

# 油桐尺蛾核型多角体病毒广西株全基因组序列

罗 辑<sup>1,2</sup>, 周国英<sup>1\*</sup>, 朱积余<sup>2</sup>

(1. 中南林业科技大学林学院, 湖南长沙 410004; 2. 广西壮族自治区林业科学研究院, 广西 南宁 530002)

**摘要:**油桐尺蛾核型多角体病毒广西株基因组序列由 Sanger 测序的方法得到。拼接后获得了 121 268 bp 序列, GC 含量为 36.76%。以长度不小于 50 aa 的序列认定为功能基因, 共预测到 131 个开放阅读框, 通过 Blast 比对, 其中 87 个能注释到功能基因。对基因组序列进重复序列分析, 共发现 SSR 位点 174 个, 同源重复序列 1 个。在 SSR 位点中, 包含 6 种长度的重复基序, 且以富含 AT 的重复基序为主。同源重复序列两端为一 59 bp 片段或其一部分组成的重复序列, 两端分别重复 12 次和 7 次, 中间则为一段与重复片段无关的 236 bp 片段。与武汉株相比, 广西株病毒在 37 个核心基因中有 4 个存在蛋白序列长度上的差异, 而在其他非核心基因中有 6 个基因仅存在于一个病毒株中, 另有 13 个基因存在蛋白序列长度上的差异。

**关键词:**油桐尺蛾; 核型多角体病毒; 株系; 基因组; 重复序列

中图分类号: S763.3

文献标识码: A

## The Complete Genome of *Buzura suppressaria* Nucleopolyhedrovirus Guangxi Strain

LUO Ji<sup>1,2</sup>, ZHOU Guo-ying<sup>1</sup>, ZHU Ji-yu<sup>2</sup>

(1. College of Forestry, Central South University of Forestry and Technology, Changsha 410004, Hu'nan, China;

2. Guangxi Forestry Science Research Institute, Nanning 530002, Guangxi, China)

**Abstract:** The genome of *Buzura suppressaria* nucleopolyhedrovirus isolate Guangxi was sequenced by Sanger's method. The size of the genome is 121 268 bp with 36.76% G + C content. It contains 131 hypothetical open reading frames (ORFs) that potentially encode at least 50 amino acids respectively. 87 ORFs were annotated by Blast. 174 SSR sites and one homologous repeated sequence (HRS) were found in the genome. The SSR sites contained 6 types of repeats, and most of them were AT-rich sequences. HRS contained 12 repeats at 5' end and 7 repeats at 3' end. The repeat was 59 bp or truncated. Between 5' end and 3' end repeats, there was a 236 bp sequence with no obvious features. Compared with Wuhan strain, 4 genes among 37 core genes were different in protein length and 6 non-core genes were found only in one strain, and another 13 genes were different in protein length.

**Key words:** *Buzura suppressaria*; nucleopolyhedrovirus; strain; genome; repeat sequence

油桐尺蛾 (*Buzura suppressaria*) 是近年来危害桉树的重要灾害性食叶害虫, 其为害率和为害面积均呈快速上升的趋势<sup>[1]</sup>。油桐尺蛾核型多角体病毒 (*Buzura suppressaria* nuclear polyhedrosis viruses, BsNPV) 属于杆状病毒科 (Rhabdoviridae) 核型多角

体病毒属 (*Nucleopolyhedrovirus*), 能专一性的侵染油桐尺蛾, 对其起到生物防治的作用, 是很有潜力的生物农药。

昆虫病毒是专性侵染节肢动物的一类微生物, 包括核型多角体病毒、颗粒体病毒、质型多角体病

收稿日期: 2014-10-20

基金项目: 国家自然科学基金项目(31400549); 中国博士后科学基金资助项目(2014M552533XB); 广西博士后专项资助项目; 广西林科院基金项目(林科 201401 号)

