

蟑螂浓核病毒全核苷酸序列与基因组结构

郭海涛 张珈敏 胡远扬*

(武汉大学生命科学院, 武汉 430072. * 联系人, Email: wuemi@whu.edu.cn)

摘要 测定了黑胸大蠊浓核病毒(*Pf* DNV)复制型(RF) DNA 的核苷酸序列, 确定 *Pf* DNV 基因组全长为 5 454 个核苷酸, 基因组两端存在典型的发夹结构和倒置重复序列. *Pf* DNV 基因组正链有 4 个大的可读框(ORF), 负链含 3 个大的可读框, 并且可读框都集中在每条链的 5' 端. 两个可能的功能性启动子分别位于图距单位 3 和 97 处. 基因组存在基因重叠现象和内含子区域. 并对 *Pf* DNV 基因组与同科其他细小病毒进行了同源性比较. 推测 *Pf* DNV 基因组正链编码非结构蛋白, 负链编码结构蛋白.

关键词 黑胸大蠊浓核病毒 核苷酸序列 基因组结构

黑胸大蠊(*Periplaneta fuliginosa*)是我国分布最广的蟑螂种类. 黑胸大蠊浓核病毒(*Periplaneta fuliginosa* densovirus, *Pf* DNV)是胡远扬等人^[1]在国内首次报道并在国内外第 1 个正式分类鉴定的蟑螂浓核病毒. 该病毒呈二十面体球状, 直径 22 nm, 基因组为单链线状 DNA, 血清学性质不同于目前已报道的其他浓核病毒^[2]. 根据其理化性质和病理学特征, 确定该病毒属细小病毒科(Parvoviridae)、浓核病毒属(Densovirus).

蟑螂早在 3 亿年前已经出现在地球上, 研究以其为宿主的蟑螂细小病毒基因组结构并与其他动物、昆虫以及人类细小病毒基因组结构进行比较, 可以了解这类病毒的分子进化, 序列同源关系和遗传变异状况. 为此, 我们构建了 *Pf* DNV 全基因组克隆、系列酶切亚克隆和完整缺失克隆^[3], 用双脱氧手工测序法以及全自动荧光测序法测定了 *Pf* DNV 全基因组的核苷酸序列为 5 454 nt, 并对其基因组的结构进行了分析.

1 材料与方法

(i) 酶和反应试剂. 限制性内切酶、Klenow 片段、T4 DNA 连接酶购自 Boehringer Mannheim 公司. 构建完整缺失克隆用 Deletion Kit 购自 Takara 生物医学公司.

(ii) 菌种和载体. 大肠杆菌 DH5 α 和质粒载体 pUC119 购自 Takara 生物医学公司.

(iii) *Pf* DNV 提纯以及 DNA 制备. *Pf* DNV 病毒粒子由本室分离、增殖和保存. 基因组 DNA 的提取依照 Jourdan 等人^[4]的方法进行并略有改进.

(iv) *Pf* DNV 全基因组克隆的构建. *Pf* DNV DNA 加 dGTPs 尾, 质粒 pUC119 用 *Pst* I 酶切后加上 dCTPs 尾, 连接、转化和筛选按文献[5]所述进行. 得到病毒基因组以不同方向插入的两种重组质粒并将其中一株命名为 *Pf* DNV-pUC119.

(v) DNA 序列测定方法. 在 *Pf* DNV-pUC119 基础上, 进一步构建一系列完整缺失克隆以及酶切亚克隆. 采用双脱氧链终止法进行序列测定. 对于部分片段连接处, 依据已测得序列合成特异性引物进行测序, 以保证测序结果的可靠性. 测序战略如图 1 所示.

(vi) DNA 序列分析. 所有 DNA 数据处理均在计算机上进行. 序列测定所得各个片段的连接、可读框和特征序列查找、ssDNA 末端二级结构预测、同源性比较由软件 DNASIS, DNAsar, RNAstructure, BLAST 完成.

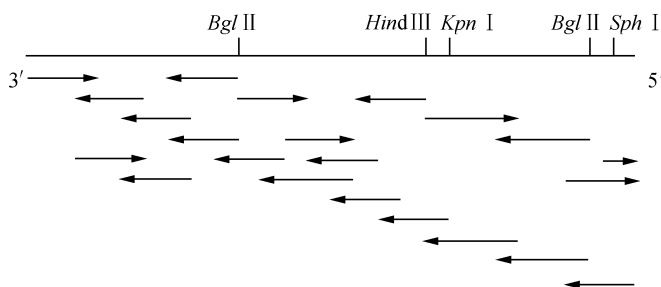


图 1 *Pf* DNV 基因组测序战略

2 结果分析与讨论

2.1 *Pf* DNV 基因组的核苷酸序列

利用计算机将各片段的测序结果连接并去除载体序列和 GC 尾巴得到 *Pf*DNV 全基因组的核苷酸序列, 共计 5 454 个核苷酸. 其中(A+T)%含量为 61.87%, (G+C)%含量为 38.13%. 本序列的 GenBank 编号为 AF192260.

