

45-51
斜纹夜蛾核型多角体病毒 *p74* 基因的克隆和序列分析*

杨洁**, 龙繁新, 吴文言, 王珣章

(中山大学昆虫所, 生物防治国家重点实验室, 广州 510275)

摘要: 用 AcNPV *p74* 基因 3' 端的 1.4 kb 片段, 通过 Southern blotting 将 SpltNPV 的 *p74* 基因定位于 XhoI(4.9 kb)、EcoRI(4.4 kb)、BamHI(3.0 kb) 片段上。进一步用 Exo III 和构建亚克隆测序法, 对长 2545 bp 的 *p74* 基因进行了序列分析, 其中 1974 bp 的编码区编码 658 个氨基酸, 蛋白分子量约 75.87 kD, 其编码蛋白与 AcMNPV 和 CfMNPV *p74* 蛋白的氨基酸序列同源性分别为 64% 和 63%。

关键词: 斜纹夜蛾核型多角体病毒 (spltNPV); *p74* 基因; 基因克隆; 序列测定

中图分类号: Q7.756 文献标识码: A 文章编号: 1003-3125(2000)01-0045-07

昆虫杆状病毒作为一种无公害生物杀虫剂越来越得到广泛的应用, 但是昆虫杆状病毒的宿主专一性较强, 杀虫谱较窄, 已成为限制其更广泛应用的一个重要因素。如何扩大其宿主范围, 提高毒力, 是应用杆状病毒防治害虫所面临的重大课题。*p74* 基因作为与昆虫杆状病毒感染力有关的基因, 其基因产物与杆状病毒在感染昆虫幼虫时与中肠的特异性吸附有关^[1,2], 这种吸附可能具有种的专一性, 因而 *p74* 基因可能是决定杆状病毒宿主专一性的一个重要因素。所以研究不同杆状病毒之间 *p74* 基因的结构及功能的异同, 具有重要的意义。斜纹夜蛾 (*Spodoptera litura*) 是危害多种农作物的重要害虫, 主要分布于中国南部、东南亚地区和日本西南部。本研究在斜纹夜蛾核型多角体病毒 (*Spodoptera litura* nuclear polyhedrosis virus, SpltNPV) *p74* 基因定位的基础上进行序列分析, 并将所测序列与苜蓿丫纹夜蛾核型多角体病毒 (*Autographa californica* multinucleocapsid nuclear polyhedrosis virus, AcMNPV) 和柞色卷蛾核型多角体病毒 (*Choristoneura fumiferana* multinucleocapsid nuclear polyhedrosis virus, CfMNPV)^[3] 的 *p74* 基因进行了核苷酸同源性和编码蛋白的氨基酸序列同源性比较。

1 材料与方法

1.1 病毒、菌株和质粒

斜纹夜蛾核型多角体病毒中山大学分离株 (Zhongshan University isolate, ZSU)^[4] 由本室保存, 苜蓿丫纹夜蛾核型多角体病毒由美国加州大学 Federici B. A. 教授提供, 载体质粒用 pBluescript M13(-) 和 pUC18, 受

收稿日期: 1998-08-28, 修回日期: 1998-11-06

* 由国家自然科学基金重点项目 (39730030) 资助

** 现工作单位: 第一军医大学南方医院传染科, 广州 510515

作者简介: 杨洁 (1970-), 男, 湖北广水人。助理研究员, 硕士, 1992 年武汉大学病毒系本科毕业。研究方向生物工程。曾获军队科技进步二等奖。

体菌株为 *E. coli* DH5 α 。

1.2 病毒基因组 DNA 的制备

参考 Summers M. D. 和 Smith G. E. 的方法^[5], 从病毒多角体中提取病毒基因组 DNA。

1.3 核酸杂交探针的制备

由于 AcMNPV 的 EcoRI-P 片段含有大部分 *p74* 基因(2.0 kb)^[6], 遂用 pBluscript M13(-)克隆该片段, 并经 EcoRI 和 HindIII 双酶切回收含 *p74* 基因 3'端的 1.4 kb 片段。按 Boeringer Mannheim Biochemica 公司的 'DIG DNA Labeling and Detection kit'说明书标记探针。