作者简介: 罗 辑(1985—), 男, 湖南长沙人, 博士后, 主要从事森林病虫害研究。

\* 通讯作者: 教授, 博士生导师, E-mail: gyzhou2118@163.com

毒、痘病毒、虹彩病毒等很多类群,已报道的有1 600多种,分属13科21属<sup>[2]</sup>。但研究较为深入的并不多,仅有不到百种,多为模式种或有较大应用价值的种类。而在基因组学方面的研究就更少,在NCBI上有报道的核型多角体病毒仅有41种<sup>[3-4]</sup>。BsNPV湖北武汉株的基因组序列由Zhu等人<sup>[4]</sup>在2014年1月报道,这是BsNPV基因组的首次报道。近期BsNPV广西株的基因组序列已经完成测序,本文在分析其序列的基础上对病毒的结构及重复序列进行了分析,并与武汉株进行了比较,以期寻找不同株系之间的差异。

## 1 材料与方法

### 1.1 供试病毒

油桐尺蛾核型多角体病毒从广西南宁市六里林场采集的油桐尺蛾上分离保存。

### 1.2 方法

1.2.1 病毒的增殖 选取生长状态相同的油桐尺蛾四龄幼虫,单条置于培养皿中饥饿24小时,用喷

有上述保存的油桐尺蛾核型多角体病毒悬液的桉树叶饲喂,待其发病死亡后收集虫尸。

1.2.2 病毒的纯化 将收集的虫尸置于研钵中,加适量PBS(PH7.4)缓冲液研磨。用夹有一层棉花的纱布过滤,4℃差速离心,收集沉淀。沉淀用0.1% SDS洗涤后,用少量PBS重悬。再用30%~65%蔗糖连续梯度离心,收集病毒带,PBS洗涤后4℃保存。

1.2.3 病毒DNA提取 向病毒悬液中加入等体积碱解液(0.1 mol·L<sup>-1</sup> Na<sub>2</sub>CO<sub>3</sub>,0.17 mol·L<sup>-1</sup> NaCl,0.01 mol·L<sup>-1</sup> EDTA,pH 10.8),37℃保温至澄清。加入10% SDS至终浓度为0.5%,加入蛋白酶K,至终浓度0.25 mg·mL<sup>-1</sup>,55℃保温3小时。酚氯仿抽提后溶于TE中,置于-20℃备用。

1.2.4 PCR扩增 在NCBI上搜索所有已报道的BsNPV的序列,在序列两端合成T<sub>m</sub>值在60℃以上的引物(表1),根据Hu等人<sup>[5]</sup>在1998年报道的BsNPV基因组部分序列确定这些基因的相对位置,对基因间序列进行长片段PCR扩增。待测序拼接完成后,根据测序结果设计引物扩增验证已知序列部分。

表1 BsNPV已知序列及本文涉及的引物

序列名称	GeneBank号	基因组中方向	引物名称	引物序列	扩增片段编号*
ph	X70844.1	+	ph-up	CAAATACCGAAGCACAAAACAAAACACTGAT	10
			ph-down	AGACTAAATTTCCAATTTCGTGTCATAAGCCAG	1
p10	AF034410.1	-	p10-up	CAAGCAAAACGCCAATTAGAAATCCAAAC	2
			p10-down	CGCCGTATTACAAAAAAGCCCACTAACAAT	1
40	AF058929.1	-	40-up	AGCGGTCTCGATTTCAGCATCACACTGG	3
			40-down	AAATCATCTGAAACTGGCGGGAAACAT	2
9	DQ231344.1	-	9-up	TTTAGTTGAGATATGTTGGATGGCGAGT	4
			9-down	ACTCGCCATCCCAACATATCTCAACTAAA	3
8	DQ231343.1	+	8-up	CGTTTTCTCGTCACACATATCGTTTTTCT	4
			8-down	TATCGCCGCACATTGCTCAACAACAAA	5
dpol	AF068184.1	+	dpol-up	GAGACCCGCACAATTTACACGACAATTC	5
			dpol-down	GCAAAGGTTACAAAATGGCTCTTAAAAAC	6
iap	AF045936.1	-	iap-up	TACAATCGGGTTCTGGTAAAGGCACACATC	7
			iap-down	TTAAAATGTAGCCCTCAATCACAGTAGGA	6
lef	U61154.1	-	lef-up	GATTTGTTTATTAGACAGTGACGGTTTCCA	8
			lef-down	TACGCCATGTCTGCCCATTTGTTTGGTAT	7
119	AF058928.1	-	119-up	AATTTATCAGTTTGAGATTTCGTATTTCGCC	9
			119-down	CCATTTGTAATCTATTGCTGCCAGGTTAAA	8
egt	U61154.1	-	egt-up	GCCGTCTCACTTTTTATATGTCACCAAAT	10
			egt-down	GCGTTTTACTTTTTACTATTTTTCCGTAGG	9

