Virology 287, 391–404 (2001)
doi:10.1006/viro.2001.1056, available online at http://www.idealibrary.com on IDEAL®

Sequence Analysis of the *Spodoptera litura* Multicapsid Nucleopolyhedrovirus Genome

Yi Pang,*[,1] Jianxiu Yu,* Lihua Wang,* Xiaohui Hu,* Weidong Bao,† Gang Li,† Chong Chen,† Hua Han,† Songnian Hu,† and Huanming Yang†

*State Key Laboratory for Biocontrol and Institute of Entomology, Zhongshan University, Guangzhou, 510275, China; and †Beijing Genomics Institute, Beijing Airport Industrial Zone, Beijing, 101300, China

Received March 12, 2001; returned to author for revision April 27, 2001; accepted June 14, 2001; published online August 6, 2001

The complete *Spodoptera litura* multicapsid nucleopolyhedrovirus (SpltMNPV) genome contained 139,342 bp with a G+C content of 42.7%, and 141 putative open reading frames (ORFs) or genes of 150 nucleotides or greater that showed minimal overlap. Ninety-six ORFs had homologues in *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV), 16 had homologues in other baculoviruses, and 29 were unique to SpltMNPV. The homologues of ubiquitin and *gp37* are fused in SpltMNPV. The genome lacked a homologue of the major budded virus glycoprotein gene *gp64*, but it contained a homologue of ORF130 of *Lymantria dispar* multicapsid nucleopolyhedrovirus (LdMNPV). There were two homologues of AcMNPV ORF2 (*bro* gene), and a DnaJ protein gene (SpltORF39) in which the N-terminus showed homologies with the J domain of DnaJ family proteins. Seventeen homologous regions (*hrs*) were identified, each containing 2–29 palindromic repeats, with an average length of 534 bp and base content (G+C%) of 33.0.

Key Words: Spodoptera litura; nucleopolyhedrovirus; genome; sequence.

INTRODUCTION

Baculoviruses are a diverse family and the members of Baculoviridae are characterized by rod-shaped, enveloped virions containing a single molecule of circular supercoiled dsDNA, 90–160 kb in size. The family is subdivided into two genera, *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV), based on their distinct occlusion bodies (Murphy *et al.,* 1995). Baculoviruses are specific pathogens for invertebrates, especially insects of the order Lepidoptera. *Spodoptera litura* multicapsid nucleopolyhedrovirus (SpltMNPV) is highly specific and infects only a single host, the cotton leaf worm (Pang, 1994). This insect is an economically important polyphagous pest in China, India, and Japan, causing considerable economic loss to many vegetable and field crops. SpltMNPV has been successfully applied in large scale as a commercial biological insecticide against the cotton leaf worm in China (Chen *et al.,* 1998a; Pang, 1998). Despite the mass production and application of this virus in the field, genetic information, including the molecular mechanism of its infection and host-specificity, has not previously been available.

In an early investigation, the SpltMNPV genome was determined to have a size of about 136.6 kb (Wei *et al.,* 1999). The sequence of a number of SpltMNPV genes, including *polyhedrin* gene (Wei *et al.,* 1999), *egt* gene (Yan *et al.,* 1999), *odv-e66* gene (Zheng *et al.,* 2000), *p10* gene (Wei *et al.,* 1998a), *pk* gene (Wei *et al.,* 1998b), *p74* gene (Chen *et al.,* 1998b), *chitinase* gene (Hu *et al.,* 2000), and *p49* gene (Li *et al.,* 2000) have been elucidated and characterized. In this report, we describe the complete sequence and organization of SpltMNPV and compare it to sequence data from other nucleopolyhedroviruses (Ayres *et al.,* 1994; Gomi *et al.,* 1999; Ahrens *et al.,* 1997; Kuzio *et al.,* 1999; Ijkel *et al.,* 1999).

RESULTS AND DISCUSSION

General characteristics of the SpltMNPV genome

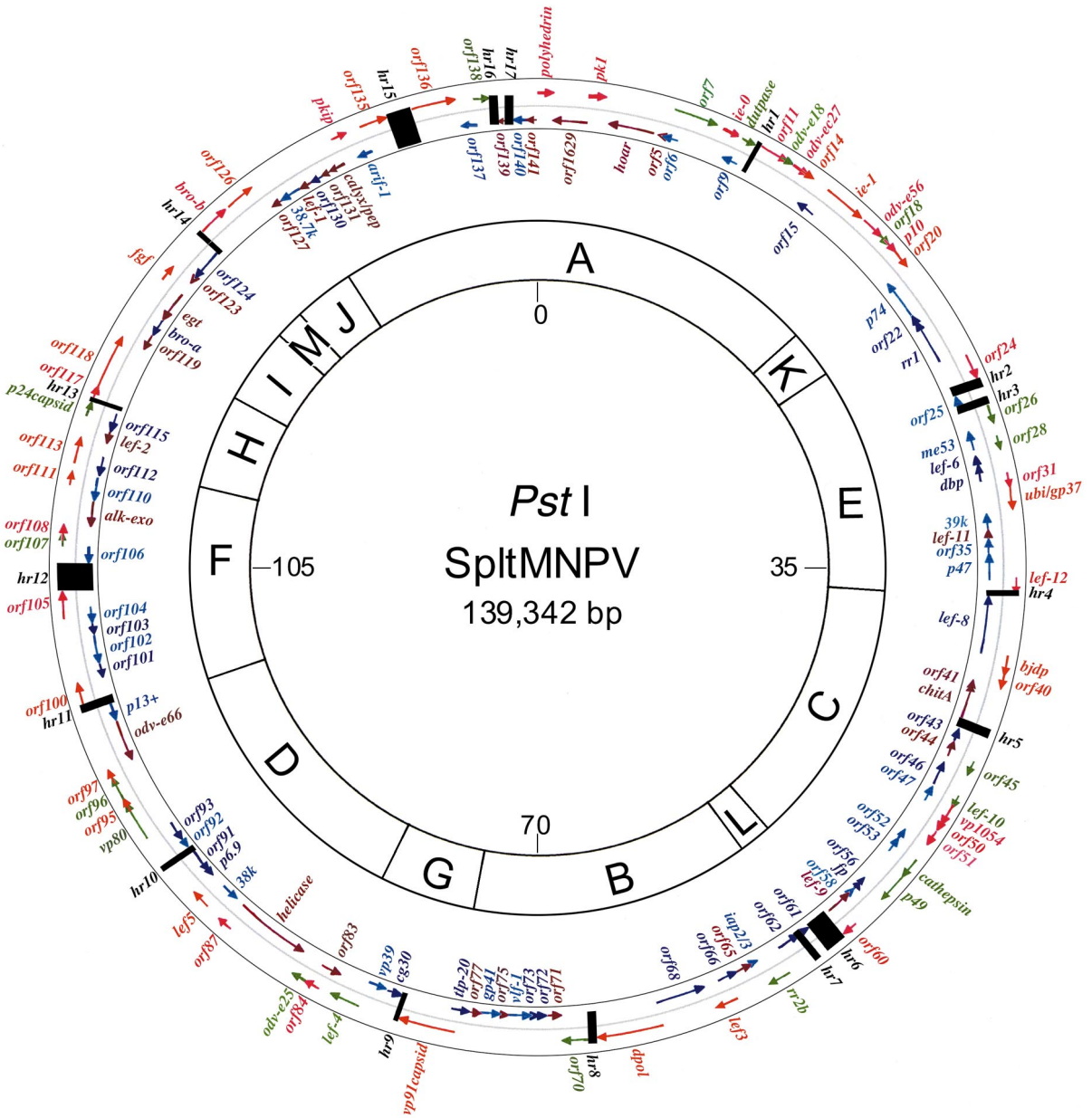
The SpltMNPV genome consisted of 139,342 bp, which was similar to that of *Spodoptera exigua* nucleopolyhedrovirus (SeMNPV) (135.6 kb), AcMNPV (133.9 kb), and *Orgyia pseudotsugata* multicapsid nucleopolyhedrovirus (OpMNPV) (132.0 kb). SpltMNPV genome had a G+C content of 42.7%, which was similar to that of SeMNPV (44%), AcMNPV (41%), and *Bombyx mori* nucleopolyhedrovirus (BmNPV) (40%), but significantly lower than that of *Lymantria dispar* nucleopolyhedrovirus (LdMNPV) (58%) and OpMNPV (55%). Using computer analysis, 141 open reading frames (ORFs) were selected for further analysis in detail. The location, orientation, and size of homologous regions (*hrs*) are summarized in Fig. 1 and detailed in Table 1.

Ubiquitin, *gp37*, and ORF fusions

Sequence analysis of SpltMNPV genome revealed that homologues of two genes found in other baculoviruses, ubiquitin and *gp37*, were fused into a single ORF (SpltORF32). Blast searches showed that the N-terminal

[1] To whom correspondence and reprint requests should be addressed. Fax: +86-20-84037472. E-mail: LS12@zsu.edu.cn.