2.2 病毒基因组末端结构分析

通过分析发现 *Pf* DNV 基因组两末端均存在具有相同核苷酸序列的 122 nt 回文序列, 形成典型的发夹结构(图 2 和 3). 已报道的其他细小病毒的发夹结构一般由 120 ~ 160 个碱基构成, 但是整个发夹结构又被两个短的回文结构中断, 后者对称地位于较大的回文结构中心的两边, 因此末端序列可折成“Y”型或“T”型结构. 而 *Pf* DNV 基因组末端回文序列不被小的回文序列所中断, 因此无“Y”型或“T”型结构. 细小病毒基因组结构显著的特点就是其两末端均有发夹结构, 这种发夹结构在细小病毒基因组 DNA 的复制过程中起着重要作用. 一般认为, 细小病毒 DNA 合成靠自我引发机制(self-priming mechanism)而起始, 这个引发机制是由于两端回文序列的存在而衍生的^[6].

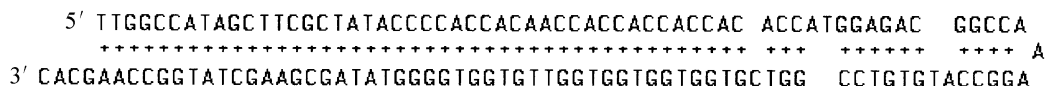


图 2 *Pf* DNV 基因组 5' 端发夹结构

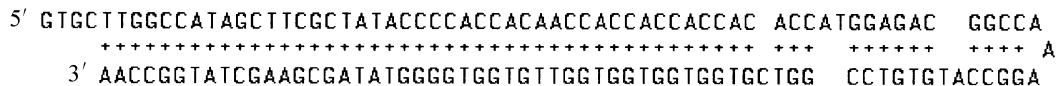


图 3 *Pf* DNV 基因组 3' 端发夹结构

对 *Pf* DNV 两个末端的碱基序列进行分析发现存在 201 nt 的倒转重复序列 (ITRs), 互补配对率达 97.6%, 不配对的碱基恰好是两末端发夹结构中不配对的碱基。

2.3 *Pf* DNV 基因组的读码框架

以 ATG 为起始密码, TAA, TAG, TGA 为终止密码, 对 *Pf* DNV 基因组正负链进行可读框查找, 共发现 7 个较大的 ORF. 正链有 4 个大的可读框(ORF1, ORF2, ORF3, ORF4), 分别编码 33, 62, 13, 31 ku 的蛋白. 负链上有 3 个大的可读框(ORF5, ORF6, ORF7), 分别编码 26, 63, 20 ku 的蛋白. 正负链上的 ORF 均集中在所在 DNA 链的 5' 端一侧, 并且有基因重叠现象存在.

2.4 可能的启动子区和 poly(A)位点

根据 Bensimhon 等人^[7]提出的真核启动子区的特点, 查找了可能的启动子区. 这些特征包括在帽子位点上游约 100 bp 处的 enabler 区, 在帽子位点上游 50 ~ 75 bp 处富含 GC 的激活区和帽子位点上游 30 bp 处的 TATA 盒. 如表 1 所示. 其中 P₃(位于图距单位 3 处的启动子)和 P₉₇ 是最具可能的功能性启动子, 因为它与鹿眼蛱蝶浓核病毒(*Junonia coenia* densovirus, *Jc* DNV)的功能性启动子的位置类似^[8], 而且含有必备的启动子元件.

表 1 可能的启动子序列分析

	图距单位	可能的序列	激活序列	TATA 盒
正链	3	GGTGGGGT(96 ^a , -102 ^b)	GGTGGGGG(151, -47)	TATAAAA(168)
	6	GGGGTGTA(272, -93)	GGGGTGTA(302, -63)	TATAAAA(335)
	36	CCGCGAAGA(1 902, -87)	CCCGTGTG(1 936, -53)	TATATCT(1 959)
负链	20	CCCGGCAT(1 215, -127)	CTGGCTGC(1 141, -53)	TATATCA(1 118)
	73	GGAGGAGC(4 085, -122)	GAGGGAAC(4 034, -69)	TATAAAT(3 993)
	93	GGGGACTC(5 144, -118)	GCGTGAGC(5 104, -78)	TATAGAA(5 056)
	97	GGTGGGGT(5 359, -102)	GGTGGGGG(5 304, -47)	TATAAAA(5 287)

