



OPEN

Genome sequence and organization of the *Mythimna* (formerly *Pseudaletia*) *unipuncta* granulovirus Hawaiian strain

Yinü Li^{1,4}, Xingjian Liu^{1,4}, Ping Tang³, Huan Zhang², Qilian Qin²✉ & Zhifang Zhang¹✉

Purified occlusion bodies (OBs) of *Mythimna* (formerly *Pseudaletia*) *unipuncta* (the true armyworm) granulovirus Hawaiian strain (MyunGV-A) were observed, showing typical GV morphological characteristics under scanning and transmission electron microscopy (EM). The genome of MyunGV-A was completely sequenced and analysed. The genome is 176,677 bp in size, with a G+C content of 39.79%. It contains 183 open reading frames (ORFs) encoding 50 or more amino acids with minimal overlap. Comparison of MyunGV-A with TnGV, XcGV, and HearGV genomes revealed extensive sequence similarity and collinearity, and the four genomes contain the same nine homologous regions (*hrs*) with conserved structures and locations. Three unique genes, 12 baculovirus repeated ORF (*bro*), 2 helicase, and 3 enhancer genes, were identified. In particular, two repeated genes (ORF39 and 49) are present in the genome, in reverse and complementarily orientations. Twenty-four OB proteins were identified from the putative protein database of MyunGV-A. In addition, MyunGV-A belongs to the *Betabaculovirus* group and is most closely related to TnGV (99% amino acid identity) according to a phylogenetic tree based on the combined amino acid sequences of 38 core gene contents.

Baculoviruses are a large family of rod-shaped, invertebrate-infecting viruses with large circular, covalently closed, double-stranded DNA genomes of between 80 and 180 kb. This family was initially taxonomically subdivided into nucleopolyhedroviruses (NPVs) or granuloviruses (GVs) based on viral occlusion morphology¹. However, when an increasing number of genome sequences became available, it was clear that lepidopteran NPVs and GV are more closely related to each other than to dipteran and hymenopteran NPVs. Therefore, a new taxonomic division that follows the evolution of the host more closely² was accepted by the International Committee on Taxonomy of Virus (ICTV). In the 10th report of the ICTV (online, 2019), the family *Baculoviridae* was still divided into four genera: *Alphabaculovirus*, *Betabaculovirus*, *Deltabaculovirus* and *Gammabaculovirus* (https://talk.ictvonline.org/ictv-reports/ictv_online_report/). To date, 85 baculovirus genomes have been sequenced (<http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=virus&taxid=10442>), including 55 from *Alphabaculovirus* (lepidopteran NPVs), 26 from *Betabaculovirus* (lepidopteran GV), 1 from *Deltabaculovirus* (dipteran NPVs) and 3 from *Gammabaculovirus* (hymenopteran NPVs).

Betabaculoviruses are granuloviruses (GVs) infecting only lepidopteran hosts, whereas alphabaculoviruses, deltabaculoviruses and gammabaculoviruses are nucleopolyhedroviruses (NPVs) isolated from a wider range of hosts, including lepidopterans, dipterans and hymenopterans.

Lepidopteran NPVs are further divided into two groups, I and II, based on gene content³. Notably, the budded virus (BV) fusion protein in Group I NPVs is GP64, whereas Group II NPVs lack *gp64* and utilize the F protein⁴. GV are classified into three types according to tissue tropism⁵. Type I GV, such as *Xestia c-nigrum* GV (XcGV), kill hosts at a slow speed by only infecting the midgut epithelium and fat body tissue⁶. Type II GV, such as *Cydia pomonella* GV (CpGV), kill hosts at a rapid speed, similar to typical lepidopteran NPVs, by infecting most of the host's major tissues⁷. Type III GV infect only the midgut epithelium. Only one GV, *Harrisina brillians* GV (HabrGV)⁸, has been identified as Type III. Phylogenetic analysis on the basis of conserved genes of GV does not show certain monophyletic origins for these different types of pathogenesis⁹.

¹Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China. ²Institute of Zoology, Chinese Academy of Sciences, Beijing, China. ³Jiangsu University of Science and Technology, Zhenjiang, China. ⁴These authors contributed equally: Yinü Li and Xingjian Liu. ✉email: qinql@ioz.ac.cn; bri-zhangzhifang@caas.cn

Mythimna unipuncta granulovirus (MyunGV-A), originally described as *Pseudaletia unipuncta* granulovirus (PsunGV) based on an isolated Hawaiian population of *Mythimna (Pseudaletia) unipuncta*¹⁰, was identified as PsunGV by the ICTV in 2002. Until 2017, PsunGV was proposed to be renamed MyunGV-A by the ICTV to reflect the fact that the new species MyunGV-B is the second distinct betabaculovirus to be isolated from the host *Mythimna (Pseudaletia) unipuncta*.