\*:片段编号相同的为扩增该片段时的上下游引物

PCR体系为病毒DNA 20 ng,引物(10 μmol·L<sup>-1</sup>)各1.6 μL,dNTP(2.5 mmol·L<sup>-1</sup> each)4 μL,Thermo Scientific Long PCR Enzyme Mix 0.5 μL,配套10(Buffer 5 μL,反应体系50 μL。反应条件按说明书中对应预计片段长度的程序设置。1.5%琼脂糖凝胶电泳后切胶回收后送上海立菲生物技术公司测序。

1.2.5 基因组分析 测序后获得的原始序列用DNASTAR软件的seqman工具进行序列拼接,得到BsNPV广西株的全基因组序列。再利用Softberry网站FGENESV0(<http://www.softberry.com/all.htm>)对其功能基因进行预测,将含有起始密码子和终止密码子且氨基酸序列长度不小于50 aa的序列认定

为功能基因。对预测的功能基因用 NCBI 的 Blastp 工具预测基因的功能。利用 SSRHunter 软件对该序列内的 SSR 位点进行检索,检索标准为:含一、二、三、四、五和六核苷酸类型的简单序列重复基序的最小重复数分别为 8、4、3、3、3 和 3 次。本文中主要搜索完全重复型 SSR 位点,循环排列或成反相互补的 SSR 被当作同一种,例如 CAT 重复基序包括 CAT、ATC、TCA、ATG、TGA 和 GAT。

## 2 结果与分析

### 2.1 基因组基本特征

油桐尺蛾核型多角体病毒广西株基因组 (Gen-

Bank Accession No. KM986882) 全长 121 268 bp, GC 含量为 36.76% (图 1)。对基因组序列进行功能基因预测,共预测到 131 个开放阅读框(ORF),基因组中 ORF 区域的 GC 含量为 38.27%,与整体 GC 含量相近。对这些 ORF 进行功能注释后,将注释为多角体蛋白(polyhedrin)的基因标注为 ORF-1,其起始密码子的 A 位点标注为基因组的 1 号位,并将其转录方向定为基因组正链,后续基因顺序编号。将能够注释的基因编号用注释的基因名替换,没有注释到的基因仍使用基因编号。最终获得了 87 个可以注释到的基因,占总基因数的 66.4%。