0042-6822/01 $35.00



1

(14)	DYYTVLGLKPTATREQIRKKFLRLTRISHPD--KAPLTSEA---FVFLQRAYSVIGLEESVRAAYDQF	SpltMNPV
(8)	DYYEILGVPRNASQEEIKKAYRRLVRKYHPDICKKPECEEK---FKEINEAYQVLSDPEK-RKLYDMY	*Aquifex aeolicus*
(4)	DYYEILGVSKNATKEEIKKAYRKLSKKYHPDVNKEPDAAEK---FKEIKEAYEVLS-DDQKRAHYDQF	*Bacillus stearothermophilus*
(110)	DYYEILGVSRGASDEDLKKAYRRLALKFHPDKNHAPGATEA---FKAIGTAYAVLSNPEK-RKQYDQF	*Homo sapiens*
(111)	DYYEILGVSRSASDEDLKKAYRKLALKFHPDKNHAPGATEA---FKAIGTAYAVLSNPEK-RKQYDQF	*Mus musculus*
(2)	DYYTILGVAKTATPEEIKKAYRKLAVKYHPDKNPGD--AEAERRFKEVSEAYEVLG-DAQKRESYDRY	*Chlamydia muridarum*
(9)	DYYELLGINEDAQDQEIHRAWRKTSLKYHPDKNPNDPK-AAEKFHMLQL-AYNALI-DVQLRKAYDSE	*Schizosaccharomyces pombe*
(17)	DYYEVLGLQKGASEDEIKRAYKRLASKHHPDKNQGSKEAEE--KFKEINEAYEVLG-DDQKRAAYDQY	*Haemophilus influenzae Rd*
(5)	DYYEVLGVNRDASDEEIKKSYRKLAMKYHPDRNPDNPKAEE--SFKEAKEAYEVLS-DEQKRAAYDQY	*Methylovorus sp. SS1*
(4)	DYYNVLNVNPSATEDDLKKSYRRLAMKWHPDKNPTSIKQEAEAKFKQISEAYDVLS-DPNKRQIYDQY	*Arabidopsis thaliana*
(6)	EYYDILGIKPEATPTEIKKAYRRKAMETHPDKHPDDPDAQA--KFQAVGEAYQVLS-DPGLRSKYDQF	*Saccharomyces cerevisiae*
(6)	DYYAILGVPRNATQEEIKRAYKRLARQYHPD-VNKSPEAEE---KFKEINEAYAVLS-DPEKRRIYDTY	*Thermus thermophilus*
(5)	DYYEILGVSKTAEEREIRKAYKRLAMKYHPDRNQGDKEAEA---KFKEIKEAYEVLTDSQK-RAAYDQY	*Escherichia coli*
3	helix Ⅰ	helix Ⅱ loop helix Ⅲ helix Ⅳ

FIG. 1. Circular map of SpltMNPV genome. The center circle shows a scale in kilobase pairs and the sites for *Pst*I restriction enzyme. Arrows represent directions of transcription for the putative genes or ORFs. Names of characterized genes or ORFs with a transcription direction the same as *polyhedrin* (clockwise) are shown in the outer circle. Names of characterized genes or ORFs with a transcription direction in the counterclockwise direction are shown in the middle circle. Solid black squares indicate the positions of *hrs*.

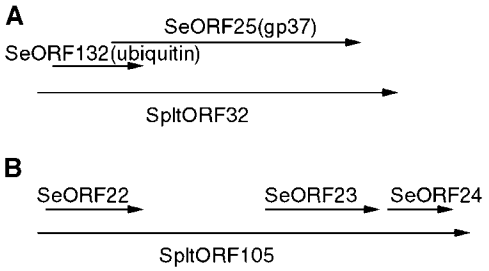


FIG. 2. ORFs fusion in SpltMNPV genome compared to Se ORFs using BLAST search. SpltORF32 shows certain identities with SeORF123 and SeORF25 (A); SpltORF105 with SeORF22-4 (B).

region of the ORF32 amino acid sequence was a ubiquitin-like moiety (about 228-bp position: 31193>31420, 76aa length) and the C-terminal, a region of GP37 (Fig. 2A). The ubiquitin protein is abundant in all eukaryotes and is one of the most highly conserved proteins known. Its principal function in cells is to signal the degradation of proteins by the 26S proteosome (Hochstrasser, 1996). In the eukaryotic cell, ubiquitin is generally encoded either as an in-frame fusion with another protein, often a ribosomal protein, or as a polyubiquitin gene, encoding from 5 to 18 ubiquitin monomers. In either case, C-terminal cleavage is necessary to release the ubiquitin monomer (O'Reilly, 1997). Ubiquitin-specific, ATP-independent proteases capable of cleaving ubiquitin from its linear or branched conjugate have been detected in all eukaryotes examined, but not in bacteria such as *Escherichia coli*, which lack ubiquitin and ubiquitin-specific enzymes. Both natural and engineered ubiquitin fusions to itself on other proteins are cleaved by processing proteases after the last (Gly^{76}) residue of ubiquitin (Baker *et al.,* 1992). Similar proteases must exist in SpltMNPV and its host system. The *gp37* (also known as *p34.8*) is homologous to spheroidin of entomopoxviruses (Yuen *et al.,* 1990), and it is involved in enhancement of virus infection *in vivo* (Phanis *et al.,* 1999). In addition, BLAST searches revealed that regions of SpltORF105 matched SeORF22, 23, and 24 (Fig. 2B).

Structural protein genes

The SpltMNPV genome possesses homologues of many AcMNPV structural protein genes. Structural protein genes of polyhedra included *polyhedrin*, polyhedron envelope (PE), or polyhedron calyx (*calyx/pep*) and *p10*. The other protein genes listed as structural protein genes by Hayakawa *et al.* (2000), encoding proteins *odv-e18*, *odv-ec27*, *odv-e56*, *p74*, *odv-e25*, and *odv-e66*, and nucleocapsid comprised *orf1629*, *pk-1*, *vp1054*, *gp41*, *vp91*, *vp39*, *p6.9*, *vp80, ld130*, and *p24*, are all present. The protein tyrosine phosphatase (*ptp*) gene and the major budded virus glycoprotein gene *gp64* were not found in SpltMNPV. The most conserved homologue of structural proteins was *polyhedrin* (ORF1, 81% identity), followed by *gp41* (ORF76, 60% identity) and *odv-e*56 (ORF17, 46% identity). Several homologues including *orf1629*, *p10*, *vp80*, *p24capsid* showed relatively low identities (less than 30%) to homologues of AcMNPV.

GP64 appears to be required for the spread of the infection to other cells and for the virus to exit from the infected cell for AcMNPV (Monsma *et al.,* 1996). Although a *gp64* gene was not identified, a homologue of *gp64* in LdMNPV ORF130, which has been shown to substitute for the lack of *gp64* (Pearson *et al.*, 2000), was found in the SpltMNPV genome (ORF136). Splt136 is predicted to encode both an N-terminal signal and a transmembrane domain and showed 31% identity to the LdMNPV ORF130.

Genes involved in DNA replication and late gene expression

Using transient assays, considerable progress has been achieved in the identification and characterization of genes that are involved in DNA replication and late gene expression. Nineteen late expression factors or *lef* genes exist in AcMNPV (Rapp *et al.,* 1998). Computer-assisted homology search of predicted amino acid sequences indicated that the SpltMNPV genome contained homologues of 17 AcMNPV genes involved in DNA replication and late gene expression including *ie-1*, *lef-1*, -2, -3, -4, -5, -6, -8, -9, -10, -11, -12, *dnapol*, *helicase*, *p47*, *39k*, and *p35*, but lacked *lef-7*. ORF80 showed weak homologies (22% identity) to OpMNPV ORF151 (*ie-2*) (data not shown in Table 1). The most conserved SpltMNPV LEF homologue was *lef-9* (ORF59, 63% identity to AcMNPV) followed by *lef-8* (ORF38, 56%) and *p47* (ORF36, 47%) (Table 1), all of which were components of the virus-specific RNA polymerase (Jin *et al.,* 1998). LEF-4, the other component of the RNA polymerase, showed 40% identity to the homologue of AcMNPV. In addition, three SpltMNPV homologues of AcMNPV LEFs including *ie-1*,

FIG. 3. Amino acid sequence alignment of J domain and the conserved J domain present in different organisms (indicated in red). The J domain of SpltMNPV ORF39 is shown at the top, followed by the sequence in other prokaryotic and eukaryotic J domains of DnaJ family proteins. The number in parentheses to the left of each sequence indicates the position in the polypeptide for the first residue listed. The polypeptide sequences were obtained from GenBank data bases by accession number as follows: *Aquifex aeolicus*, E70361; *Bacillus stearothermophilus*, JC4739; *Homo sapiens*, BAA90896.1; *Mus musculus*, BAA88308.1; *Chlamydia muridarum*, AAF39450.1; *Schizosaccharomyces pombe*, CAB85447.1; *Haemophilus influenzae* Rd, C64112; *Methylovorus sp.* SS1, AAC95379.1; *Saccharomyces cerevisiae*, S48085; *Thermus thermophilus*, AAB04678.1; *Escherichia coli*, AAC73126.1.