MyunGV-A (PsunGV-H) was first discovered by synergistic factors (described later as enhancin)¹⁰. Subsequent studies on MyunGV-A mostly focused on the mechanisms of enhancement and the enhancin gene. The enhancin of MyunGV-A can interact with viral particles and increase the binding of viral particles to insect midgut microvilli, thereby dramatically promoting the oral infectivity of *Mythimna unipuncta* NPV and decreasing the larval survival time¹¹. The enhancin of MyunGV-A comprising 901 amino acids have been purified and characterized¹². Overall, high-throughput sequencing of baculovirus genomes appears to be essential for analysing the molecular mechanisms of baculovirus infection and understanding baculovirus genome evolution. In this study, the morphological characteristics of MyunGV-A were observed by electron microscopy (EM). We present the complete sequence and organization of the MyunGV-A genome and compare it with other baculoviruses by genomic and phylogenetic analysis. A total of 24 OB proteins of MyunGV-A were identified.

Materials and methods

Virus preparation and DNA extraction. MyunGV-A (PsunGV-H) was obtained from Tanada Y. and kept at the Institute of Zoology, Chinese Academy of Sciences¹³. The virus was propagated in laboratory stocks of healthy second-instar *M. separate* larvae by per os infection. The occlusion bodies (OBs) produced in larval cadavers were purified by a standard method¹⁴.

To extract viral DNA, the purified OBs were resuspended in 0.1 M sodium carbonate solution [0.1 M Na₂CO₃, 0.17 M NaCl, 0.01 M EDTA (pH 10.5)] and incubated at 37 °C for 1 h. The pH was adjusted to 7.0 with 0.1 M HCl. Sarkosyl 0.5% and proteinase K 0.25 mg/mL were added to the sample and incubated at 37 °C for 2 h and 65 °C for 2 h. Genomic DNA was extracted with an equal volume of phenol and chloroform. The DNA was precipitated with two volumes of 100% ethanol, washed with 70% ethanol, and dissolved in TE buffer [10 mM Tris-HCl (pH 8); 1 mM EDTA].

Electron microscopy observation. OBs of MyunGV-A were observed by scanning electron microscopy (SEM; Hitachi S3400N) and transmission electron microscopy (TEM; JEOL JEM1230) according to standard methods¹⁵.

DNA sequencing and analysis. A random genomic library of MyunGV-A was constructed according to the “partial filling-in” method¹⁶. A total of 831 recombinant plasmids containing 1.5 to 5.0 kb viral DNA fragments were prepared for sequencing using a BigDye Terminator v3.1 (ABI) and a 3130XL Genetic analyser (ABI). The combined sequence generated from these clones represented sixfold genomic coverage. The gaps and ambiguities in the assembled sequence were resolved by PCR. All sequences were assembled into contigs using SeqMan from the DNASTAR 7.0 software package.

ORFs were defined using ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). The criterion for defining an ORF was a size of 50 or more codons with minimal overlap. DNA and protein comparisons were performed using BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). For protein homology detection, we used the HHpred webserver for the translated ORFs^{17,18}. Multiple alignments and percentage identities were obtained using ClustalW. Promoter motifs present upstream of the putative ORFs were screened as described previously¹⁹. Identity among homologous genes was determined with MegAlign software using ClustalW with default parameters. Homologous repeat regions (*hrs*) were analysed by Tandem Repeats Finder (<https://tandem.bu.edu/trf/trf.html>). GeneParityPlot analysis was performed as described by Hu et al.²⁰.

Protein analysis of OBs of MyunGV-A. Fresh purified OBs of MyunGV-A suspended in ddH₂O were incubated with an equal volume of lysis buffer (0.1 M Na₂CO₃, 0.17 M NaCl, 0.01 M EDTA, pH 10.6) at 4 °C for 1 h. The pH was adjusted to 8.0 with 0.1 M HCl. The samples were added to 10 mM Tris-HCl containing β-mercaptoethanol (0.2%) and sodium dodecyl sulfate (SDS) at 95 °C for 10 min. The proteins of MyunGV-A OBs were separated by SDS-PAGE using an 8% to 15% gradient gel. The protein bands were excised into 29 samples according to molecular weight from small to large for LC-MS/MS analysis (LCQ Deca Xp plus, ThermoFinnigan). LC-MS/MS analysis and protein identification were performed as described by Shi XF²¹. The raw files of MS spectra were searched against the putative protein database of MyunGV-A (NC_013772.1).