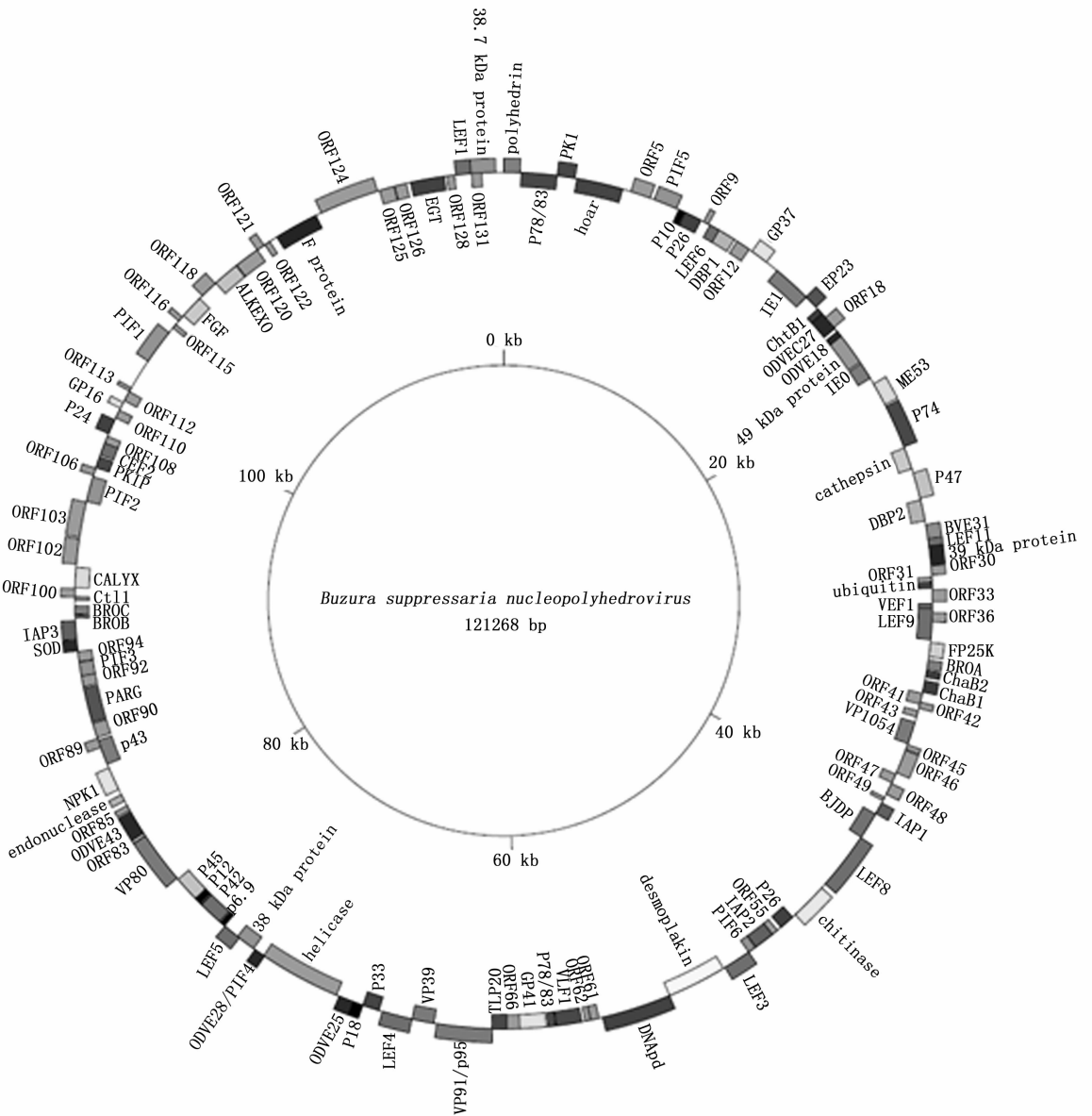


图 1 油桐尺蛾核型多角体病毒基因组物理图谱

广西株基因组中绝大部分基因都是顺序排列在环状 DNA 上,基因间有一定的间距,但其中有 22 处存在重叠的区域。重叠区域可以分为两类,一类是相邻的两个基因位于 DNA 同一条链上,由于阅读框的差异形成的重叠,这样的区域有 10 处;另一类是相邻的两个基因分别位于 DNA 正负两条链上,由于转录方向的差异形成的重叠,这样的区域有 12 处。值得一提的是,在 12 处分别位于两条链上的重叠中,有 2 处是完全重叠,即一条链上的基因完全包含在另一条链上对应位置的基因中。而这两处完全重叠的区域均表现为一个较长的基因是已知的功能基因,另一个较短的基因为功能未知的基因。