a) DNA 序列的数目; b) 与推断的帽子位点的距离

在 ORF1, ORF2, ORF5, ORF6 内的可能的启动子区(图距单位 6, 36, 93, 73)之中, P₉₃ 距离下游的第 1 个 ATG 起始密码子有 543 nt; P₃₆ 以及 P₇₃ 的第 1 个同框 ATG(2 056, 4 809 nt)之前有 2 或 3 个 ATG 密码, 但都很快终止于其他读码框架中, 除非有 mRNA 剪切存在, 否则翻译时就需要核糖体越过多个上游 ATG 到达第 1 个同框 ATG 位置; 而 P₂₀ 下游无大的可读框存在; 在 P₆ 下游最近的一个 ATG(335 nt)可以起始编码一个氨基端比 ORF1 编码蛋白少 6 个氨基酸的多肽. 由于缺少 *Pf*DNV 基因组转录的精确数据, 所以不能断定这些可能的启动子是否具有功能.

以 AATAAA 作为一致序列在正链上发现有 3 个可能的 poly(A)位点. 分别位于图距单位 50, 69, 91 处. 其中图距单位 69 具有附近的 CAYTG 序列(3 816 nt)以及下游的 G/T 富含区(3 819 nt)^[9], 并且附近序列可以折叠成较为稳定的二级结构, 因此是最具可能的 poly(A)位点. 图距单位 50 缺少 CAYTG 序列且与 ORF2 的终止密码 TAA 重叠, 这种现象也出现在 *Jc* DNV 基因组中^[8]; 图距单位 91 缺少 CAYTG 序列, 并且距离正链上最后一个可读框 ORF3 的终止密码 TAG(3 115 nt)有近 2 000 碱基的距离. 同样推理负链上的 poly(A)位点位于图距单位 38 处.

以上分析结果表明: 在启动子 P₃ 和 P₉₇ 的控制下, *Pf* DNV 基因组将起始转录两条大的 mRNA 链.

2.5 5' 端供体和 3' 端受体剪切位点

以 5' 端供体点的一致序列(A/CAGGTA/GAGT)以及 3' 端受体点的一致序列 {(C/T)_n XCAGGC/T}^[10]在计算机上查找 *Pf* DNV 基因组中的 5' 和 3' 剪切位点, 结果如表 2.

表 2 5' 供体和 3' 受体剪切位点的位置

正链				负链			
5' 供体点		3' 受体点		5' 供体点		3' 受体点	
位置	序列	位置	序列	位置	序列	位置	序列
2 491	<u>CAGGTGGTG</u>	3 290	<u>TCTGAACAGGT</u>	4 533	<u>CAGGTACGT</u>	4 399	<u>CCTCTTTCAGGT</u>
2 830	<u>AAGGTAGTG</u>			4 400	<u>CAGGTGACG</u>	4 180	<u>TCTCACTCAGGT</u>
3 562	<u>CAGGTGGAC</u>			4 181	<u>CAGGTGTAC</u>		
				4 118	<u>CAGGTGGTG</u>		

正负链基因组转录出的 mRNA 在剪切机制调控下, 可能翻译产生的蛋白见表 3.

表 3 mRNA 剪切后推测表达的多肽

	5' 供体点/nt	3' 受体点/nt	内含子长度/nt	表达的蛋白/ku
正链	2 491	3 290	798	51, 31, 33
	2 830	3 290	459	62, 31, 33
负链	4 533	4 399	133	85
	4 533	4 180	352	77
	4 400	4 180	219	26, 55

2.6 Pf DNV 基因组编码蛋白与其他细小病毒同源性的比较

通过 BLAST 程序将 *Pf* DNV 基因组编码的蛋白与 NBRF 蛋白数据库进行同源性搜索, 发现 ORF2, ORF5, ORF6 编码的蛋白与其他细小病毒有同源性. 其中 ORF2 编码蛋白 82 ~ 531 aa 与蔗螟浓核病毒(*Diatraea saccharalis* densovirus, *Ds* DNV), *Jc* DNV, 大蜡螟浓核病毒(*Galleria mellonella* densovirus, *Gm* DNV)的非结构蛋白(NS 蛋白)同源性较高; 331 ~ 507 aa 与埃及伊蚊浓核病毒(*Aedes* DNV)NS 蛋白同源性较高; 380 ~ 499 aa 与腺伴随病毒 AAV-2, 人红细胞细小病毒 B19, 猪细小病毒 PPV, 牛细小病毒 BPV NS 蛋白的 GKRN 保守区有较高同源性. ORF5 编码蛋白 148 ~ 241 aa 与 *Ds* DNV, *Jc* DNV, *Gm* DNV, 腺伴随病毒 AAV-5, 貂肠炎病毒 MEV, PPV, CPV 结构蛋白(VP 蛋白)的 PGY 保守区同源性较高. ORF6 编码蛋白 78 ~ 570 aa 与 *Jc* DNV VP 蛋白有较高同源性; 143 ~ 570 aa 与 *Ds* DNV, *Gm* DNV VP 蛋白同源性较高. 由此可以推测 *Pf* DNV 基因组的 ORF2 可能编码非结构蛋白, ORF5 和 ORF6 可能编码结构蛋白.