Phylogenetic analysis of MyunGV-A. The amino acid sequences encoded by the 38 core genes described for all members of family Baculoviridae²² of 82 complete baculovirus genomes (excluding 3 incomplete genomes) in the NCBI genome database (<https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=virus&taxid=10442>) were joined together according to a consistent order (ORF order of AcMNPV) and aligned using MAFFT with default parameters. A phylogenetic tree based on these sequences was constructed using MEGA 7 MEGA 7.0.14²³. Maximum likelihood (ML) tree construction methods were used with 1000 bootstrap resamples. The GTR + G + I substitution model was used for ML analysis.

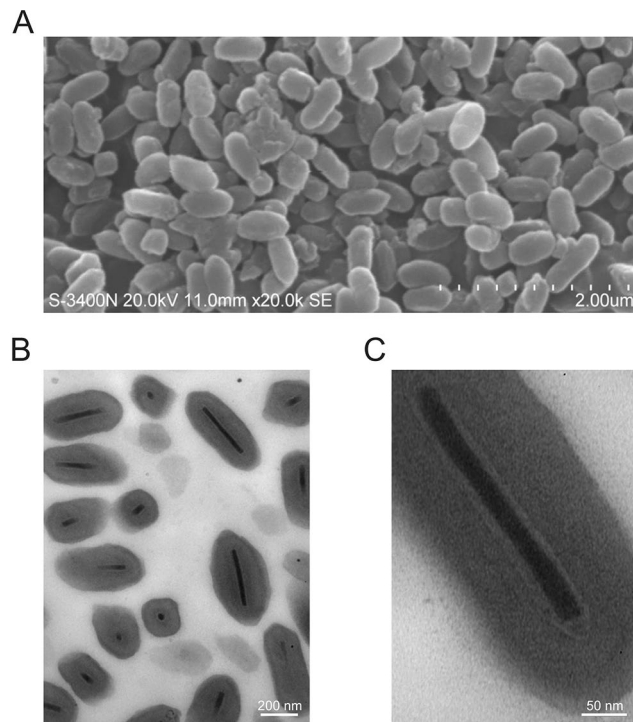


Figure 1. A scanning electron micrograph of MyunGV-A (A) and transmission electron micrograph of MyunGV-A (B,C).

Results and discussion

Electron microscopy observation. SEM revealed that the purified OBs of MyunGV-A have elongated ellipse shapes, with a length of approximately 0.5 μm and a width of approximately 0.3 μm (Fig. 1A). TEM showed a single rod-shaped ODV of approximately 300 nm in length and 40 nm in width embedded in a granular OB (Fig. 1B,C). These are typical GV morphological characteristics.

Sequence and genome characteristics of MyunGV-A. The size of the MyunGV-A genome is 176,677 bp (GenBank accession no. NC_013772), with a G+C content of 39.79%. MyunGV-A is the second largest GV sequenced to date, with XcGV (178,733 bp)⁶ being larger. Computer-assisted ORF analysis detected 372 ORFs of 50 or more codons and 9 homologous regions (*hrs*) in the MyunGV-A genome; 189 ORFs overlap significantly or are completely contained within other MyunGV-A ORFs. The deduced protein sequences of these 189 ORFs show no significant homology to protein sequences in GenBank. The remaining 183 ORFs and 9 *hrs* are shown in Table 1 according to location, orientation, size of the predicted amino acid sequence, potential baculovirus homologues, best matched baculovirus ORF and BLAST score (bits).

The first nucleotide of the granulin start codon was defined as nucleotide 1, and the ORF encoding granulin was accordingly designated as the first ORF. The putative ORFs were numbered sequentially in this orientation. Ninety-nine ORFs are in the granulin-sense orientation and 84 in the opposite orientation. A total of 183 putative ORFs of MyunGV-A were searched for promoter motifs at 180 bp upstream of the initiation codon of each ORF; only 42 were found to have a canonical baculovirus early gene promoter motif (a TATA box followed by a CAGT or CATT motif 20 to 40 bp downstream)^{24,25}. Seventy-five ORFs only possess a late promoter motif ((A/T/G) TAAG); 75 contain both early and late promoter motifs, which might allow transcription during both early and late stages of infection. Thirty-four lack any recognizable canonical promoter motif.