1.4 Southern blotting 及杂交片段的克隆

用上述标记的探针在宽松条件下(5 \times SSC, 56 $^{\circ}$ C, 无甲酰胺)与 SpltNPV 基因组 DNA 的 EcoRI、BamHI 和 XhoI 酶切片段杂交, 按试剂盒说明进行显色。然后用 pBluscript M13(-)和 pUC18 克隆阳性片段。具体方法参考文献[7]。克隆有杂交片段的质粒分别命名为 pBluep74-E、pBlup74-B、pBlueP74-X。

1.5 DNA 序列分析及同源性比较

1.5.1 构建双链连续测序模板

参照 Pharmacia 公司 Double-stranded Nested Deletion Kit 说明, 使用 BstXI 和 SmaI 对质粒 pBluep74-E 双酶切, 然后用 ExoIII 消化(75 mmol/L NaCl, 30 $^{\circ}$ C)每 3 min 取样一次, S1 核酸酶去除单链 DNA, 终止反应, T4 连接酶自连, 转化筛选。

1.5.2 双链 DNA 序列分析及同源性比较

参照 Pharmacia 公司双链 T7 DNA 序列分析 Kit 推荐的方法, 进行序列分析, 并应用计算机软件 DNASIS、PROSIS, 将所测序列与其它两种杆状病毒的 *p74* 基因及其推导的氨基酸序列进行同源性比较。

2 结果与讨论

2.1 SpltNPV *p74* 基因的定位、克隆及酶切分析

以地高辛标记的 AcMNPV *p74* 基因为探针, 与 SpltNPV 基因组酶切片段进行杂交, 结果表明, 阳性带分别为 SpltNPV 基因组 DNA EcoRI 4.4 kb、XhoI 4.9 kb、BamHI 3.0 kb 片段。将这三个片段分别克隆后用内切酶进行酶切分析, 以确定其相对位置, 结果显示, 三个片段都含有 0.95 kb 的 BamHI 和 EcoRI 片段的共同部分, 提示 AcMNPV 的 *p74* 基因探针与这三个片段都是特异性杂交。我们先测出这 0.95 kb 片段的序列, 以确定 *p74* 基因的精确位置及编码方向。将质粒 pBluep74-E 用 ExoIII 构建双链连续测序模板, 利用 T7 primer 从一端进行连续测序。

2.2 SpltNPV *p74* 基因的序列测定

根据上述结果对 EcoRI 4.4 kb 片段进行双链 DNA 序列分析, 所测序列长 2 545 bp。对该序列进行读码框分析, 结果推测 SpltNPV *p74* 基因编码区为 222~2 195 nt, 编码 658 个氨基酸(图 1), 分子量为 75.87 kD。比 AcMNPV 和 CfMNPV 的 P74 蛋白多 12 个氨基酸。分析结果显示, SpltNPV 的 *p74* 基因符合 Kozak 规则, 即 ATTATGA。该基因含有杆状病毒典型晚期启动子特征序列, 即 DTAAG(D 代表 A、T 或 G)序列^[8~10]。和 AcMNPV *p74* 基因一样^[2], SpltNPV 的 TTAAG 距起始密码子很近, 只有 9 个核苷酸, 前者为 11 个核苷酸。此外, SpltNPV 的 *p74* 基因又具有早期启动子元件——TATA 和早期启动子 5' 加帽位点——CAGT^[10, 11]。在 SpltNPV 的 *p74* 基因的 3' 端部分, 有典型的 poly(A) 加尾信号 AATAAA。

2.3 SpltNPV 与 AcMNPV 和 CfMNPV 的 *p74* 基因及其编码蛋白的同源性比较

应用计算机软件 PROSIS, 将 SpltNPV *p74* 基因分别与 AcMNPV、CfMNPV、OpMNPV 和家蚕核型多角体病毒(*Bombyx mori* nuclear polyhedrosis virus, BmNPV) (GenBank 133180) 的