TABLE 1

Putative ORFs in SpltMNPV Strain G2

ORF[a]	Name[b]	Position/length (aa)[c]	Predicted Mr(D)[d]	Motifs[e]	Homologous ORFs/identity (%)[f]					BLAST[g]	
					Ac	Se	Ld	Op	Bm	Bestmatch	Score (bits)
1	*polyhedrin*	1 > 750 (249)	29,250	L	8/81	1/81	1/77	3/81	1/77	gb|AAC09246.1| (AF037262) [SpliNPV]	509
2	*orf1629*	747 < 2393 (548)	60,872	e	9/22	2/21	2/27	2/24	2/25	emb|CAA68047.1| (X99711) [SpliNPV]	252
3	*pk1*	2392 > 3204 (270)	31,512	—	10/38	3/41	3/45	1/39	3/38	gb|AAB94757.1| (AF039272) [SpliNPV]	494
4	*hoar*	3519 < 5714 (731)	83,086	EC	—	4/21	—	—	—	gb|AAF33535.1|AF169823_4 (AF169823) [SeMNPV]	59.3
5		6027 < 6230 (67)	7786	—							
6		6227 < 6511 (94)	11,043	—						gb|AAB96623.1| (AF019224) spheroidin [Heliothis armigera entomopoxvirus]	32.8
7		6447 > 8600 (717)	80,994	—						gb|AAG02962.1|AF250284 [Amsacta moorei entomopoxvirus]	48.0
8	*ie-0*	8851 > 9720 (289)	33,385	L	141/30	138/37	21/36	138/29	117/30	gb|AAF33667.1|AF169823_138 [SeMNPV]	183
9		9674 < 9829 (51)	6069	L							
10	*dutpase*	9774 > 10268 (164)	18,452	—	—	55/43	116/48	—	—	pir‖T10819dUTP pyrophosphatase (EC 3.6.1.23)	128
	hr1	10301 ~ 10800									
11		10801 > 12210 (469)	55,639	L	142/45	137/49	20/48	139/46	118/45	gb|AAB54096.1| (U67264) [HzNPV]	547
12	*odv-e18*	12234 > 12485 (83)	9222	e, L	143/47	136/60	19/61	140/48	119/73	gb|AAB54097.1| (U67264) [HzNPV]	52.7
13	*odv-ec27*	12517 > 13368 (283)	32,815	EC, L	144/45	135/49	18/51	141/44	120/45	gb|AAB54098.1| (U67264) [HzNPV]	301
14		13400 > 13681 (93)	10,802	L	145/48	134/52	17/51	142/50	121/44	gb|AAF33663.1|AF169823_134 [SeMNPV]	114
15		13726 < 14328 (200)	22,144	—	146/33	133/31	16/32	144/26	122/34	gb|AAA66776.1| (L22858) [AcMNPV]	83.5
16	*ie-1*	14501 > 16567 (688)	77,916	EC	147/26	132/32	15/30	145/25	123/27	gb|AAB54100.1| (U67264) [HzNPV]	233
17	*odv-e56*	16693 > 17808 (371)	40,075	L	148/46	6/43	14/47	146/45	124/47	gb|AAA98967.1| (U09501) [AcMNPV]	294
18		17823 > 18374 (183)	21,063	e, L	34/33	124/55	42/43	26/39	25/34	emb|CAA67757.1| (X99377) ORF 552 [SpliNPV]	363
19	*p10*	18439 > 18756 (105)	11,266	e, L	137/27	130/62	41/50	133/20	114/30	emb|CAA67758.1| (X99377) p10 [SpliNPV]	138
20		18734 > 19678 (314)	36,384	—						emb|CAA67759.1| (X99377) ORF 945 [SpliNPV]	511
21	*p74*	19706 < 21679 (657)	75,844	EC, L	138/53	131/50	27/50	134/53	115/54	emb|CAA67755.1| (X99376) p74 protein [SpliNPV]	1181
22		21682 < 22026 (114)	13,189	—	150/30	96/26	30/28	—	126/26	gb|AAF52851.1| (AE003627) [*Drosophila melanogaster*]	47.6
23	*rr1*	22066 < 24378 (770)	86,713	—	—	139/54	148/27	32/26	—	emb|CAA67423.1| ribonucleotide reductase [SpliNPV]	1304
24		24586 > 25785 (399)	47,302	—						gb|AAA29743.1| reticulocyte binding protein 1 [*Plasmodium vivax*]	38.7
	hr2	25805 ~ 26264									
25		26567 < 26746 (59)	7094	—							
	hr3	26751 ~ 27151									
26		27162 > 28055 (297)	34,822	—						gb|AAC14694.1 putative transcriptional repressor E2F-6 [*Homo sapiens*]	32.5
27	*me53*	28076 < 28981 (301)	36,293	L	139/21	7/25	23/25	137/23	116/21	gb|AAF33538.1|AF169823_7 (AF169823) [SeMNPV]	101
28		29013 > 29273 (86)	10,529	—	29/31	128/43	39/49	39/34	20/31	gb|AAC70224.1|AAC70224 (AF081810) [LdMNPV]	60.9
29	*lef-6*	29287 < 29724 (145)	16,957	L	28/29	127/29	38/32	40/25	19/34	gb|AAC70223.1|AAC70223 (AF081810) [LdMNPV]	50.4
30	*dbp*	29788 < 30660 (290)	33,296	e	25/22	126/28	47/24	43/29	16/24	gb|AAF33655.1|AF169823_126 (AF169823) [SeMNPV]	101

TABLE 1—*Continued*

ORF[a]	Name[b]	Position/length (aa)[c]	Predicted Mr(D)[d]	Motifs[e]	Homologous ORFs/identity (%)[f]					BLAST[g]	
					Ac	Se	Ld	Op	Bm	Bestmatch	Score (bits)
31		30718 > 31167(149)	17,300	—	26/39	125/42	36/35	42/41	17/37	dbj|BAA24257.1| (AB009614) [LsNPV]	166
32	ubi/gp37	31136 > 32191(351)	39,923	—	35/80	123/83	43/79	25/82	26/80	emb|CAA39250.1| (X55717) ubiquitin [*Phytophthora infestans*]	106
33	*39k*	32246 < 33214(322)	36,560	e	36/33	120/34	44/30	24/30	27/33	dbj|BAA24260.1| (AB009614) [LsNPV]	191
34	*lef-11*	33075 < 33509(144)	17,233	—	37/33	119/39	45/36	23/30	28/32	dbj|BAA24261.1| (AB009614) [LsNPV]	123
35		33479 < 34141(220)	26,557	EC, e, L	38/45	118/48	46/51	22/47	29/44	gb|AAC70231.1|AAC70231 (AF081810) [LdMNPV]	204
36	*p47*	34211 < 35479(422)	49,355	—	40/47	115/55	48/54	45/46	31/47	gb|AAC70233.1|AAC70233 (AF081810) [LdMNPV]	450
37	*lef-12*	35506 > 36111(201)	23,596	—	41/30	—	—	46/23	32/31	gb|AAA66671.1| (L22858) [AcMNPV]	51.9
	hr4	36102 ~ 36301									
38	*lef-8*	36316 < 39072(918)	106,196	—	50/56	112/60	51/57	54/53	39/56	emb|CAA71676.1| (Y10669) [SpliMNPV]	1615
39	*bjdp*	39071 > 39979(302)	34,631	—	51/27	111/31	—	55/28	40/27	dbj|BAA91724.1| (AK001496) [*Homo sapiens*]	110
40		40001 > 40570(189)	21,636	—			50/31			gb|AAC70235.1|AAC70235 Ld-helicase-2 [LdMNPV]	36.4
41		40590 < 40754 (54)	6225	L							
42	*chitA*	40767 < 42461(564)	62,841	L	126/60	19/55	70/60	124/62	103/60	gb|AAD31762.1|AF121457_1 [*Hyphantria cunea* NPV]	688
	hr5	42502 ~ 43051									
43		43021 < 43635(204)	24,533	EC						emb|CAB38995.1| (AL034558) predicted hexExon [*Plasmodium falciparum*]	40.6
44		43779 < 44405(208)	25,153	—	52/24	109/27	53/32	—	41/24	gb|AAC70238.1|AAC70238 [LdMNPV]	84.7
45		44475 > 44888(137)	16,283	L	53/47	108/55	54/50	56/42	42/47	gb|AAF33637.1|AF169823_108 [SeMNPV]	153
46		44926 < 46194(422)	47,242	L	—	107/23	55/24	—	—	gb|AAF33636.1|AF169823_107 [SeMNPV]	63.6
47		46199 < 46426 (75)	9088	L						gb|AAD35851.1|AE001746_12 phosphomannomutase [*Thermotoga maritima*]	32.1
48	*lef-10*	46386 > 46640 (84)	9169	EC, L	53a/37	106/39	56/38	57/32	42a/39	gb|AAF33635.1|AF169823_106 [SeMNPV]	55.0
49	*vp1054*	46474 > 47532(352)	40,888	e	54/37	105/44	57/49	58/33	43/38	gb|AAF33634.1|AF169823_105 [SeMNPV]	310
50		47653 > 47868 (71)	8209	—	55/37	104/40	58/37	59/35	44/29	gb|AAF33633.1|AF169823_104 [SeMNPV]	56.2
51		48172 > 48693(173)	20,391	EC, L	57/36	102/34	60/36	61/34	46/36	gb|AAA66687.1| (L22858) [AcMNPV]	98.3
52		48722 < 49246(174)	19,451	L	59/42	101/48	61/61	62/41	47/43	gb|AAF33630.1|AF169823_101 [SeMNPV]	81.1
53		49271 < 49519 (82)	9614	L	60/44	100/59	62/40	63/40	48/42	gb|AAF33629.1|AF169823_100 [SeMNPV]	82.3
54	*cathepsin*	49566 > 50579(337)	38,095	e, L	127/46	16/49	78/48	125/46	104/47	gb|AAF33546.1|AF169823_16 [SeMNPV]	284
55	*p49*	50628 > 51947(439)	51,943	—	135/30	—	—	—	112/30	emb|CAA07223.1| (AJ006751) [SpliNPV]	725
56		51886 < 52053 (55)	6708	—						gb|AAD20057.1| (AF131842) [*Homo sapiens*]	30.1
57	*fp*	52075 < 52668(197)	23,009	e, L	61/60	98/71	63/57	64/61	49/60	gb|AAF33627.1|AF169823_98 [SeMNPV]	276
58		52622 < 52825 (67)	7733	—							
59	*lef-9*	52839 < 54335(498)	57,413	—	62/63	97/69	64/66	65/61	50/63	gb|AAF33626.1|AF169823_97 [SeMNPV]	720
60		54253 > 54429 (58)	6542	L							
	hr6	54401 ~ 55642(413)									
61		55717 < 55950 (77)	9375	EC						emb|CAC09070.1| (AX026013) [*Lycopersicon esculentum*]	31.7
	hr7	55992 ~ 56166									
62		56221 < 57324(367)	42,871	—						gb|AAC66310.1| (AE000791) [*Borrelia burgdorferi*]	38.3

