸟枪法测序可能存在一定的局限性，必须结合其他一些测序策略才可能完成 PaP1 的全基因组测序。

大片段克隆测序是一种克服上述困难的行之有效的策略，已经广泛应用于植物基因组学、人类基因组计划等重大研究领域[5,6]。用该方法对 PaP1 基因组进行测序，可以有效避免重复序列在拼接时带来困难的缺陷，有望获得 PaP1 的全基因组序列。根据鸟枪法测序所获得的最大 26 152 bp 序列，通过电子酶切分析我们发现在这 26 152 bp 序列中没有 *Aat* Ⅱ的酶切位点。如果用此酶切全基因组 DNA，必然会产生 1 条含此 26 152 bp序列在内的大片段以及其他的片段（图 1），且这些片段应该是彼此独立的。这样，从 26 152 bp 序列两侧设计引物，向两侧做 DNA 步移，可望测完最大的那条约 30 kb 的片段。再分别从凝胶中分离其余片段，分别克隆，逐个独立测序，就应该可以测完全序列。各片段之间即使出现重复序列，也应该独立存在，而不应该视为同一序列。

本研究所用的大片段克隆载体为 pYLTAC7，是 Liu 等[7]于 1999 年结合 PAC 和 BAC 载体的特点，构建的一个具有转化功能的人工染色体（transformation-competent artificial chromosome，TAC）。该载体全长 22.9 kb，具有克隆 100 kb 以内大片段 DNA 的容量。但由于该载体 MCS 内仅含有 *Bam*H Ⅰ及 *Hind* Ⅲ等少数酶切位点，故我们设计了在酶切片段两端加上特异接头的策略，具体如图 4 所示。

PaP1 基因组 DNA 经 *Aat* Ⅱ酶切消化后，形成的是 3′突出末端，我们先利用 T4 DNA 聚合酶对各酶切片段进行平端化，分离纯化各片段后连接 HSB 接头（此接头含有 *Hind* Ⅲ、Sma Ⅰ和 *Bam*H Ⅰ 3 个酶切位点的序列，一端为平端，另一端为 *Hind* Ⅲ突出末端），然后就可以和线性化的 pYLTAC7/*Hind* Ⅲ载体进行连接了。如果是含有插入片段的重组质粒，我们就可以利用插入片段两端接头上的 *Bam*H Ⅰ酶切位点将插入片段切出来。在对各酶切片段进行平端化、加接头和磷酸化等处理过程中，考虑到无论是胶回收还是酚氯仿抽提纯化都会伴有目的片段不同程度的流失，我们采

用了低熔点胶回收以及利用糖原作为 DNA 分子的协沉剂[8]，一方面提高了回收效率，另一方面保证了大片段 DNA 分子构象的稳定性以利于连接反应的进行。此外，由于载体的分子量比较大(22.9 kb)，我们在实验中还采用了电转化的方法将重组分子导入感受态细胞中，最大限度地提高了转化效率，为获得阳性克隆及后续测序工作提供了可靠的保证。

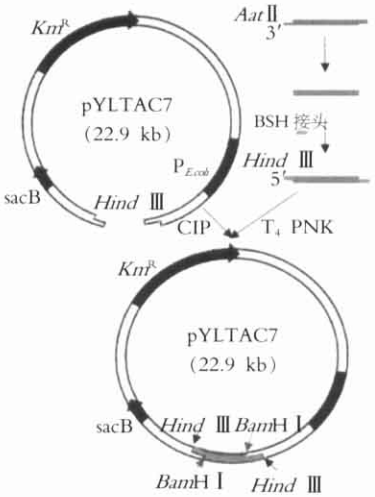


图4 加接头法大片段克隆 *Aat* Ⅱ 酶切片段示意图

通过大片段克隆的方法，我们目前获得了 PaP1/*Aat* Ⅱ的8个酶切片段中3个片段的阳性克隆，分别含有片段4、片段5和片段8。测序结果表明片段4和片段8质粒的实际插入片段大小分别为9 805 bp和3 465 bp，与已测得的最大的26 152 bp重叠群中部分序列完全相同。通过前面的理论分析我们知道在这26 152 bp序列中是没有 *Aat* Ⅱ 酶切位点的，PaP1 基因组经 *Aat* Ⅱ 酶切后，这个26 152 bp的完整序列被包含在1个长约30 kb的大片段中，而其他的片段中就必然含有尚未测序到的序列片段，不管它们是否与26 152 bp序列相同，它们都应该是彼此独立的未测到的新序列。同时，这两个质粒的插入片段都能够在PaP1/*Aat* Ⅱ的酶切图谱中找到相应大小的条带，据此可以说明两者均为 PaP1 基因组中独立于26 152 bp重叠群之外的大片段重复序列。而片段5质粒经测序后显示其插入片段的实际大小为8 335 bp，经 BLAST 发现此序列是已测到26 152 bp序列之外的不同序列，充分说明了按照大片段克隆的方法做下去，是完全有可能解决噬菌体 PaP1 基因组测序的难题的；同时也说明在噬菌体 PaP1 基因组中确实存在着较多的大片段重复序列。

我们知道，原核生物基因组中没有像真核生物基因组中那样高拷贝的、全基因组范围分布的重复序列。在已经完成测序的所有原核生物基因组中，尚未发现有任何一个基因组中有如此大的重复序列存在。那么这些大片段重复序列具有什么样的生物学意义？在最初的实验中，我们同时对 PaP1/*Aat* Ⅱ 的8个酶切片段都分别进行了克隆，结果只成功克隆到了其中的3个片段，大小分别为9 805、8 335 bp和3 465 bp，共21 605 bp，结合前面已测得的26 152 bp序列，目前我们一共在 PaP1 基因组中获得了47 757 bp的序列。在后续的实验工作中，我们将进一步优化实验条件，对剩下的 *Aat* Ⅱ 酶切片段进行克隆，以期获得噬菌体 PaP1 完整的基因组序列。解开 PaP1 的特殊结构之谜，揭示这些大片段重复序列存在的生物学意义，将具有非常重要的科学价值。

参考文献：

[1] ANKENY R A. Sequencing the genome from nematode to human: changing methods, changing science[J]. Endeavour, 2003, 27(2): 87-92.

[2] BABA Y. Development of novel biomedicine based on genome science [J]. Eur J Pharm Sci, 2001, 13(1): 3-4.

[3] 彭 丽，罗永艾. 噬菌体治疗细菌感染的研究[J]. 医药导报，2005, 24(6): 500-502.

[4] 张克斌，陈志瑾，金晓琳，等. 铜绿假单胞菌噬菌体的分离鉴定及耐噬菌体突变频率测定[J]. 微生物学通报，2002, 29(1): 40-45.

[5] MOON D A, MAGOR K E. Construction and characterization of a fosmid library for comparative analysis of the duck genome[J]. Anim Genet, 2004, 35(5): 417-418.

[6] BARBOSA A, DEMEURE O, URIEN C, *et al*. A physical map of large segments of pig chromosome 7q11-q14: comparative analysis with human chromosome 6p21[J]. Mamm Genome, 2004, 15(12): 982-995.

[7] LIU Y G, SHIRANO Y, FUKAKI H, *et al*. Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning[J]. Proc Natl Acad Sci USA, 1999, 96(11): 6535-6540.

[8] HARJU S, PETERSON K R. Sensitive ribonuclease protection assay employing glycogen as a carrier and a single xinactivation/precipitation step[J]. Biotechniques, 2001, 30(6): 1198-1204.

（编辑 张大春）