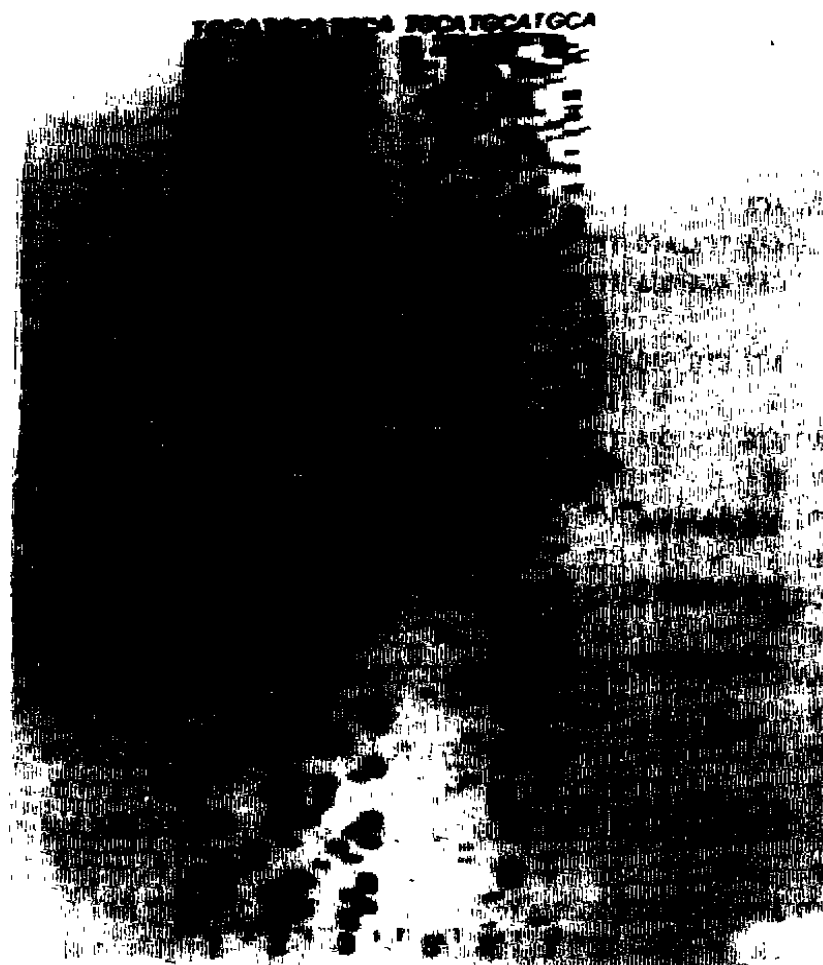


图 1a SpltNPV *p74* 基因的部分测序照片

Fig. 1a Sequencing of partial SpltNPV *p74* gene

1, 2, and 3: sequencing for the first loading;

1', 2', and 3': sequencing for the second loading.

p74 基因及其推导的编码蛋白进行了同源性比较,结果表明,SpltNPV *p74* 基因推导的编码蛋白与 AcMNPV、CfMNPV、OpMNPV、BmNPV *p74* 基因编码蛋白同源性分别为 64%、63%、64% 和 63%。N-端及中间区域的氨基酸同源性较高,有些区域连续几十个氨基酸的同源性在 90% 以上,这些区域的高保守性很可能与其结构和功能有密切关系。值得注意的是已发表的 *p74* 基因所编码蛋白的氨基酸在它们靠 N-端和中间部分都有很高的同源性,在其 C-端则同源性较低^[2,3]。SpltNPV *p74* 基因的编码蛋白也有这种现象,鉴于 *p74* 基因产物很可能与昆虫杆状病毒在感染昆虫幼虫时对中肠的吸附有关^[2,5,8],这种吸附可能具有种的特异性,因而可以设想 *p74* 基因的这些特异区域可能在某种程度上决定不同杆状病毒对昆虫中肠吸附的宿主专一性。