TABLE 1—*Continued*

ORF[a]	Name[b]	Position/length (aa)[c]	Predicted Mr(D)[d]	Motifs[e]	Homologous ORFs/identity (%)[f]					BLAST[g]	
					Ac	Se	Ld	Op	Bm	Bestmatch	Score (bits)
63	*rr2b*	57477 > 58478 (333)	38,445	—	—	45/57	120/72	34/21	—	gb|AAC70306.1|AAC70306 [LdMNPV]	484
64	*iap3/2*	58556 < 58966 (136)	15,590	L	—	110/36	139/53	35/48	—	gb|AAF33639.1|AF169823_110 [SeMNPV]	69.1
					71/48	88/50	79/27	74/31	58/48		
65		58899 < 59840 (313)	35,519	—	69/38	89/39	—	—	57/38	gb|AAF33618.1|AF169823_89 [SeMNPV]	198
66		59800 < 60198 (132)	15,120	—	68/50	90/61	80/57	73/43	56/49	gb|AAF33619.1|AF169823_90 [SeMNPV]	124
67	*lef-3*	60203 > 61276 (357)	41,503	e	67/24	91/28	81/26	72/25	55/25	gb|AAB53262.1| [SpliNPV]	615
68		61552 < 63924 (790)	91,676		66/21	92/25	82/26	71/22	54/21	gb|AAF33621.1|AF169823_92 [SeMNPV]	136
69	*dpol*	63926 > 66994 (1022)	118,024	EC	65/42	93/52	83/49	70/40	53/42	gb|AAF61904.1|AF215639_1 [SpliNPV]	1774
	hr8	67002 ~ 67499									
70		67509 > 68651 (380)	44,707	—						gb|AAF30799.1|AE002136_6 ATP/GTP-binding protein [*Ureaplasma urealyticum*]	40.6
71		68727 < 69353 (208)	23,449	L	74/22	—	135/29	77/25	60/23	gb|AAC70321.1|AAC70321 [LdMNPV]	49.2
72		69412 < 69795 (127)	14,647	L	75/23	94/35	84/37	78/25	61/24	gb|AAA58701.1| (U11242) [HzNPV]	132
73		69814 < 70068 (84)	9715	L	76/52	95/42	85/52	79/55	62/55	gb|AAC70271.1|AAC70271 [LdMNPV]	67.5
74	*vlf-1*	70133 < 71287 (384)	45,173	L	77/66	82/63	86/67	80/62	63/65	gb|AAC70272.1|AAC70272 [LdMNPV]	478
75		71307 < 71669 (120)	13,136	e, L	78/29	81/26	87/60	81/80	64/80	gb|AAA66708.1| [AcMNPV]	43.8
76	*gp41*	71666 < 72658 (330)	36,895	e, L	80/60	80/58	88/62	83/53	66/60	gb|AAA46748.1| [HzNPV]	441
77		72633 < 73331 (232)	27,160	—	81/39	79/43	89/43	84/39	67/39	gb|AAF33608.1|AF169823_79 [SeMNPV]	171
78	*tlp-20*	73210 < 73803 (197)	21,745	L	82/30	78/44	90/34	85/28	68/27	gb|AAF33607.1|AF169823_78 [SeMNPV]	92.4
79	*vp91capsid*	73772 > 76357 (861)	97,913	L	83/32	77/35	91/31	86/31	69/32	gb|AAF33606.1|AF169823_77 [SeMNPV]	468
	hr9	76362 ~ 76581									
80	*cg30*	76639 < 77391 (250)	28,987	—	88/20	76/32	—	89/33	71/21	gb|AAA66718.1| [AcMNPV]	50.0
81	*vp39*	77450 < 78358 (302)	33,874	EC, L	89/37	75/44	92/44	90/41	72/36	pir||JQ1582 [LdMNPV]	258
82	*lef-4*	78360 > 79787 (475)	54,706	—	90/40	74/47	93/43	91/33	73/39	gb|AAF33603.1|AF169823_74 [SeMNPV]	438
83		79833 < 80600 (255)	30,633	e	92/44	73/51	94/39	93/40	75/43	gb|AAF33602.1|AF169823_73 [SeMNPV]	286
84		80599 > 81147 (182)	21,429	EC	93/43	72/48	95/50	94/41	76/43	gb|AAF33601.1|AF169823_72 [SeMNPV]	152
85	*odv-e25*	81144 > 81827 (227)	24,895	e, L	94/33	71/55	96/55	95/29	77/33	gb|AAF33600.1|AF169823_71 [SeMNPV]	251
86	*helicase*	81918 < 85625 (1235)	144,746	L	95/36	70/41	97/42	96/32	78/36	gb|AAC70283.1|AAC70283 [LdMNPV]	1024
87		85594 > 86106 (170)	19,202	L	96/44	69/52	98/50	97/40	79/42	gb|AAF33599.1|AF169823_69 [SeMNPV]	196
88	*38k*	86114 < 87028 (304)	36,047	EC	98/42	67/51	99/42	99/40	82/43	gb|AAF33597.1|AF169823_67 [SeMNPV]	311
89	*lef-5*	86924 > 87832 (302)	34,909	—	99/37	66/40	100/41	100/33	83/36	gb|AAF33596.1|AF169823_66 [SeMNPV]	194
90	*p6.9*	87850 < 88104 (84)	9986	L	100/36	65/54	101/60	101/54	84/56	gb|AAC70287.1|AAC70287 [LdMNPV]	85.8
91		88162 < 89253 (363)	41,230	EC, L	101/37	64/43	102/35	102/31	85/37	gb|AAF33594.1|AF169823_64 [SeMNPV]	323
	hr10	89262 ~ 89651									
92		89657 < 90022 (121)	13,429	L	102/28	63/32	103/28	103/28	86/28	gb|AAF33593.1|AF169823_63 [SeMNPV]	56.6
93		90019 < 91140 (373)	44,039	L	103/48	62/49	104/49	104/42	87/49	gb|AAF33592.1|AF169823_62 [SeMNPV]	440
94	*vp80*	91167 > 93101 (644)	73,887	—	104/26	61/33	105/30	105/22	88/23	dbj|BAA24250.1| [SeMNPV]	232
95		93101 > 93268 (55)	6621	—	110/32	60/44	106/56	111/34	—	pir||T00199 hypothetical protein 110 [LsNPV]	80.8
96		93300 > 94385 (361)	41,598	e, L	109/44	59/48	107/52	109/42	92/44	dbj|BAA24252.1| [LsNPV]	489
97		94466 > 94804 (112)	13,184	e	108/33	58/28	108/31	108/41	91/31	dbj|BAA24253.1| [LsNPV]	80.4