## 2.2 重复序列分析

在杆状病毒基因组中存在两种重复序列,一种为简单重复序列 (simple sequence repeats, SSR), 其为 1~6 个核苷酸为重复基序的重复序列;另一种为同源重复序列 (homologous repeated sequences, HRS), 其是由一个或几个长的重复基序连续或间隔或反向排列而成的重复序列。

在油桐尺蛾核型多角体病毒的基因组中,共搜索到符合条件的 SSR 位点 174 个(表 2)。其中包括单核苷酸重复基序 66 个,二核苷酸重复基序 25 个,三核苷酸重复基序 77 个,四核苷酸重复基序 2 个,五核苷酸重复基序 3 个,六核苷酸重复基序 1 个。这些序列中最多的重复基序是 A/T, 然后依次是 ATT/AAT, TGT/ACA, CAT/ATG 和 AT/TA, 这些重复基序占总数的 81.6%。三个及以下核苷酸重复基序占 SSR 位点总数的绝大多数,达到 96.6%;而四个及以上核苷酸重复基序总共仅有 6 个,占总数 3.4%。在所有 SSR 位点中,重复基序的碱基组成存在一定的碱基偏好性,在六类重复基序中均以富含 AT 的重复基序为主。特别是单核苷酸、四核苷酸及六核苷酸基序,均为 A、T 两种核苷酸组成。从重复基序的重复次数上看,这些 SSR 的重复度并不高,单核苷酸的重复次数为 8~10 次,二核苷酸的重复次数为 4~11 次,三核苷酸的重复次数为 3~11 次,四核苷酸的重复次数均为 3 次,五核苷酸的重复次数为 3~4 次,六核苷酸的重复次数为 3 次。

在油桐尺蛾核型多角体病毒基因组中只有一处 HRS 位点,其位于 *orf-113* 和 *pif-1* 基因之间(101 070-102 317),长度 1 248 bp, GC 含量仅有 30.04%, 低于基因组的整体水平。通过对其序列的分析发现,此段序列的两端为一 59 bp 片段或其一部分组