SEQUENCE REGION : 1 - 2545
 TRANSLATION REGION : 222 - 2195

1	TA AGA AGA AGA AGC TAA TAA ATT GTG TAA AAA TGT AAT TTT AGA AAC	47
48	GTC GCC GAT CCA AAT TTT TGT GAC ATA TTC CAT TTA TGT <u>ATA AAT GGA</u>	95
96	AAA TTA ACA ATT ATC ATA TGT TTG CCC AGT TGG TGA AGC <u>TTA TGA TGA</u>	143
144	GGA GCA AAA TCA ATG TTT ACC ATT GAA AAT TGT CGA TTG TGG TGA TAA	191
192	AAG TTT GAT ACA AGA ATT AAG TCT CTA ATT ATG AGC GCG CAC GTA AAT	239
1		
	M S A H V N	6
240	TCT CCG ACT CTG CAG GAT CTG CTA AAC GCC CAA AAG TAT GCG TAC CAT	287
7	S P T L Q D L L N A Q K Y A Y H	22
288	CAG TAC ATG TTG CGA TTC ATT CCC CGA TGG CGC CGT CGC TTT CCT CAC	335
23	Q Y M L R F I P R W R R R F P H	38
336	ATA TTG ATC AAT TAC ACG ATC CCG CCG GCG GGC AAC GAA GAC TTT TAC	383
39	I L I N Y T I R P A G N E D F Y	54
384	GTT CCG AGA AAA ATC AGA GAC GGC GCC GTG TTC GTC GAC ATA AAG TTT	431
55	V P R K I R D G A V F V D I K F	70
432	AGC AAA CCG GGA TGC GAG AGT ATG TCC TGC TAT CCG TTT CCG GAG ACG	479
71	S K R G C E S M S C Y P F R E T	86
480	GGT CCC ATT TCT AGA AAC ACG CCG TTT GGC TAT ACA CAA ACC GCG GAC	527
87	G P I S R N T P F G Y T Q T A D	102
528	ACG CCG CAG GCG TAC GCG CAG CCG GCA TGT TAC CAT CTC GAT CCG TCG	575
103	T P Q A Y A Q P A C Y H L D R S	118
576	GCG GCG ATG CCG GAG GGG GAA GAG AAC AGG GTG CAA TCG GTC GAG TTG	623
119	A A M R E G E E N R V Q S Y E L	134
624	CGG TAC ACG GAC AAC GAA CCG TGC ATC ATG ATG GAC ACG TAC TCT AAA	671
135	R Y T D N E R C I M M D T Y S K	150
672	ATG TAT CTA AAT TCC CCG TAC CTA CCG ACC GAC GAG CAT CTG ATC AAA	719
151	M Y L N S P Y L R T D E H L I K	166
720	GGT ATC GAC GAC GTG CCC GGT TTC AAT GTC GAA TAC GGA TCG GAT CCT	767
167	G I D D V P G F N V E Y G S D P	182
768	CAT TTT CCC GAA AAG TTT AAA GGC ACG TTT AAC GAA GCG TAT TGT CCG	815
183	H F P E K F K G T F N E A Y C R	198
816	CGA TTC GGT CCG GAG ATG ATG AAC GGA ACA TGT CTG ATC CGA TGG TGG	863
199	R F G R E M M N G T C L I R W W	214