TABLE 1—*Continued*

ORF[a]	Name[b]	Position/length (aa)[c]	Predicted Mr(D)[d]	Motifs[e]	Homologous ORFs/identity (%)[f]					BLAST[g]	
					Ac	Se	Ld	Op	Bm	Bestmatch	Score (bits)
98	*odv-e66*	94794 < 96872 (692)	78,129	L	46/36	57/37	131/47	50/35	37/37	gb|AAF05263.1|AF162221_149 [XcGV]	597
99	*p13+*	96875 < 97744 (289)	33,976	L	—	56/53	—	—	—	gb|AAF33586.1|AF169823_56 [SeMNPV]	309
	hr11	97753 ~ 98049									
100		98090 > 99055 (321)	37,193	EC	—	54/33	138/29	31/22	—	gb|AAF33584.1|AF169823_54 [SeMNPV]	122
101		99106 < 99816 (236)	27,162	e, L	106 +107/55	53/53	140/60	107/59	90/52	gb|AAF33583.1|AF169823_53 [SeMNPV]	242
102		99836 < 101209 (457)	51,990	L	—	52/22	141/23	—	—	gb|AAC70327.1|AAC70327 [LdMNPV]	43.0
103		101236 < 101775 (179)	20,891	—						gb|AAD07884.1| (AE000594) [*Helicobacter pylori* 26695]	32.1
104		101855 < 102055 (66)	6998	e						gb|AAB41388.1| (U51081) [*Drosophila paulistorum*]	35.6
105		102193 > 103449 (418)	46,655	—	—	22/32, 23/47, 24/51	—	—	—	gb|AAF33553.1|AF169823_23 [SeMNPV]	118
	hr12	103533 ~ 104801									
106		104792 < 105595 (267)	31,843	—						gb|AAF54120.1|CG10267 gene product [*Drosophila melanogaster*]	32.1
107		105624 > 106226 (200)	22,635	L	115/42	50/49	143/50	115/38	95/43	gb|AAC70329.1| (AF081810) [LdMNPV]	212
108		106234 > 106587 (117)	13,219	—						gb|AAB41487.1| (U51723) V-SERA 3 [*Plasmodium vivax*]	31.3
109	*alk-exo*	106605 < 107831 (408)	47,345	L	133/36	41/41	157/40	131/34	110/35	gb|AAF33571.1|AF169823_41 [SeMNPV]	285
110		107915 < 109045 (376)	43,375	e				98/20		gb|AAD41662.1| (AF074888) resistance protein [*Oryza sativa*]	35.6
111		109007 > 109165 (52)	6011	L							
112		109168 < 109551 (127)	14,817	EC, L	19/27	42/27	159/31	18/45	11/28	gb|AAF33572.1|AF169823_42 [SeMNPV]	37.9
113		109553 > 110758 (401)	47,128	EC	18/21	43/34	158/22	17/29	10/20	gb|AAF33573.1|AF169823_43 [SeMNPV]	189
114	*lef-2*	110817 < 111581 (254)	29,069	e	6/34	12/38	137/41	6/35	135/34	gb|AAC77814.1| (AF060564) [BusuNPV]	159
115		111433 < 111777 (114)	12,371	e, L						gb|AAF22572.1|AF136502_3 [HaNPV]	30.9
116	*p24capsid*	111825 > 112559 (244)	27,374	L	129/28	10/39	—	127/29	106/31	gb|AAF78933.1|AF266697_3 [HaNPV]	159
	hr13	112542 ~ 112961									
117		112963 > 113280 (105)	12,121	—						gb|AAF99266.1|AF232689_165 [rat cytomegalovirus Maastricht]	34.8
118		113090 > 115849 (919)	105,550	—	—	30/26	129/34	—	—	gb|AAC70315.1|AAC70315 (AF081810) [LdMNPV]	324
119		115866 < 116585 (239)	27,425	—	17/32	29/29	128/28	16/27	9/32	emb|CAA05885.1| (AJ003131) [SpliNPV]	358
120	*bro-a*	116653 < 117213 (186)	21,895	e	2/25	—	153/19	116/32	131/21	emb|CAA05886.1| (AJ003131) [SpliNPV]	318
121	*egt*	117465 < 119033 (522)	60,033	e	15/41	27/48	125/46	14/40	7/41	prf||2207425A [SpliNPV]	890
122	*fgf*	119165 > 119905 (246)	27,545	e	32/30	38/30	156/27	27/27	24/30	emb|CAA05888.1| (AJ003131) [SpliNPV]	372
123		119928 < 120161 (77)	9222	L	120/32	37/32		120/29	98/33	gb|AAF33567.1|AF169823_37 [SeMNPV]	32.8
124		120184 < 121764 (526)	59,840	L	119/39	36/43	155/45	119/46	97/40	emb|CAA05889.1| (AJ003131) [SpliNPV]	536
	hr14	121771 ~ 121926									
125	*bro-b*	122062 > 123498 (478)	55,601	—	2/28					gb|AAF05174.1|AF162221_60 [XcGV]	696
126		123737 > 125107 (456)	54,883	EC						gb|AAF05179.1|AF162221_65 [XcGV]	531

TABLE 1—*Continued*

ORF[a]	Name[b]	Position/length (aa)[c]	Predicted Mr(D)[d]	Motifs[e]	Homologous ORFs/identity (%)[f] Ac	Se	Ld	Op	Bm	BLAST[g] Bestmatch	Score (bits)
127		125316 < 125834 (172)	19,857	EC						gb|AAC34739.1| alpha-glycerophosphate oxidase [*Enterococcus casseliflavus*]	32.8
128	38.7 k	125891 < 126919 (342)	40,746	—	13/22	13/29	122/34	12/26	5/22	gb|AAF33543.1|AF169823_13 [SeMNPV]	150
129	*lef-1*	126906 < 127601 (231)	27,424	—	14/41	14/43	123/42	13/39	6/41	gb|AAF78939.1|AF266697_9 [HaNPV]	217
130		127582 < 127950 (122)	14,224	L	—	15/32	124/32	—	—	gb|AAF33545.1|AF169823_15 [SeMNPV]	36.4
131		127947 < 128477 (176)	20,999	e						gb|AAC24399.1|K11H12.4 gene product [*Caenorhabditis elegans*]	32.8
132	*calyx/pep*	128481 < 129515 (344)	38,340	e, L	131/31	46/50	136/37	129/32	108/35	gb|AAF78935.1|AF266697_5 [HaNPV]	196
133	*pkip*	129544 > 130161 (205)	23,957	—	24/30	32/25	110/22	44/26	15/32	gb|AAF33562.1|AF169823_32 [SeMNPV]	60.9
134	*arif-1*	130199 < 130936 (245)	27,640	—	21/23	34/27	118/29	19/24	12/26	gb|AAF33564.1|AF169823_34 arif1 [SeMNPV]	76.1
135		130910 > 132187 (425)	48,474	—	22/56	35/59	119/57	20/54	13/54	gb|AAF33565.1|AF169823_35 [SeMNPV]	474
	hr15	132105 ~ 133451									
136		133451 > 135499 (682)	76,839	L	23/24	8/35	130/31	21/19	14/23	gb|AAF33539.1|AF169823_8 [SeMNPV]	394
137		135545 < 136240 (231)		L						gb|AAC74109.1| (AE000204) putative outer membrane protein [*E. coli*]	32.5
138		136338 > 137117 (259)	31,186	EC						emb|CAB72227.1| (AL138854) [*Schizosaccharomyces pombe*]	34.0
	hr16	137151 ~ 137486									
139		137436 < 137618 (60)	7594	—						gb|AAF52223.2| (AE003608) [*Drosophila melanogaster*]	
	hr17	137619 ~ 138104									
140		138104 < 138952 (282)	34,027	EC						pir‖JQ1045 arylphorin precursor [*Calliphora vicina*]	34.8
141		138975 < 139160 (61)	7348	—						gb|AAD26382.1 glycoprotein [*Kaposi's sarcoma-associated herpesvirus*]	29.0

[a] ORF number of SpltMNPV.
[b] Name of the ORF in other NPV.
[c] ORF's position and direction in SpltMNPV (number of amino acids).
[d] Molecular mass of the predicted ORF.
[e] Motif of ORF where EC = early promoter motif (TATA box followed by CAGT motif 20–25 bp downstream) within 180 bp of the initiation codon; e = CGTGC motif within 210 bp of the initiation codon; L = late promoter motif (A/T/G)TAAG within 120 bp of the initiation codon.
[f] Homology ORF (above slant) and amino acid sequence identity to AcMNPV, SeMNPV, LdMNPV, OpMNPV, and BmNPV, respectively.
[g] BLAST result including bestmatch Accession number, name of species, (in parentheses), and bestmatch score. For ORF BLAST searches where the identity was not found, similarities (maximum match) were calculated by DNASTAR and those results are underlined.
Abbreviations: SpliNPV, *Spodoptera littoralis* NPV; HzNPV, *Helicoverpa zea* NPV; LsNPV, *Leucania separata* NPV; BusuNPV, *Buzura suppressaria* NPV; HaNPV, *Helicoverpa armigera* NPV.

lef-3, and *lef-6* showed relatively low identities (less than 30%).

The very late gene expression appears to be modulated by *vlf-1* (Todd *et al.,* 1996). The product of the gene VLF-1 has sequence motifs characteristic of a family of integrases and resolvases (McLachlin and Miller, 1994). The *vlf-1* gene is expressed as a late gene, and VLF-1 is an essential and limiting factor in *polh* expression (Yang and Miller, 1999). In SpltMNPV the genome *vlf-1* was well conserved, with 66% amino acid sequence identity to the homologue of AcMNPV.

Inhibitor of apoptosis (*iap*) genes

Baculoviruses possess two genes with antiapoptotic activity: *p35* (also known as *p49*, Du *et al.,* 1999) and *iap*, which can suppress apoptosis induced by virus infection or by diverse stimuli in vertebrates or invertebrates (Miller, 1997). In the SpltMNPV genome, a homologue of

TABLE 2

SpltORF125 Different Identities to Homologues of Other NPVs

ORFs	Identity (%)	Length (aa)	Score (bits)	bro-group
Ld-bro-c	(240/487) 49	528	446	III
Ld-bro-d	(222/488) 45	510	417	III
Ld-bro-i	(132/271) 48	403	252	III
Ld-bro-n	(62/229) 27	338	76.1	I
Ld-bro-p	(64/250) 25	337	75.3	I
Ac-bro-a	(64/225) 28	328	74.9	I
Ld-bro-a	(60/234) 25	350	74.1	II
bm-bro-d	(76/322) 23	349	73.4	I
Bm-bro-c	(53/174) 30	318	73.4	I
Ld-bro-o	(73/281) 25	336	72.2	II
Bm-bro-a	(50/176) 28	317	70.2	I
Ld-bro-j	(78/349) 22	403	69.5	I
Ld-bro-b	(51/187) 27	323	68.7	I
Ld-bro-l	(41/124) 33	353	64.0	IV
Bm-bro-e	(31/89) 34	241	53.9	I
Ld-bro-k	(38/113) 33	238	52.3	I
Bm-bro-b	(31/104) 29	239	51.5	I
Ld-bro-g	(33/115) 28	222	48.4	IV
Ld-bro-m	(37/115) 32	243	47.3	II
Ld-bro-f		129		IV
Ld-bro-h		85		IV

Note. Identity and score data obtained from BLAST results. Length and bro-group as according to Ayres *et al.* (1994), Kuzio *et al.* (1999), and Gomi *et al.* (1999).

p35 was identified. It has a predicted amino acid sequence of 51.9 kDa that showed 30% identity to the AcMNPV *p35* gene and 79% identity to the *p49* gene of *Spodoptera littoralis* nucleopolyhedrovirus (SpliMNPV). In the SpltMNPV genome, only one *iap* gene was identified (ORF64 in Table 1). IAP homologues often contain two tandem baculovirus IAP repeats (BIR) and a carboxyl-terminal C3HC4 zinc finger-like motif (Birnbaum *et al.,* 1994). Alignment of SpltMNPV ORF64 revealed that it contained one BIR motif and one zinc finger-like motif (data not shown). This suggested that ORF64 was similar to the IAP-3 of LdMNPV (Kuzio *et al.,* 1999).