表 2 油桐尺蛾核型多角体病毒 SSR 类型及数量

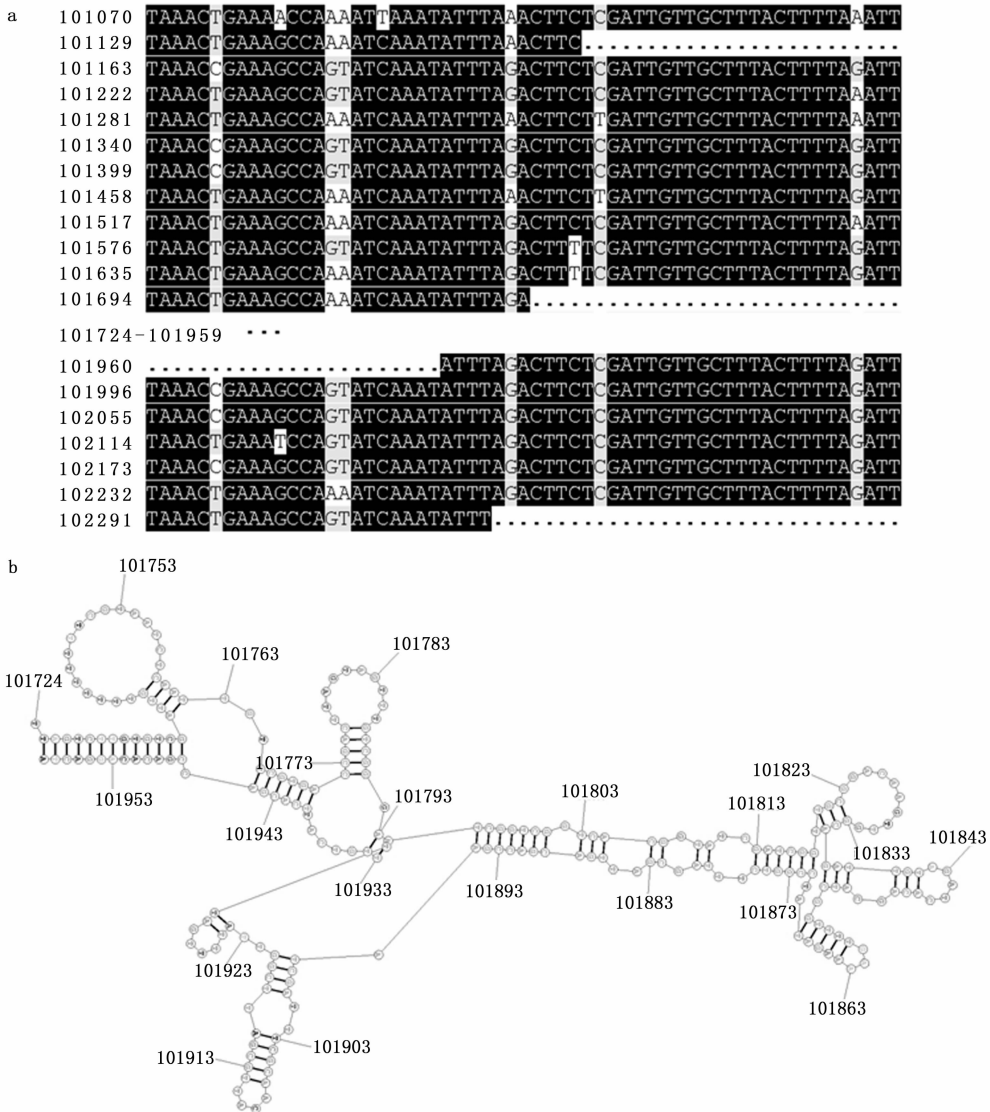
SSR 重复基序	SSR 数量	重复基序的比例
单核苷酸		37.9%
A/T	66	100%
二核苷酸		14.4%
AT/TA	15	60%
AC/GT	9	36%
GC/CG	1	4%
三核苷酸		44.3%
ATT/AAT	28	36.4%
TGT/ACA	17	22.1%
CAT/ATG	16	20.8%
GGC/GCC	7	9.1%
TGC/GCA	4	5.2%
TCG/CGA	3	3.9%
GGT/ACC	2	2.6%
四核苷酸		1.1%
TTTA/TAAA	2	100%
五核苷酸		1.7%
TATTT/AAATA	1	33.3%
AATTC/GAATT	1	33.3%
AATTT/AAATT	1	33.3%
六核苷酸		0.6%
ATATAA/TTATAT	1	100%

成的重复序列,各重复序列顺序相连(图 2a)。5' 端有 12 个重复序列,3' 端有 7 个重复序列。此段序列并不是完全重复序列,在 59 bp 的片段中,有 9 个位点存在一定的变异,但这些位点的变异大多都只有两种碱基的差异。虽然每个差异位点的变异较少,而组合成重复片段后则变异较多,也就是说每个差异位点的变化并不与其他差异位点相关,在具有完整序列的 15 个重复片段中,有 10 种不同的重复序列。

位于 HRS 中部的一段序列(101 724-101 959),并不是由上述重复片段组成,而是由一段与重复片段无关的 236 bp 片段组成。此片段的首尾各有 13 bp 的片段为反向互补序列,且序列中有多处 3~7 bp 的反向互补序列,因此其可能形成一结构较为稳定的发卡结构(图 2b),如形成预测的结构其自由能可达  $-47.1 \text{ kcal} \cdot \text{mol}^{-1}$ 。

## 3 讨论

核型多角体病毒的核酸为双链环状 DNA,大小一般在 80~180 kb 之间。与其他生物相比,其基因组很小,因此基因在基因组中的排列较为紧凑。相对于真核细胞生物存在大量的内含子和非编码 DNA<sup>[6]</sup>,杆状病毒的基因组中仅有少量的非编码区,而绝大部分序列都是基因的编码区<sup>[7-8]</sup>。油桐尺蛾



a: 重复序列的比对 (黑色为同源性大于70%的碱基, 灰色为同源性大于50%的碱基);

b: HRS 中部非重复序列部分二级结构预测

图2 油桐尺蛾核型多角体病毒基因组 HRS 位点结构

核型多角体病毒的基因组也不例外, 无论是用高通量测序的方法获得的武汉株<sup>[4]</sup>, 还是本研究用 Sanger 测序获得的广西株, 都具有相同的特点。