864	GAA AGC ATG ATC GGT TTC GTA CTC GGC GAC ACG ATT TAC ATC ACG CTT	911
215	E S M I G F V L G D T I Y I T L	230
912	AAG TTG CTC GTC AAC AAT GTT TTT AGC GAA CTA CGC AAC TTT GAC TAT	959
231	K L L Y N N V F S E L R N F D Y	246
960	AGA CGG CCG TCC GAA ATG CTG CCA CCG CCG CCC CAA GTC GAC TCG GCT	1007
247	R R P S E M L P P P P Q V D S A	262
1008	CGC GTT CTA GAC GAA TGG AGA AAC GTG OGC GAT CCG GCC GTA GAC GTT	1055
263	R V L D E W R N V R D P A V D V	278
1056	GAT TTT GAG AAT AAA TTT TCA ACT TTC AAG TCT ATA TCC GAA CTG GAC	1103
279	D F E N K F S T F K S I S E L D	294
1104	ATT ACG CAA AAC GTT CGA CTC GTG TAC AAA GCC GAA GTG GGC TAC ACG	1151
295	I T Q N V R L V Y K A E V G Y T	310
1152	CGC GAA CAG TTG CCC GCG TCA AAG TTT GCG TCC GCG TAT CGA ATC GGC	1199
311	R E Q L P A S K F A S A Y R I G	326
1200	GAC AAG ACC GGT GAT CGA AAA AGT TTT AAT CGC GAC TAC GAC GAC GAC	1247
327	D K T G D R K S F N R D Y D D D	342
1248	GAG TTG GAC GCA ATC ATT GCG AGT TTC TTC GAA GAC CAC GCC ATG TTA	1295
343	E L D A I I A S F F E D H A M L	358
1296	TTC GGC ATC TTT ACC GAT ATG GGA TTC GAA GCG CTA ATG AGC AAT TTC	1343
359	F G I F T D M G F E A L M S N F	374
1344	AAA TCA ATG TTG AAA AAA ATC AAC ACG ACA CTT TTG CCA AAA TTG CCG	1391
375	K S M L K K I N T T L L P K L R	390
1392	GCT ATA CTA CTG TCG ACG ACG AGC CCG GTT AGC GCC CGA TTT TTG TCG	1439
391	A I L L S T T S R V S A R F L S	406
1440	TAC GTT TAC AAG GGA GCC GTG ATC AAC ATG ATG TCT CGT CTC GCC ATC	1487
407	Y V Y K G A V I N M M S R L A I	422
1488	AAG ACG GTG ACG GCT ATG GCG AAG GCA CTG ACT CCG ATG GCA ATT CAG	1535
423	K T V T A M A K A L T R M A I Q	438
1536	GCG TCG TCC GTG ATC GGT ATA GTA TTG ATC GCG ATG ACT CTT GTC GAT	1583
439	A S S V I G I V L I A M T L V D	454
1584	CTA CTG CTA ATG TTC TGG GAT CCG TTC GGT TAC AAC AAC ATG TTT CCG	1631
455	L L L M F W D P F G Y N N M F P	470
1632	CGA GAA TTT CCC GAC GAT CTG GTC GAT TCG TTT CTG ATC TCG TAC TAT	1679
471	R E F P D D L V D S F L I S Y Y	486

1680	GAA TCC ATG GGC GAA GGC AGC AGA GAC ATG ATA GAG TTT GAA CCG ATG	1727
487	E S M G E G S R D M I E F E P M	502
1728	TTT TTT GGC GAC ATG GTC GAA GAC ATG GAC GAA TAC CAA CTC GAC TCG	1775
503	F F G D M V E D M D E Y Q L D S	518
1776	ATG CTC TAC ATA TGG GAA TAC TTG AAC GGT TTA GAA GTC AAT TCG GAC	1823
519	M L Y I W E Y L N G L E V N S D	534
1824	GGA CAA CTG TTG CTC TGG GAA GAG GAC GAG CCC GTG TCC GAT TTT GAC	1871
535	G Q L L L W E E D E P V S D F D	550
1872	GAG ATC ACG CTC GTG GGC GCG GCT CTC GCT TCG AGC GGC ATG TAC ACG	1919
551	E I T L V G A A L A S S A M Y T	566
1920	CGC GCC GAG TTT CAC GCG TAC ACC GAA CGA CAC AAT CCG ATC ATG GCC	1967
567	R A E F H A Y T E R H N R I M A	582
1968	AAC TCT GAA CGA AAT TGC ACA CCG CTA GTC GAT CAG ATT CTA ACG TTT	2015
583	N S E R N C T P L V D Q I L T F	598
2016	CTG CTT TTG GTA TGC GCT GGA ATC GCG TGG ATC TCC GGC GGC GGT CCC	2063
599	L L L V C A G I A W I S G G G P	614
2064	GTC ATG TCC AGT CTC GTC ATA ATC TTT ATA CTT ATG ATA GTT TTC AAA	2111
615	Y M S S L V I I F I L M I V F K	630
2112	TTG ATC AAT GCT CCG TTC ATG TAT TTT GTG GGA ATG CAA CCG TTC GCA	2159
631	L I N A P F M Y F V G M Q R F A	646
2160	GCC GGT GAT CCG CTG CTC TGG TAC ATA AAT TAT TAA AAA ACA CAA TTT	2207
647	A G D P L L W Y I N Y *	658
2208	TTT TTC AAA AGT TTA TTA TTC ATT AGC <u>AAA TAA AAA</u> TTT TTT GTA CAA	2255
2256	AAT CTG GCA CAA CAA TAT TGT GAT CGA TCG CGT TTC GGT ACG TAA TGA	2303
2304	CAT CCG CGT CCG CCG CGA GCA TGA CCG CTC CCC GTT TAA CCG TGC AGT	2351
2352	GTT GGT ATA AAT TTA TTT TGT CGT CCG CCG CAC CCG GAC AAT CAA TGT	2399
2400	TTG CAA GCC TGT CCG GCA ACG TGA ACA AAA CCG CAT ACG GTA GAG TCG	2447
2448	ATC TAG GAT GAG AGT AAA AGA AGT AAC ACC GCC GAT CTC GCT CGT CTC	2495
2496	TCT CTC CTG ATC ACC ATC GAG TCG GGC ACA AAC TCG TGC CTG ATG AAT	2543
2544	TC	

