

# 杆状病毒基因组结构及其特征\*

高建明 彭建新\*\* 洪华珠

(华中师范大学昆虫学研究所 武汉 430079)

摘要: 综述杆状病毒基因组结构、特征、组织方式。

关键词: 杆状病毒, 基因组, 结构, 特征, 组织方式

中图分类号: Q93 文献标识码: A 文章编号: 0253-7654 (2003) 04-0099-05

杆状病毒是一类 DNA 病毒, 病毒粒子呈杆状, 基因组为双链环状 DNA 分子, DNA 以超螺旋形式压缩包装在杆状衣壳内, 大小在 90 ~ 180Kb 之间。杆状病毒是节肢动物病原体且主要寄生在鳞翅目昆虫中。目前杆状病毒已开发作为高效、安全的杀虫剂广泛应用于害虫防治, 也已开发作为应用最广泛的表达载体系统, 用于各种外源基因的表达。由于 DNA 测序技术、分子生物学技术、计算机技术的发展与应用, 杆状病毒基因组结构分析正取得飞速发展。已有苜蓿银纹夜蛾多粒包埋核型多角体病毒 (*Autographa californica* Multinucleocapsid Nuclear Polyhedrosis Virus, AcMNPV)、家蚕核型多角体病毒 (*Bombyx mori* nucleopolyhedrovirus, BmNPV)、黄杉毒蛾多粒包埋核型多角体病毒 (*Orgyia pseudotsugata* Multinucleocapsid Nuclear Polyhedrosis Virus, OpMNPV)、中国棉铃虫单粒包埋核型多角体病毒 (*Helioerpa armigera* single nucleocapsid nucleopolyhedrovirus, HaSNPV)、甜菜夜蛾多粒包埋核型多角体病毒 (*Spodoptera exigua* Multinucleocapsid Nuclear Polyhedrosis Virus, SeMNPV)、舞毒蛾多粒包埋核型多角体病毒 (*Lymantria dispar* Multinucleocapsid Nuclear Polyhedrosis Virus, LdMNPV)、斜纹夜蛾多粒包埋核型多角体病毒 (*Spodoptera litura* Multinucleocapsid Nuclear Polyhedrosis Virus, SiMNPV) 和 XcGV (*Xestia c-nigrum* Granulovirus) 等 8 种杆状病毒基因组全序列得到测定。本文综述杆状病毒基因组结构、特征、组织方式。

## 1 杆状病毒基因组结构特征

杆状病毒基因排列非常紧凑, 基因间除了同源重复区 (homologous repeat region, hr) 序列外, 只有很少的间隔, 除了 *bro* (baculovirus repeat open reading frames related to Ac2) 基因外, 重复基因很少, 也几乎没有插入序列 (内含子)。这些特征显然有利于增大基因组表达量。另外, 杆状病毒基因间有少量的重叠, 其早、晚基因分布在整个基因组, 并无成簇现象。除了这些共同特征以外, 其它性质见表 1<sup>[1-6]</sup>。

从表 1 可以看出, 基因组大小从 BmNPV 128.4kb 到 XcGV 178.7kb。除了 LdMNPV 和 XcGV 以外, 其它病毒基因组大小均相似, 差别主要是由于 *hr* 和 *bro* 基因数量差异引起的。比如 XcGV 基因组最大是因为具有许多重复的基因: 3 个 *p10*、7 个 *bro*、4 个 *AcM145/150* 的同源基因、4 个增强蛋白基因, 这些重复基因占了整个基因组 17%, 基因间序列 21 Kb, 约占基因组 12%, 其中有约 5 Kb *hr*。LdMNPV 也有 13 个 *hr* 和 16 个

\* 湖北省自然科学基金资助项目 (No. 2000 J 112)

\*\* 联系人 Tel: 027-87863522, E-mail: jianxinpeng@21cn.com

收稿日期: 2002-07-26, 修回日期: 2002-10-30

*bro*。HaSNPV 的 G + C 含量 (%) 最低, 与 SIMNPV、BmNPV、AcMNPV 和 XcGV 相近, G + C (%) 的不同反映了密码子的使用频率不同。