经过前人的研究发现, 杆状病毒虽然种类繁多, 不同种类的基因数量、基因种类、基因分布也不尽相同, 但有 37 个基因是这些基因中的核心基因, 在不同杆状病毒中都存在, 并且起着相同的作用<sup>[9-10]</sup>。这些基因参与病毒生命循环的一些关键环节, 包括病毒经宿主吞服后的侵染过程、与宿主细胞的相互作用、核酸复制与转录、病毒粒子的包装、组装、释放等。与文献中的核心基因比对后发现油桐尺蛾核型多角体病毒中这 37 个基因均存在, 但其在基因组中

的分布并不均匀, 有超过一半的核心基因分布于基因组 40 ~ 81 kb 之间, 其余基因在基因组中基本上均匀分布。由于这些基因是病毒的核心基因, 所以其在不同种类间具有一定的差异, 而在相近种类间相对保守, 因此可以用这些核心基因作为构建进化树的序列, 分析不同杆状病毒间的亲缘关系。通过相同种类不同地理株系之间的序列比较发现, 即便是这些参与重要过程的核心基因, 其不同地理株系间仍存在一定的差异<sup>[11]</sup>。与武汉株相比, 广西株油桐尺蛾核型多角体病毒虽然基因组同源性高达 98.72%, 但其核心基因也存在着一定的差异: 有 4 个核心基因出现了蛋白长度上的差异, 包括 *lef-9*、

*vf-1*、*orf-66*、*lef-5*,其余33个核心基因均没有长度上的差异。在这4个核心基因中,*vf-1*、*orf-66*两个基因分别与武汉株仅相差3 bp和9 bp,差异不大;而*lef-9*和*lef-5*两个基因则差异较大,前者与武汉株相差123 bp,后者则相差204 bp。序列比对后发现*vf-1*、*orf-66*以及*lef-5*三个基因的差异均是由于序列中简单重复序列的插入与缺失造成的,而前两者是分别差异一个和三个三核苷酸基序,*lef-5*则是由于单核苷酸重复基序中相差一个基序,从而导致的移码突变。*lef-9*基因与武汉株相比具有一段123 bp的缺失,而这一段序列并没有类似于发卡、重复等的特殊结构。

除了这些相对保守的核心基因外,杆状病毒的其他基因则表现出较大的差异,这很大程度上归因于病毒的高适应性以及快速的进化,而这样的特点在以鳞翅目幼虫为宿主的杆状病毒中表现得尤为突出<sup>[12-14]</sup>。即便是同一个基因,在感染同属宿主的不同杆状病毒中也可能存在着较大的差异。以多角体蛋白基因为例,在甜菜夜蛾核型多角体病毒(*Spodoptera exigua* NPV, SeNPV)、草地贪夜蛾核型多角体病毒(*Spodoptera frugiperda* NPV, SrNPV)以及海灰翅夜蛾核型多角体病毒(*Spodoptera littoralis* NPV, SlNPV)中就有较大差异<sup>[15-17]</sup>。此三种病毒侵染的宿主同属于灰翅夜蛾属,但据其多角体蛋白基因构建进化树后发现,SlNPV与其他两种病毒并不位于进化树的同一分支上,而且相距较远。差异也不仅仅存在于不同种类间,在同种病毒的不同株系间也会存在差异。对比家蚕核型多角体病毒的6个株系,在预测到的135个基因中有45个基因至少有一个株系存在蛋白长度上的差异,而差异最大的基因相差210 bp<sup>[11]</sup>。油桐尺蛾核型多角体病毒广西株与武汉株之间也存在这样的差异,其中有6个基因仅在一个株系中存在,另外,存在蛋白长度差异的基因达13个,差异最大的达360 bp。这些差异的存在一方面可能是由于病毒宿主的地理差异,另一方面可能是由于病毒与宿主间协同进化过程中的相互作用导致的。

## 参考文献:

[1] 廖旺姣,盛双,常明山,等. 油桐尺蠖幼虫对3种桉树无性系的取食选择性初探[J]. 广西林业科学, 2011, 40(4): 328-329, 335.