Baculovirus-repeated ORFs (*bro* genes)

The *bro* genes are present in a number other baculoviruses with 1 to 16 copies. In the SpltMNPV genome, two homologues of AcMNPV ORF2, *bro-a* and *bro-b*, were identified. Their predicted molecular masses were 21.9 and 55.6 kDa, respectively. The *bro-a* had low identity to the homologues of other baculoviruses (Table 1). BLAST search revealed that the predicted amino acid sequence had 21% identity (139 amino acids overlap) to the ORF131 of BmNPV(*bro-d*). Motif search showed *bro-a* had a CAGT early-gene start motif (upstream 79–82 nt), which is the early gene start site for all *bro* genes transcription (Kang *et al.,* 1999). The *bro-b* had identities in various degrees to the homologues in AcMNPV, LdMNPV, and BmNPV (Table 2). On average, *bro-b* (ORF125) showed about 32% amino acid sequence identity to the *bro* LdMNPV genes except *bro-f* and *bro-h*. The amino acid sequence of *bro-b* had the maximum identity to the *bro-d* (49%), followed by *bro-i* (48%) and *bro-d* (45%) of LdMNPV.

Other genes

In the SpltMNPV genome, there are a number of auxiliary genes (O'Reilly, 1997). Homologues of chitinase (*chitA*), cathepesin, and ecdysteroid UDP-glucosyltransferase (*egt*) were conserved, with 60, 46, and 40% amino acid sequence identities to the homologues of AcMNPV, respectively. The fibroblast growth factor (*fgf*) gene had weak homologies to AcMNPV (30% identity). Strikingly, it lacked a homologue of superoxide dismutase (*sod*) gene.

Genes involved in nucleotide metabolism, including the large and small subunits of ribonucleotide reductase (*rr1* and *rr2b*) and dUTPase, were present in SpltMNPV genome. BLAST searches showed ORF63 (*rr2b*) had 72% identity to the homologue of LdMNPV, which is closely related to eukaryotic *rr2* sequences (Hayakawa *et al.,* 2000).

DnaJ is one of the heat-shock proteins in *E. coli* (Hartl, 1996). DnaJ family proteins have a modular organization called the J domain, which is highly conserved within 70

TABLE 3

Lengths, Base Content, and Repeats in SpltMNPV hrs

Name	Position/length (bp)	G + C content (%)	P-I[a]	P-II[b]
hr1	10,301–10,800 (500)	37.5	4	
hr2	25,805–26,262 (460)	39.5	6	
hr3	26,751–27,151 (401)	32.9		4
hr4	36,102–36,301 (300)	27.5	2	
hr5	42,501–43,051 (550)	28.5	2	
hr6	54,401–55,642 (1242)	29.2	3	20
hr7	55,992–56,166 (175)	33.1	2	
hr8	67,002–67,499 (498)	31.7	4	
hr9	76,362–76,581 (220)	33.6	2	
hr10	89,262–89,651 (390)	32.6	4	
hr11	97,753–98,049 (297)	39.3	4	
hr12	103,533–104,801 (1269)	36.3	9	8
hr13	112,542–112,961 (420)	37.3	3	
hr14	121,771–121,926 (156)	36.6	2	
hr15	132,105–133,451 (1347)	30.5	3	26
hr16	137,151–137,486 (366)	37.5	4	
hr17	137,619–138,104 (486)	29.1		13
Average	534	33.4	54[c]	71[c]

[a] P-I: *G*AAAAGTCgGcCA*C*GTT*CG*AtT*CG*AAC*GT*GtCtGACTTTT*C*.

[b] P-II: AACATGTTtATgAAC*A*TG*TT* (lowercase letters indicate no complementary bases, italicized uppercase letters indicate relatively conserved bases, and bold uppercase letters indicate highly conserved bases).

[c] Sum of repeat in all hrs.

FIG. 4. Complete consensus sequence of a representative SpltMNPV *hrs,* hr6 (54,401–55,642), showing the two imperfect palindromic repeats. P-I is indicated by the boxed regions and P-II is indicated by the gray regions.

amino acids near the N-terminus (Kelley and Georgopoulos, 1997). It has been demonstrated that DnaJ stimulates the rate of hydrolysis of DnaK-bound ATP in *E. coli* (Tomoyasu *et al.,* 1998; Suh *et al.,* 1998). Several eukaryotic homologues of DnaJ have been found in various organisms including *Saccharomyces cerevisiae*, *Homo sapiens*, *Cucumis sativus*, mouse (M27 murine lung carcinoma cell line), and *Drosophila melanogaster* (Luke *et al.,* 1991; Raabe and Manley, 1991; Preisig-Muller and Kindl, 1993; Brightman *et al.,* 1995; Iliopoulos *et al.,* 1997). BLAST homology search revealed that the N-terminus (about 70 aa) of ORF39 that we named *bjdp* (baculovirus J-domain protein, bJDP), exhibited different levels of homology to the J domain of different DnaJ proteins (Fig. 3). The highly conserved tripeptide histidine-proline-aspartate (HPD) existed in a loop between helices II and III. In addition to HPD, helix-I was well conserved. Although the predicted *bjdp* amino acid sequence had 31% identity to SeMNPV ORF111, the homologous region was located in the middle of the peptide (position: 147>242), not in the N-terminus (data not shown in Table 1).

Unique ORFs to SpltMNPV

Twenty-nine ORFs in the SpltMNPV genome were unique to this virus. Most of them showed slight similarities with ORFs or genes of some species (BLAST score about 30–50 bits), but seven ORFs did not share any significant homology to any sequence in the GenBank (Table 1).

Homologous regions

An important feature of many baculovirus genomes described previously is the dispersal of several *hrs* over

the genome. The *hrs* are usually composed of repeated sequences encompassing both imperfect palindromes and direct repeats that are homologous to one another (Ayres *et al.,* 1994; Kool *et al.,* 1995; Ahrens *et al.,* 1997; Kuzio *et al.,* 1999). The *hrs* have been demonstrated as *cis*-acting enhancers of RNA polymerase II-mediated transcription and can also act as the origin of DNA replication in transient replication assay (Broer *et al.,* 1998; Pearson *et al.,* 1995). In addition, evidence suggests that *hrs* are the binding sites of the baculovirus transactivator, *ie-1* (Rodems and Friesen, 1995; Choi and Guarino, 1995). It has also been suggested that *hrs* may be sites of recombination within or between baculovirus genomes (Hayakawa *et al.,* 2000).

Seventeen *hrs* were identified in SpltMNPV genome, each containing 2–29 palindromic repeats, with an average length of 534 bp and base content (G+C%) of 33.0 (Table 3). Instead of composing a single palindromic motif, which is a major feature of other baculovirus *hrs*, SpltMNPV *hrs* contain two typical imperfect palindromic repeats (designated as P-I and P-II, respectively; Table 3, Fig. 4). Remarkably, 41-bp P-I in those *hrs* had no obvious conserved flanking sequence. SeMNPV *hrs* is the only other baculovirus in which this feature has been reported (Ijkel *et al.,* 1999). Other 20-bp P-II (especially in *hr*6 and *hr*15) were flanked by 9–20 bp of conserved sequence, which was similar to other MNPVs except SeMNPV. Sequence analysis of these two palindromes showed no close relation to the *hrs* repeats of AcMNPV, SeMNPV, LdMNPV, OpMNPV, and BmMNPV. However, the genome context of SpltMNPV *hrs* exhibited some similarity to several other baculoviruses. For instance, SpltMNPV *hr9* was located between ORF79 and ORF80 that were the homologues of AcMNPV ORF83 and ORF88, respectively. Moreover, AcMNPV *hr*3 follows AcMNPV ORF83.

GeneParityPlot analysis and evolution

Conserved gene clusters and their distribution along the genome were located using GeneParityPlot analysis. The homologues, which exist in all five genomes, were chosen for comparison between SpltMNPV and AcMNPV, SeMNPV, LdMPV, OpMNPV, and BmNPV, respectively. Comparisons of the gene arrangement of the selected ORFs are shown in Figs. 5A–5E.