Comparison of MyunGV-A ORFs to other baculoviruses. Comparison of gene organization and homology between MyunGV-A and other baculovirus genomes provides insight into gene conservation and implications for the diversity of baculoviruses. MyunGV-A shares 88 ORFs with AcMNPV, 166 with XcGV, 169 with HearGV and TnGV, 139 with MyunGV-B and 104 with CpGV (Table 1). The average amino acid sequence identities of homologous ORFs between MyunGV-A and AcMNPV, XcGV, HearGV, TnGV, MyunGV-B and CpGV are 34%, 79%, 79%, 98%, 62% and 44%, respectively. A total of 180 ORFs were assigned a function or are homologous with other baculoviruses, of which three ORFs (68, 69 and 147) have homologues only with TnGV. ORF68 and ORF147 share 100% homology with TnGV but ORF69 94%. In addition, ORF69 has 37% homology with a kind of bacterium, *Zooshikella ganghwensis*. Three ORFs, ORF113, -133 and -166, were identified as unique to MyunGV-A.

ORF	Name	Position	Length (aa)	Promoter	Homologous ORF#/amino acid identity (%)					
					AcMNPV	XcGV	HearGV	TnGV	MyunGV-B	CpGV
1	Granulin	1>747	248	L	8/55	1/100	1/100	1/99	1/98	1/87
2	1629 capsid	796<1518	240	L	9/38	2/72	2/72	2/98	2/44	2/39
3	pk	1499>2353	284	L	10/33	3/91	3/92	3/100	3/70	3/47
4	Unknown	2399<3274	291	L	–	4/58	4/57	4/99	4/31	–
5	p10	3287>3646	119	L	137/33	5/90	5/92	5/100	5/65	–
6	Unknown	3682<4245	187	E	–	7/90	6/89	6/100	6/73, 7/50	4/35
7	Unknown	4235>4495	86	L	–	8/81	7/80	7/99	8/54	5/21
8	ie-1	4499<5923	474	E	147/27	9/77	8/76	8/100	9/58	7/28
9	Unknown	5946>6533	195	L	146/27	10/86	9/85	9/100	10/58	8/35
10	Unknown	6581<6880	99	L	145/39,150/44	11/97	10/96	10/100	11/83	9/53
11	odv-e18	6889<7140	83	L	143/44	12/90	11/90	–	12/81	14/61
12	p49	7144<8502	452	L	142/32	13/85	12/84	11/99	13/79	15/43
13	Unknown	8576<9265	229	E	–	14/88	13/88	12/99	14/60	–
14	odv-e56	9279<10,340	353	L	148/38	15/83	14/86	13/99	15/73	18/57
15	Unknown	10,370>10,576	68	–	–	16/81	15/78	14/100	16/58	19/35
16	pep	10,613<11,185	190	L	131/26	17/95	16/95	15/100	18/61	20/44
17	pep	11,262<11,723	153	L	–	18/93	17/92	16/99	19/84	23/53
18	pep/p10	11,744<12,907	387	E/L	–	19/94	18/94	17/100	20/80	22/49
19	Unknown	12,992>13,258	88	L	145/26,150/34	105/39,20/36	107/37,19/35	18/100	–	–
20	p94	13,339>15,777	812	E/L	134/36	21/70	20/71	19/99	–	–
21	Unknown	15,834<17,309	491	–	–	22/73	160/45	20/99	137/41	–
22	Unknown	17,390<18,463	357	–	–	23/55	22/56	21/99	23/44	–
23	Unknown	18,582<18,920	112	E/L	–	24/75	23/75	22/100	24/59	–
24	Unknown	20,373>21,686	437	E	–	25/61	24/60	23/99	26/24	29/56
25	Unknown	22,224>23,210	328	E/L	–	26/56	25/57	24/98	28/39	30/22
26	efp	23,278>25,023	581	L	23/20	27/80	26/79	25/99	29/52	31/32
27	Unknown	25,152>25,628	158	E	–	28/25	27/27	–	–	–
28	Unknown	26,027<26,737	236	L	–	29/80	28/78	26/100	31/55	33/27
29	Unknown	26,756<27,334	192	L	–	30/81	29/81	27/99	32/72	34/47
30	pif-3	27,364>27,948	194	L	115/41	32/91	30/92	28/100	33/66	35/46
31	Unknown	27,955>28,287	110	L	–	33/53	31/52	–	–	–
32	Unknown	28,313>28,654	113	L	–	34/99	32/99	29/100	35/85	39/49
33	lef-2	28,656>29,240	194	E/L	6/23	35/82	33/81	30/97	36/57	41/40
34	Cp35Ra	29,244>29,513	89	L	–	36/87	34/87	31/100	37/43	42/36
35	Unknown	29,549<29,959	136	L	–	–	–	32/100	–	–
36	Unknown	30,049<30,369	106	E	–	38/89	36/86	–	38/60	–
37	Unknown	30,494<30,952	152	E/L	–	39/75	37/74	33/100	39/45	45/20
38	mp-nase	31,036<32,817	593	L	–	40/81	38/77	34/99	40/53	46/31
39	Unknown	33,277>34,992	571	E	–	–	53/86,157/85	43/97	–	–
40	Unknown	35,052<36,023	323	–	–	42/60	39/58	–	41/25	–
41	p13	36,074>36,910	278	E/L	–	43/91	40/91	35/100	42/66	47/50
42	Unknown	37,028>37,621	197	E/L	–	44/59	41/61	36/100	–	–
43	pif-2	37,638>38,804	388	L	22/51	45/93	42/93	37/100	43/82	48/55
<i>hr1</i>	–	38,812–39,431	–	–	–	–	–	–	–	–
44	Unknown	39,435<39,677	80	L	–	46/76	43/71	38/99	44/38	49/33
45	Unknown	39,702>43,343	1213	L	–	47/82,48/51	44/52	39/99	45/43	50/29
46	Unknown	42,743<43,168	141	–	–	49/48	–	40/100	–	–
47	Unknown	43,330<44,112	260	L	106/44	50/84	45/83	41/100	46/68	52/53
48	pif-7	44,157>44,321	54	–	110/27	51/90	46/90	42/100	47/75	53/40
49	Unknown	44,456<46,171	571	E	–	–	53/86,157/85	43/97	–	–
50	Ubiquitin	46,360<46,593	77	L	35/76	52/96	47/96	44/100	48/95	54/82
51	odv-cc43	46,695>47,756	353	L	109/30	53/83	48/83	45/99	50/67	55/44
52	Unknown	47,783>48,115	110	L	108/37	54/92	49/94	46/100	51/76	56/29
53	39 k	48,181<49,077	298	L	36/25	55/89	50/90	47/100	52/63	57/24
54	lef-11	49,058<49,345	95	E	37/28	56/89	51/87	–	53/78	58/53