黑胸大蠊浓核病毒全基因组序列的测定, 从分子水平上为其分类鉴定提供了依据. 细小病毒基因组线性单链 DNA 末端结构是其标志性的结构, 它们都有回文序列, 但不同的病毒基因组末端结构又有差别, 大多数的浓核病毒和腺伴随病毒都有末端倒置重复序列, 而绝大多数自主型细小病毒没有 ITRs^[11]. *Pf* DNV 基因组末端回文结构和倒转重复序列与其他已报道的细小病毒基本一致, 这类结构与细小病毒的 DNA 复制相关, 由此我们推测 *Pf* DNV 的复制采取“自身引物, 发夹转移”的方式^[6], 首先以折叠的 3' 端回文序列为引物合成病毒 DNA 中间体, 最后通过特异性位点的切割而得到子代病毒 DNA. *Pf* DNV 正负链的蛋白编码方式和编码能力基本一致, 编码区都集中在正负链的 5' 端; 基因之间存在重叠现象; 正负链基因组都存在内含子区域; *Pf* DNV 基因组在位于正负链 5' 端的两个启动子控制下分别起始转录两条大的 mRNA 链. 这些特点与 *Jc* DNV 非常相似^[8]. 从基因组编码蛋白与其他细小病毒的同源性比较来看, 基因组正链可能主要编码病毒的 NS 蛋白, 而负链可能主要编码 VP 蛋白, NS 蛋白和 VP 蛋白的种类和比例同时受到 mRNA 剪切方式的调控. 这一结论正在进一步论证之中.

致谢 本工作为国家自然科学基金资助项目(批准号: 39470032).

参 考 文 献

1 胡远扬, 梁东瑞, 黄远达, 等. 发现黑胸大蠊(*Periplaneta fuliginosa*)非包涵体病毒. 武汉大学学报, 1991, 1: 114

2 Hu Y, Zheng J, Iizuka T, et al. A densovirus newly isolated from the smoky-brown cockroach *Periplaneta fuliginosa*. Arch Virol, 1994, 138: 365 ~ 372

3 胡远扬, 张珈敏, 朱丽华, 等. 黑胸大蠊浓核病毒全基因组克隆载体及酶切亚克隆、完整缺失克隆的构建. 中国病毒学, 1996, 11: 12 ~ 17

- 4 Jourdan M, Jousset F X, Gervais M, et al. Cloning of the genome of a densovirus rescue of infection virions from recombinant plasmid in the insect host *Spodoptera littoralis*. *Virology*, 1990, 179: 403 ~ 409
- 5 萨姆布鲁克 J, 弗里奇 E F, 曼尼阿蒂斯 T, 著. 分子克隆实验指南. 第 2 版. 金冬雁, 黎孟枫, 等译. 北京: 科学出版社, 1995
- 6 Bando H, Choi H, Ito Y, et al. Terminal structure of a densovirus implies a hairpin termini transfer replication which is similar to the model for AAV. *Virology*, 1990, 179: 57 ~ 63
- 7 Bensimhon M J, Gabarro-Arpa J, Reiss C, et al. Physical characteristics in eucaryotic promoters. *Nucleic Acids Res*, 1983, 11: 4521 ~ 4540
- 8 Dumas B, Jourdan M, Pascaud A M, et al. Complete nucleotide sequence of the cloned infectious genome of *Junonia coenia* densovirus reveals an organization unique among parvoviruses. *Virology*, 1992, 191: 202 ~ 222
- 9 Birnstiel M L, Busslinger M, Strub K. Transcription termination and 3' processing: the end is in site. *Cell*, 1985, 41: 349 ~ 359
- 10 Mount S M. A catalogue of splice junction sequence. *Nucleic Acids Res*, 1982, 10: 459 ~ 472
- 11 Chen K C, Shull B C, Moses E A, et al. Complete nucleotide sequence and genome organization of bovine parvovirus. *J Virol*, 1986, 55: 574 ~ 582

(1999-10-22 收稿, 2000-03-01 收修改稿)