图 1b SpltNPV *p74* 基因的核苷酸序列及其编码的氨基酸序列, 下划线的部分为可能的转录起始和终止信号。

Fig. 1b Nucleotide sequence of SpltNPV *p74* gene and amino acid sequence it coded, the underlined parts show the TATA Box, the possible initiation site and the termination site.

参 考 文 献

- [1] Leisy D J, Rohrmann G F, Nesson M *et al.* Nucleotide sequence and transcriptional mapping of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus p10 gene [J]. Virology, 1986, 153:157~167
- [2] Kuzio J, Jaques R, Faulkner P. Identification of *p74*, a gene essential for virulence of baculovirus occlusion bodies [J]. Vi-

- rology, 1989, 173:759-763
- [3] Hill J E, Kuzio J, Wilson J A *et al.* Nucleotide sequence of the *p74* gene of a baculovirus pathogenic to the spruce budworm, *Choristoneura fumiferana* multicapsid nuclear polyhedrosis virus [J]. *Bioch & Biophys Acta*, 1993, 1172:187-189
- [4] 中山大学昆虫专业昆虫病毒研究组. 斜纹夜蛾核型多角体病毒的初步研究[J]. 中山大学学报(自然科学版), 1976, 3: 92-97
- [5] Summers M D Smith G E. A manual of methods for baculovirus vectors and insect cell culture procedures [M]. 1987.
- [6] Kool M, Valk J M. The structural and functional organization of the *Autographa californica* polyhedrosis virus genome [J]. *Arch Virol*, 1993, 130:1-16
- [7] Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning, A Laboratory Manual* [M], 2nd ed. Cold Spring Harbor Laboratory Press, 1989.
- [8] Wilsn M E, Maimprize T H, Fiesen P D *et al.* Location, translation and sequence of a baculovirus gene encoding a small arginine-rich polypeptide [J]. *J Virol*, 1987, 61:661-666
- [9] Thiem S M, Miller L K. Identification, sequence, and transcriptional mapping of the major capsid protein gene of the baculovirus *Autographa californica* nuclear polyhedrosis virus [J]. *J Virol*, 1989, 63:2008-2018
- [10] Ahrens C H, Russell R L Q, Funk C J *et al.* The Sequence of *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome [J]. *Virol*, 1997, 229:381-399
- [11] Ayres M D, Howard S C, Kuzio J *et al.* The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus [J]. *Virol*, 1994, 202:586-605

Sequencing of the *p74* Gene of *Spodoptera litura* Multinucleocapsid Polyhedrosis Virus

YANG Jie, LONG Qing-xin, WU Wen-yan, WANG Xun-zhang

(Institute of Entomology and State Key Laboratory for Biological Control,
Zhongshan University, Guangzhou 510275, China)

Abstract: The *p74* gene of SpltNPV was located at EcoRI-4.4 kb, XhoI-4.9 kb and BamHI-3.0 kb fragments by Southern blotting with AcMNPV-derived probe. After cloning of the three fragments, restriction enzyme analysis was used to construct their restriction map. A region of 2 545 bp including *p74* gene of SpltNPV was sequenced. The open reading frame of SpltNPV *p74* gene was decided to be 1 974 bp long, which codes 658 amino acids. Comparison of the homology with other baculovirus *p74* genes shows that SpltNPV *p74* is 64% and 63% identical to that of AcMNPV and CfMNPV at the amino acid level.

Key words: SpltNPV; *p74* gene; Gene cloning; Sequencing