表 1 杆状病毒基因组结构特征

特征类别	HaSNPV	SIMNPV	AcMNPV	BmNPV	OpMNPV	LdMNPV	SeMNPV	XcGV
总碱基对数 (kb)	131.4	139.3	133.9	128.4	132.0	161.0	135.6	178.7
G + C 含量 (%)	39	42.7	41	40	55	58	44	41
总的 ORF 数	135	141	154	136	152	163	139	181
独有 ORF 数	20	29	11	1	16	29	17	82
同源重复区数	5	17	8	7	5	13	6	8
早期基因个数	33	11	65	12	61	12	34	13
晚期基因个数	60	52	72	78	64	79	72	84
早 + 晚基因个数	9	8	29	7	26	6	14	2
未鉴定启动子	49	53	47	35	58	78	53	84
<i>bro</i> 基因数	3	2	3	5	3	16	无	7

分析杆状病毒基因组, 病毒 DNA 含有同源重复区 (*hr*)。 *hr* 是杆状病毒基因组极具特色的结构, 由重复序列构成, 包括顺向重复序列和不完全反向重复序列 (回文结构), 不同杆状病毒 *hr* 序列同源性一般都比较低。 *hr* 有两方面的功能: 一是其增强子功能, 二是在病毒 DNA 复制过程中作为复制原点<sup>[1-8]</sup>。

## 2 杆状病毒基因类型

**2.1 按表达时间分类** 杆状病毒基因按表达时间可分为极早期基因、早期基因、晚期基因和极晚期基因 (或超量表达晚期基因)。前面两类称为早期基因, 后面两类则称为晚期基因。

早期基因的表达不需要新的蛋白质合成, 且先于 DNA 复制, 其转录由宿主细胞 RNA 聚合酶催化。早期基因主要为病毒 DNA 复制、晚期基因表达提供必需的蛋白因子。早期基因最显著的结构特点是含有共同的基元序列 CAGT 或者 CGTCC, 这些保守的序列通常为转录起始位点, 它们含有转录信号。杆状病毒早期基因启动子相对多样性基元结构, 主要是与宿主 RNA 聚合酶及其转录因子间相互作用和适应的结果。已鉴定出的早期基因有: *egt*、*ie-1*、*39K*、*gp67*、*pcna*、*me53*、*PE-38*、*ie-0*、*lef-1* (late expression factor 1)、*-2*、*-3*、*-4*、*-7*、*-8*、*-10*<sup>[9]</sup> 及 DNA 聚合酶、解旋酶基因。

晚期基因表达依赖于病毒基因组 DNA 的复制, 并通过病毒编码或病毒改造过的 RNA 聚合酶转录, 同时需要 *lef* 类调控其表达。晚期基因启动子的特征序列是 TAAG, 也是转录起始位点。晚期基因有: *e66*、*p87*、*ptp*、*p47*、*vp39*、*p10*、*sod*、*p74*、*p6.9*、*p25*、*gp37*、*p41* 及多角体蛋白、芋螺毒素、增强蛋白、组织蛋白酶、几丁质酶、泛素等的编码基因<sup>[9]</sup>。

杆状病毒有的基因含有 2 个甚至多个启动子, 这些基因有早、晚期启动子基元序列。表达往往从早期持续到晚期, 如 *pk* (蛋白激酶)、*lef-6*、*-7*、*ie-1*、*p35*、*39k* 等。有的基因有 2 个或者 3 个晚期启动子, 如 *vp39*、*gp67*<sup>[9,10]</sup>。

**2.2 按功能分类** 杆状病毒基因的功能是多种多样的, 根据它们功能的不同, 可进行以下分类。

与病毒复制有关的基因<sup>[11,12]</sup>: DNA 聚合酶及解旋酶基因、*lef-1*、*-2* (XcGv 无)、

-3、-7、*ie-1*、*p35*、*pe-38*、*hcf-1*。前6种是DNA复制必须的基因，后4种对DNA复制有促进作用，仅存在于部分杆状病毒。编码连接酶、拓扑异构酶的基因还未发现，因而很可能由宿主细胞提供。但在LdMNPV、XcGV中有与其它生物连接酶基因的同源序列<sup>[6-8]</sup>。