Ten potential clusters conserved in all of the baculovirus genomes compared were identified. These clusters were numbered according to their sequential appearance in the GeneParityPlots: 1: AcORF142-147; 2: AcORF36-38; 3: AcORF52-53-53a-54-55; 4: AcORF61-62; 5: AcORF68-65; 6: AcORF75-76; 7: AcORF77-104; 8: AcORF109-108; 9: AcORF13-14; 10: AcORF21-22. On the whole, the genomic organization of SpltMNPV was similar to other NPVs except for the SeMNPV genome, which was reversed (Fig. 5B). GeneParityPlots demonstrated that the chitinase gene has drifted out of alignment compared with other baculovirus genomes. This interesting characteristic of SpltMNPV indicates that this gene was acquired at a different time in evolution and warrants further research. The data suggest that the genome organization of SpltMNPV is more closely related to SeMNPV, which coincides with the BLAST results (in Table 1). In addition, BLAST homology searches showed that many ORFs or genes of SpltMNPV had high identities to SpliMNPV with maximum identity reaching 96% (*polyhedrin* gene).

MATERIALS AND METHODS

Cloning of viral genotype

The ZSU strain of SpltMNPV was originally separated from dead larvae of cotton leaf worm, *S. litura*, in the suburbs of Guangzhou in 1976 (Pang, 1994). This strain was later selected to be developed as a commercial bioinsecticide registered under the trademark "Chongwen No. 1" in 1996. SpltMNPV genotype strain G2 was isolated from this strain following a modification of the *in vivo* method described by Smith and Crook (1988).

DNA sequencing

SpltMNPV DNA was extracted according to the methods described by Summers and Smith (1986). The DNA was treated by sonication (Ultrasonic processor JY92-II, 650 W, 4–6 s) and 2–3 kb fragments recovered by electrophoresis on an agrose gel and filled in. The DNA with blunt terminus was ligated to a pUC18 (*Sma*I) and transformed into *E. coli* DH5α by electroporation. Bacteria were incubated on prepared LB/antibiotic plates containing X-gal and ITPG overnight at 37°C. Recombinants were selected and inoculated in 1ml 2× YT medium in 96-well plates, incubated overnight at 37°C. Shotgun clones were sequenced to generate five- to sixfold coverage of total DNA base of SpltMNPV. Sequencing reactions were run on MegaBACE1000 sequencers using the tetracolor fluorescently labeled terminator method. In total 2400 reactions were carried out to contig the sequence of the SpltMNPV. The gaps were then closed using PCR and resequencing of individual shotgun clones. To confirm the assembly of the entire SpltMNPV sequence, predicted and actual restriction digest patterns were compared.

DNA sequence analysis

Genomic DNA composition structure, homologous regions, motif search, and restriction enzyme patterns were analyzed using DNASIS and DNASTAR. Open reading frames (ORFs) were selected using the same criteria as Gomi *et al.* (1999) but only the fragments larger than 150 bp were used. BLAST searching was performed

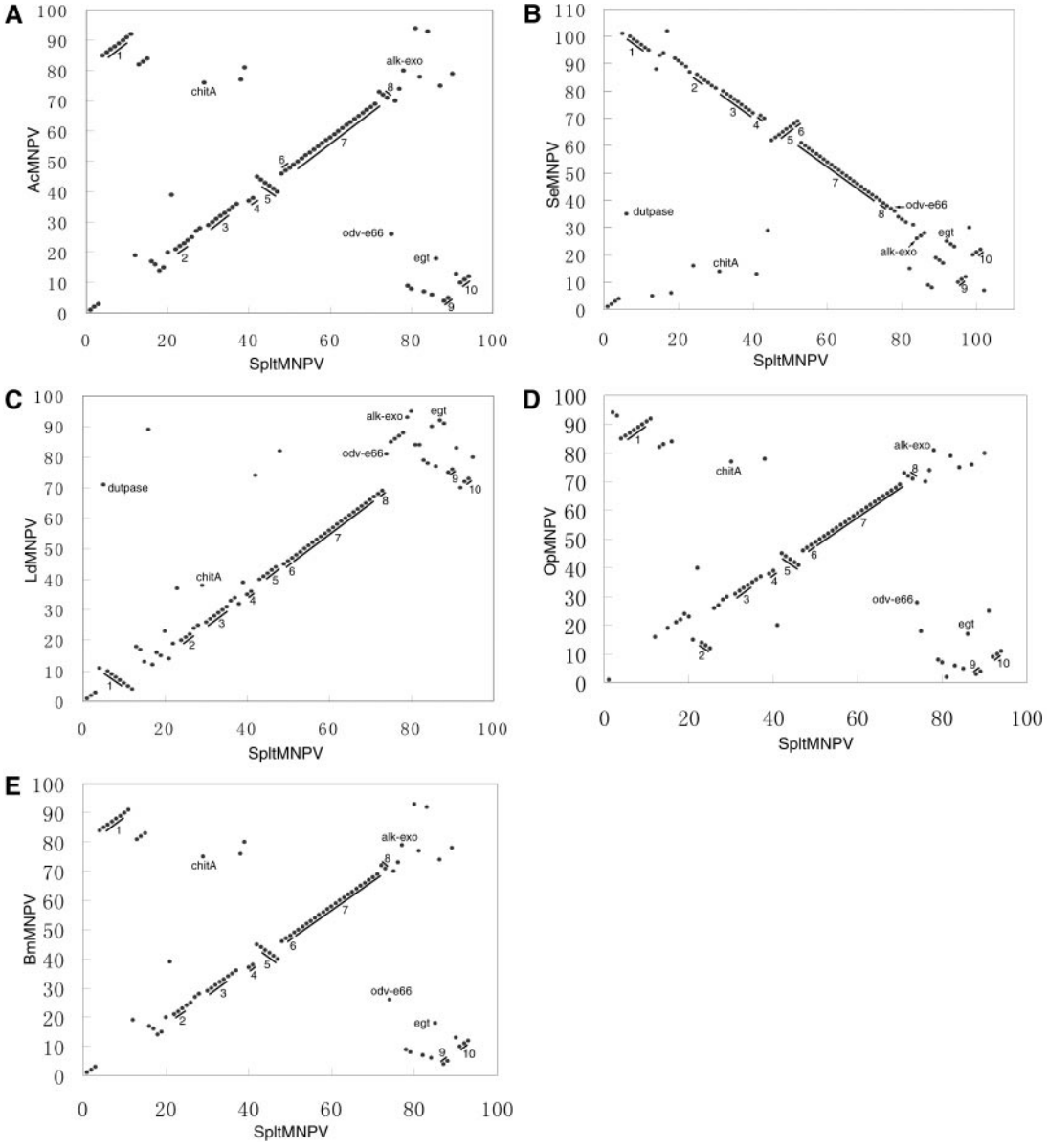


FIG. 5. GenParityPlots of SpltMNPV vs AcMNPV (A), SeMNPV (B), LdMNPV (C), OpMNPV (D), and BmNPV(E). The plots are graphic representations of the collinearity of baculovirus genomes obtained by GeneParityPlot analysis. Lines indicate 10 putative conservative gene clusters. The positions of the *chitA*, *odv-e66*, *alk-exo*, and *egt* genes are indicated.

using the Gap-BLAST search engine (Altschul *et al.,* 1997). GeneParityPlot analysis was performed on the SpltMNPV genome vs the genomes of AcMNPV, SeMNPV, LdMNPV, OpMNPV, and BmNPV, as described by Hu *et al.* (1998).

ACKNOWLEDGMENTS

This research was partially supported by the National Natural Science Foundation of China (39730030) and the National Major Basic Research Project ("973") of China (G2000016209). We thank Dr. Debbie Rae for comments on the manuscript. GenBank Accession No. of the SpltMNPV genomic sequence reported in this paper is AF325155.

REFERENCES

Ahrens, C. H., Russell, R. L., Funk, C. J., Evans, J. T., Harwood, S. H., and Rohrmann, G. F. (1997). The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **229,** 381–399.

Altschul, S. F., Madden, T., Schäffer, A. A., Zhang, J, Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M., and Possee, R. D. (1994). The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202,** 586–605.

Baker, R. T., Tobias, J. W., and Varshavsky, A. (1992). Ubiquitin-specific proteases of *Saccharomyces cerevisiae. J. Biol. Chem.* **267,** 23364–23375.

Birnbaum, M. J., Clem, R. J., and Miller, L. K. (1994). An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *J. Virol.* **68,** 2521–2528.

Brightman, S. E., Blatch, G. L., and Zetter, B. R. (1995). Isolation of a mouse cDNA encoding MTJ1, a new murine member of the DnaJ family of proteins. *Gene* **153,** 249–254.

Broer, R., Heldens, J. G., van Strien, E. A., Zuidema, D., and Vlak, J. M. (1998). Specificity of multiple homologous genomic regions in *Spodoptera exigua* nucleopolyhedrovirus DNA replication. *J. Gen. Virol.* **79,** 1563–1572.

Chen, Q. J., Pang, Y, and Li, G. H. (1998a). *Spodoptera litura* nucleopolyhedrovirus insecticide: Chongwen-I. *J. Wuhan Univ.* **44,** 183. [Chinese]

Chen, S. W., Wei, Y. J., Long, Q. X., Xu, A. L., and Wang, X. Z. (1998b). Cloning and sequencing of *p74* gene of *Spodoptera litura* nuclear polyhedrosis virus. *Acta Sci. Nat. Univ. Sunyatseni* **37,** 65–69. [Chinese with English abstract]

Choi, J., and Guarino, L. A. (1995). The baculovirus transactivator IE1 bands to viral enhancer elements in the absence of insect cell factors. *J. Virol.* **69,** 4548–4551.