[2] Pringle C R. Virus taxonomy-1999. The universal system of virus taxonomy, updated to include the new proposals ratified by the Inter-

national Committee on Taxonomy of Viruses during 1998 [J]. Archives of Virology, 1999, 144(2): 421-429.

[3] 罗辑,周国英,朱积余. 核型多角体病毒基因组学研究进展[J]. 林业科学研究, 2014, 27(4): 551-556.

[4] Zhu Z, Yin F, Liu X, et al. Genome Sequence and Analysis of Buzura suppressaria Nucleopolyhedrovirus: A Group II Alphabaculovirus [J]. Plos One, 2014, 9(1): e86450.

[5] Hu Z H, Arif B M, Jin F, et al. Distinct gene arrangement in the Buzura suppressaria single-nucleocapsid nucleopolyhedrovirus genome [J]. Journal of General Virology, 1998, 79 (Pt 11): 2841-2851.

[6] Ohno S. So much "junk" DNA in our genome [J]. Brookhaven symposia in biology, 1972, 23: 366-370.

[7] Ayres M D, Howard S C, Kuzio J, et al. The complete DNA sequence of Autographa californica nuclear polyhedrosis virus [J]. Virology, 1994, 202(2): 586-605.

[8] Gomi S, Majima K, Maeda S. Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus [J]. The Journal of general virology, 1999, 80(Pt 5): 1323-1337.

[9] Garavaglia M J, Miele S A B, Iserte J A, et al. The *ac53*, *ac78*, *ac101*, and *ac103* Genes Are Newly Discovered Core Genes in the Family Baculoviridae [J]. Journal of Virology, 2012, 86(22): 12069-12079.

[10] Belenmiele S A, Garavaglia M J, Belaich M N, et al. Baculovirus: Molecular Insights on Their Diversity and Conservation [J]. International Journal of Evolutionary Biology, 2011, 2011: 379424.

[11] Xu Y P, Cheng R L, Xi Y, et al. Genomic diversity of Bombyx mori nucleopolyhedrovirus strains [J]. Genomics, 2013, 102(2013): 63-71.

[12] Gettig R R, McCarthy W J. Genotypic variation among wild isolates of Heliothis spp nuclear polyhedrosis viruses from different geographical regions [J]. Virology, 1982, 117(1): 245-252.

[13] Graham R I, Tyne W I, Possee R D, et al. Genetically variable nucleopolyhedroviruses isolated from spatially separate populations of the winter moth Operophtera brumata (Lepidoptera: Geometridae) in Orkney [J]. Journal of Invertebrate Pathology, 2004, 87(1): 29-38.

[14] Cory J S, Green B M, Paul R K, et al. Genotypic and phenotypic diversity of a baculovirus population within an individual insect host [J]. Journal of Invertebrate Pathology, 2005, 89(2): 101-111.

[15] Ijkel W, Strien E V, Heldens J, et al. Sequence and organization of the Spodoptera exigua multicapsid nucleopolyhedrovirus genome [J]. Journal of General Virology, 1999, 80(Pt 2): 3289-3304.

[16] Harrison R L, Puttler B, Popham H J. Genomic sequence analysis of a fast-killing isolate of Spodoptera frugiperda multiple nucleopolyhedrovirus [J]. Journal of General Virology, 2008, 89(Pt 3): 775-790.

[17] Breitenbach J E, El-Sheikh E-S A, Harrison R L, et al. Determination and analysis of the genome sequence of Spodoptera littoralis multiple nucleopolyhedrovirus [J]. Virus Research, 2013, 171(1): 194-208.