与病毒DNA表达调节有关的基因：*vlf*（AcMNPV除外）及*ie-2*、*lef-4*、-5、-8、-9（AcMNPV无*lef-8*、-9）<sup>[13]</sup>、*p47*等*lef*类<sup>[9,10]</sup>。

与辅助功能有关的基因：这类基因不是病毒生存、DNA复制所必需的，包括*sod*（SIMNPV除外）、*ptp*、*egt*（XcGV除外）及几丁质酶、组织蛋白酶、成纤维细胞生长因子、蛋白激酶的编码基因（HaSNPV除外）。这里特别提到组织蛋白酶及几丁质酶，它们一起作用使虫体液化，有利于病毒的传播<sup>[14]</sup>。而EGT催化葡萄糖基团转移到昆虫蜕皮激素或其它蜕皮甾皮内激素，使激素失活，阻碍幼虫的蜕皮和蛹化。*egt*是杆状病毒中目前发现唯一在昆虫虫体水平通过对激素作用调控受感染宿主的基因<sup>[9,15]</sup>。

与病毒结构有关的基因：8种杆状病毒共有的与病毒结构有关的基因有*gp41*、*vp39*、*p87/80*（XcGV无）、*p6.9*、*PDV-e66*、*p10*<sup>[1-8]</sup>。

与核苷酸代谢有关的基因：在LdMNPV、SIMNPV、OpMNPV、SeMNPV中存在与核苷酸代谢有关的基因：*dUTPase*、*rr1*（编码核苷酸还原酶大亚基）和*rr2*（编码核苷酸还原酶小亚基）。3个基因编码的酶对病毒DNA的复制是有利的，因为细胞处于不分裂状态，其dUTP合成途径不畅通，病毒可利用此酶合成核苷酸。*rr1*和*rr2*催化rNDP还原为dNDP，然后dNDP被磷酸化产生dATP、dGTP、dCTP、dUTP。*dUTPase*催化dUTP为dUMP，从而排除dUTP插入DNA引起诱变的可能，而且dUMP是dTTP合成前体物。但另4种病毒无此类基因，说明并不是所有杆状病毒复制都需要病毒自身编码此类酶<sup>[1-8]</sup>。

与抗细胞凋亡有关的基因：*Lsp40*（LsNPV）、*p35*（BmNPV、AcMNPV）、*iap*。

与宿主域有关的基因：*hcf-1*（LdMNPV和BmNPV）、*p35*、解旋酶基因。

### 3 不同杆状病毒基因组的比较

3.1 杆状病毒基因组同源性比较 为了全面反映8种杆状病毒基因组同源性，从以下两个方面进行比较：病毒之间同源基因数量以及编码蛋白质氨基酸序列同源性（表2）。

表2 杆状病毒基因组同源基因数量及编码蛋白质氨基酸序列同源性（%）

C	SIMNPV	AcMNPV	BmNPV	OpMNPV	LdMNPV	SeMNPV	XcGV	HaSNPV
HaSNPV	ND	100	98	94	94	103	69	/
XcGV	ND	84	80	76	93	72	/	40
SeMNPV	ND	103	99	102	104	/	ND	47
LdMNPV	ND	94	91	95	/	45	ND	46
OpMNPV	ND	126	121	/	ND	40	34	41
BmNPV	ND	115	/	55	ND	41	ND	41
AcMNPV	96	/	93	56	41	41	33	41
SIMNPV	/	ND	ND	ND	ND	ND	ND	ND

注：“/”以上数字表示同源ORF数目，“/”以下数字表示氨基酸序列同源性；ND，not determined.