Du, Q., Lehavi, D., Faktor, O., Qi, Y., and Chejanovsky N. (1999). Isolation of an apoptosis suppressor gene of the *Spodoptera littoralis* nucleopolyhedrovirus. *J. Virol.* **73,** 1278–1285.

Gomi, S., Majima, K., and Maeda, S. (1999). Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus *J. Gen. Virol.* **80,** 1323–1337.

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature* **381,** 571–579.

Hayakawa, T., Ko, R., Okano, K., Seong, S. I., Goto, C., and Maeda, S. (1999). Genome sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* **262,** 277–297.

Hayakawa, T., Rohrmann, G. F., and Hashimoto Y. (2000). Patterns of genome organization and content in Lepidopteran baculoviruses. *Virology* **278,** 1–12.

Hochstrasser, M. (1996). Protein degradation or regulation: Ub the judge. *Cell* **84,** 813–815.

Hu, G. D., Pang, Y., Yang, K., Li, C. B., and Su, D. M. (2000). Cloning and sequence analysis of the chitinase gene of *Spodoptera litura* nucleopolyhedrovirus. *Acta Biochim. Biophys.* **32,** 537–540. [Chinese with English abstract]

Hu, Z. H., Arif, B. M., Jin, F., Martens, J. W., Chen, X. W., Sun, J. S., Zuidema, D., Goldbach, R. W., and Vlak, J. M. (1998). Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **79,** 2841–2851.

Ijkel, W. F., van Strien, E. A., Heldens, J. G., Broer, R., Zuidema, D., Goldbach, R. W., and Vlak, J. M. (1999). Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **80,** 3289–3304.

Iliopoulos, I., Torok, I., and Mechler, B. M. (1997). The DnaJ60 gene of *Drosophila melanogaster* encodes a new member of the DnaJ family of proteins. *Biol. Chem.* **378,** 1177–1181.

Jin, J., Dong, W., and Guarino, L. A. (1998). The LEF-4 subunit of baculovirus RNA polymerase has RNA 5′-triphosphatase and ATPase activities. *J Virol.* **72,** 10011–10019.

Kang, W., Suzuki, M., Zemskov, E., Okano, K., and Maeda, S. (1999). Characterization of baculovirus repeated open reading frames (*bro*) in *Bombyx mori* nucleopolyhedrovirus. *J. Virol.* **73,** 10339–10345.

Kelley, W. L., and Georgopoulos, C. (1997). The T/t common exon of simian virus 40, JC, and BK polyomavirus T antigens can functionally replace the J-domain of the *Escherichia coli* DnaJ molecular chaperone. *Proc. Natl. Acad. Sci. USA* **94,** 3679–3684.

Kool, M., Ahrens, C. H., Vlak, J. M., and Rohrmann, G. F. (1995). Replication of baculovirus DNA. *J. Gen. Virol.* **76,** 2103–2118.

Kuzio, J., Pearson, M. N., Harwood, S. H., Funk, C. J., Evans, J. T., Slavicek, J. M., and Rohrmann, G. F. (1999). Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar. Virology* **253,** 17–34.

Li, Z., Long, Q. X., Zhang, Y. G., Wang, X. Z., and Pang, Y. (2000). Cloning and sequence analysis of *p49* gene from *Spodoptera litura* nucleopolyhedrovirus (*Splt*NPV). *Acta Sci. Natl. Univ. Sunyatseni* **39,** 73–76. [Chinese with English abstract]

Luke, M. M., Sutton, A., and Arndt, K. T. (1991). Characterization of SIS1, a *Saccharomyces cerevisiae* homologue of bacterial dnaJ proteins. *J. Cell Biol.* **114,** 623–638.

McLachlin, J. R., and Miller, L. K. (1994). Identification and characterization of vlf-1, a baculovirus gene involved in very late gene expression. *J. Virol.* **68,** 7746–7756.

Miller, L. K. (1997). Baculovirus interaction with host apoptotic pathways. *J. Cell. Physiol.* **173,** 178–182.

Monsma, S. A., Oomens, A. G. P., and Blissard, G. W. (1996). The GP64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. *J. Virol.* **70,** 4607–4616.

Murphy, F. A., Fauquet, C. M., Biship, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A., and Summers, M. D. (1995). Virus Taxonomy. *Sixth Report of the International Committee on Taxonomy of Viruses.* Springer-Verlag Wien, New York.

O'Reilly, D. R. (1997). Auxiliary genes of baculoviruses. *In* "The Baculoviruses" (L. K. Miller, Ed.), pp. 267–295. Plenum Press, New York.

Pang, Y. (1994). Viral diseases. *In* "Insect Pathology" (Z. L. Pu, Ed.), pp. 85–216. Guangdong Science and Technology Press, Guangzhou. [Chinese]

Pang, Y. (1998). Study and application of insect pathogens. *In* "Biological Control in China" (J. Z. Bao *et al*., Eds.), pp. 85–216. Shanxi Science and Technology Press, Taiyuan. [Chinese]

Pearson, M. N., and Rohrmann, G. F. (1995). *Lymantria dispar* nuclear polyhedrosis virus homologous regions: Characterization of their ability to function as replication origins. *J. Virol.* **69,** 213–221.

Pearson, M. N., Groten, C., and Rohrmann, G. F. (2000). Identification of the *lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the *Baculoviridae*. *J Virol.* **74,** 6126–6131.

Phanis, C. G., Miller, D. P., Cassar, S. C., Tristem, M., Thiem, S. M., and O'Reilly, D. R. (1999). Identification and expression of two baculovirus *gp37* genes. *J. Gen. Virol.* **80,** 1823–1831.

Preisig-Muller, R., and Kindl, H. (1993). Plant *dnaj* homologue: Molecular cloning, bacterial expression, and expression analysis in tissues of cucumber seedlings. *Arch. Biochem. Biophys.* **305,** 30–37.

Raabe, T., and Manley, J. L. (1991). A human homologue of the *Escherichia coli* DnaJ heat-shock protein. *Nucleic Acids Res.* **19,** 6645.

Rapp, J. C., Wilson, J. A., and Miller, L. K. (1998). Nineteen baculovirus open reading frames, including LEF-12, support late gene expression. *J. Virol.* **72,** 10197–10206.

Rodems, S. M., and Friesen, P. D. (1995). Transcriptional enhancer activity of hr5 requires dual-palindrome half sites that mediate binding of a dimeric form of the baculovirus transregulator IE1. *J. Virol.* **69,** 5368–5375.

Smith, I. L., and Crook, N. E. (1988). *In vivo* isolation of baculovirus genotype. *Virology* **166,** 240–244.

Suh, W. C., Burkholder, W. F., Lu, C. Z., Zhao, X., Gottesman, M. E., and Gross, C. A. (1998). Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. *Proc. Natl. Acad. Sci. USA* **95,** 15223–15228.

Summers, M. D., and Smith, G. E. (1986). "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures." Texas A&M Univ., College Station, Texas.

Todd, J. W., Passarelli, A. L., Lu, A., and Miller, L. K. (1996). Factors regulating baculovirus late and very late gene expression in transient-expression assays. *J. Virol.* **70,** 2307–2317.

Tomoyasu, T., Ogura, T., Tatsuta, T., and Bukau, B. (1998). Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli. Mol. Microbiol.* **30,** 567–581.

Wei, Y. J., Long, Q. X., Chen, S. W., and Wang, X. Z. (1998a). Nucleotide sequence and characterization of the *p10* gene of *Spodopt-*

era litura nuclear polyhedrosis virus. *Acta Biochim. Biophys. Sin.* **30,** 550–555.

Wei, Y. J., Long, Q. X., Chen, S. W., and Wang, X. Z. (1998b). Partial nucleotide sequence of protein kinase gene of *Spodoptera litura* nuclear polyhedrosis virus. *Acta Sci. Nat. Univ. Sunyatseni* **37,** 119–121. [Chinese with English abstract]

Wei, Y. J., Long, Q. X., Chen, S. W., and Wang, X. Z. (1999). Restriction patterns and nucleotide sequence of the polyhedrin gene of *Spodoptera litura* nuclear polyhedrosis virus. *Microbiology* **26,** 88–92. [Chinese with English abstract]

Yan, Q. S., Pang, Y., Yang, J., Nong, G., Ouyang, X. G., and Dai, X. J. (1999). Characterization of the ecdysteroid UDP-glucosyltranstferase gene of *Spodoptera litura* multinucleocapsid nuclear polyhedrosis virus. *Chin. J. Biotechnol.* **15,** 113–119.

Yang, S., and Miller. L. K. (1999). Activation of baculovirus very late promoters by interaction with very late factor 1. *J. Virol.* **73,** 3404–3409.

Yuen, L., Dionne, J., Arif, B., and Richardson, C. (1990). Identification and sequencing of the spheroidin gene of *Choristoneura biennis* entomopoxvirus. *Virology* **175,** 427–433.

Zheng, J., Li, Z., Long, Q. X., Wei, Y. J., and Wang, X. Z. (2000). Nucleotide sequence of a 4730 base pairs region of the *Spodoptera litura* nucleopolyhedrovirus genome. *Acta Biochim. Biophys. Sin.* **32,** 445–450. [Chinese with English abstract]