由表2可知HaSNPV与LdMNPV、SeMNPV蛋白质氨基酸序列同源性较高，分别达到46%和47%，而AcMNPV、BmNPV和OpMNPV的较高，尤其是AcMNPV和BmNPV竟达93%。BmNPV仅缺乏AcMNPV下列基因：*Ac3*、7、48（*etm*）、49（*pcna*）、70（*hcf*-

1)、86、134 (p94)。差异大多是由于 BmNPV 基因组中有较多的连续的 3bp 插入<sup>[2]</sup>。

已测序的 8 种杆状病毒基因组总共有 383 个不同的基因, 其中 205 个是独有的, 65 个是很保守的。有人认为大约 70 个基因是杆状病毒生存的最低要求。参与病毒 DNA 复制、晚期基因转录、病毒结构、病毒传播的基因约占保守基因的 1/3。其中又以多角体蛋白基因最为保守, 同源性达到 83%, 其次是泛素, 75%, 再次为 *sod*, 70%。早期基因保守性较低, 如: *ie-1*、*ie-0* 等同源性仅 35%, 晚期基因如: RNA 聚合酶基因、*lef-4*、*-8*、*-9*、*p47*、*ulf-1* 保守性很高<sup>[4]</sup>。联想到早期启动子多样性基元结构, 人们有理由相信: 宿主域由早期基因决定, 病毒不得不从一开始就适应宿主的转录系统。

**3.2 不同杆状病毒特异的基因** 并不是每种杆状病毒都具有上述所有基因, 这里特别提及两个重要的基因: *ie-2* 和 *gp64/67*。这两个基因仅在 OpMNPV、BmNPV、AcMNPV 有。*ie-2* 能反式调节早期基因、晚期基因和极晚期基因的表达, 估计缺乏 *ie-2* 的 5 种病毒体内有其同源序列, 这还有待于研究。而 *gp64/67* 对于病毒粒子吸附与侵入等起着十分重要的作用。不过人们已经发现 *Ha133*、*Ld130*、*Se8*、*SU136* 具有与 *gp64/67* 同样的功能, 并且发现前 3 者同源性达 40% 以上<sup>[4-7]</sup>。在 SIMNPV 中, 泛素蛋白基因与 *gp37* 基因重叠, 其编码的融合蛋白 N 端是泛素, C 端为 GP37。*SU105* 和 *Se22*、*23*、*24* 一些片断的序列同源。其 hr 有两类回文序列, 不同于其他杆状病毒<sup>[7]</sup>。在 XcGV 基因组中, 有 1 个与 DNA 连接酶同源的基因、2 个 DNA 解旋酶基因 (1 个与酵母线粒体解旋酶相似, 另 1 个与 AcMNPV 解旋酶相似)、4 个增强蛋白基因。*XcGV 5*、*19*、*83* 与 *NPV10* 有一些同源性<sup>[8]</sup>。在 SeMNPV 基因组中有 2 个 *odv-e66* 的同源序列, *Se114* 和 *Se57*。*Se114* 两侧分别是 *Ac40* 和 *Ac43* 的同源序列, 而 *Se57* 两侧则是 *Ac108* 和 *LsNPV13* 的同源序列。*Se57* 和 *Se114* 的同源性仅有 32%, 因为不同杆状病毒的同源序列出现在 *Se57* 位置附近, 它很可能经历了较大的重排, 是独立进化而来, 其和 LdMNPV 亲缘关系比与 *Se114* 更密切。通过分析还发现, 有两个基因编码 P26, 分别是 *Se87*、*129*, 它们位置不同<sup>[5]</sup>。HaSNPV 缺失 *ie-2*、*gp64/67*、*p35*、*lef-12* 基因, 这 3 个基因在 LdMNPV、SeMNPV 和 XcGV 中也不存在。而且只有 HaSNPV 缺乏 *ptp* 基因。除了 3 个 *bro*、2 个 *iap* 基因以外, 无其它基因重复, 其 hr 序列和结构都相当独特<sup>[4]</sup>。在 LdMNPV 中具有 2 个增强蛋白、1 个芋螺毒素、1 个 *DUTPase*、1 个 *rr1*、2 个 *rr2*、1 个 DNA 连接酶同源基因, 1 个酵母线粒体解旋酶同源基因。LdMNPV 是 NPVs 中唯一发现有增强蛋白的杆状病毒。另外, *Ld37* 和 *47* 均同源于 *Ac25*, 但彼此同源性仅 26%。在 OpMNPV 中有 3 个 *iap*、2 个 *ptp* 及 2 个芋螺毒素基因<sup>[3]</sup>。在 BmNPV, *Bm22* 和 *81* 仅同源于 *Ac2* N 端的 136 个氨基酸, 彼此同源性为 69%<sup>[2]</sup>。

### 3.3 XcGV 和 NPVs 基因的比较

其同源基因数量见表 2。GV 和 NPVs 基因组中有许多同源性很高的基因。首先是泛素, 同源性达 79%, 其它依次是几丁质酶、*sod* 和 *lef-9*。*XcGV* 具有 13 个 AcMNPV *lef* 类基因, 包括 *lef-1*、*-2*、*-4*、*-5*、*-8*、*-9*、*-10*、*-11*、DNA 多聚酶、*p47*、*39k* 等, 缺失 *lef-3*、*-6*、*-7*、*-12*、*p35*、*ie-2* 和 *hcf-1*。另外, *XcGV* 有 NPV16 个结构蛋白基因。这些结构蛋白基因是包含体蛋白基因、蛋白激酶基因、*p10*、*odv-e18*、*-e25*、*-ec27*、*-e56*、*-e66*、*p74*、*p24*、*p6.9*、*vp39*、*p91*、*gp41*、*vp1054* 等。平均来说, 这些序列与 AcMNPV/OpMNPV 同源性为 33%, 但缺失 *pp34*、*p80/87*、*ptp* 和 *gp64*<sup>[8]</sup>。

#### 4 杆状病毒基因组组织方式

选取杆状病毒中的保守基因,按照它们在基因组物理图谱中的位置和转录方向进行比较,发现在线性DNA中间区域,保守性最高,基因组左侧(基因1~30)有许多倒位和移位,右侧(70~100)有高度的漂移(scrambling)。HaSNPV、SeMNPV、LdMNPV基因组组织方式相似,而AcMNPV、BmNPV、OpMNPV的相似。综合单基因、同源基因数量、编码蛋白氨基酸序列同源性、基因组组织方式的比较,可以得出如下结论:HaSNPV、SeMNPV、LdMNPV亲缘关系较近,而OpMNPV、AcMNPV、BmNPV亲缘关系较近。

#### 参考文献

- [1] Ayres M D, Howard S C, Kuzio J. 1994, 202: 586 ~ 805.
- [2] Gomis, Majima K, Maeda S M. J Virology, 1999, 80: 1323 ~ 1337.
- [3] Ahrens C H, Russell R L Q, Funk C J. Virology, 1997, 229 ~ 399.
- [4] Chen X W, Ijkel W F J, Turchini R. Virology, 2001, 82: 241 ~ 257.
- [5] Wilfred F J, van Strien E A, Helden J G, M. J Virology, 1999, 80: 3289 ~ 3304.
- [6] Kuzio J, Pearson M N, Harwood S H. Virology, 1999, 253: 17 ~ 34.
- [7] Pang Y, Yu J X, Wang L. Virology, 2001, 287: 391 ~ 404.
- [8] Hayakawa T, Ko R, Okano K. Virology, 1999, 262: 277 ~ 297.
- [9] 彭建新. 杆状病毒分子生物学. 武汉: 华中师范大学出版社, 2001. 82 ~ 84.
- [10] 刘德立, 齐义鹏. 病毒学报, 2001, 17: 188 ~ 191.
- [11] Kool M, Voeten J T M, Goldbach R W, et al. J Virology, 1994, 198: 680 ~ 689.
- [12] Kool M, Ahrens C H, Goldbach R W, et al. J Proc Natl Acad Sci USA, 1994, 91 (23): 11212 ~ 11216.
- [13] Lu A, Miller L K. J Virology, 1995, 69 (2): 975 ~ 982.
- [14] Hawinton R E, Arnold K, Ayres M D, et al. J Virology, 1995, 212: 673 ~ 685.
- [15] Reilly D R, Miller L K. J Science, 1989, 245: 1110 ~ 1112.