Continued

ORF	Name	Position	Length (aa)	Promoter	Homologous ORF#/amino acid identity (%)					
					AcMNPV	XcGV	HearGV	TnGV	MyunGV-B	CpGV
<i>hr2</i>	–	49,441–49,811	–	–	–	–	–	–	–	–
55	Unknown	49,955>50,803	282	E/L	–	57/72	52/71	48/100	57/55	–
56	<i>bro-a/bro-f</i>	50,984<52,480	498	E	–	60/28,131/30	54/31	49/100	–	63/29
57	Unknown	52,710<52,922	70	E	–	63/47	–	50/98	–	–
58	Unknown	53,074>53,775	233	E/L	–	151/50	–	–	64/74	–
59	Unknown	53,865>54,401	178	–	–	–	–	–	–	64/34
60	Unknown	54,541<56,619	692	E/L	–	64/64	59/64	51/64	–	–
61	Unknown	56,674<58,119	481	–	–	65/80	60/81	52/71	–	–
62	<i>bro</i>	58,220<58,879	219	E/L	2/31	62/45	58/44	53/40	21/42	–
63	<i>he65</i>	59,144<60,775	543	E	105/37	67/82	62/81	54/99	–	–
64	<i>sod</i>	60,936<61,397	153	L	31/57	68/89	63/89	55/99	58/74	59/59
65	<i>cath</i>	61,441<62,457	338	L	127/46	58/59	–	56/100	–	11/54
66	<i>bro-a/bro-f</i>	62,525<64,015	496	E	2/35	60/40,131/43	55/45	57/95	21/43	–
67	Unknown	64,111<64,281	56	L	–	70/73	64/73	58/100	–	–
<i>hr3</i>	–	64,322–64,816	–	–	–	–	–	–	–	–
68	Unknown	64,845<65,033	62	–	–	–	–	59/100	–	–
69	Unknown	65,205<65,459	84	E	–	–	–	60/94	–	–
70	Unknown	65,594>66,247	217	E	–	71/63	65/62	61/98	60/37	–
71	Unknown	66,293>67,327	344	E/L	–	72/79	66/79	62/99	–	–
72	Unknown	67,465>68,862	465	L	–	73/82	67/79	63/99	61/48	–
73	Unknown	68,948<70,021	357	L	–	74/69	68/70	64/99	65/40, 99/29	–
74	Unknown	70,125>70,412	95	L	79/38	75/87	69/86	65/100	66/63	65/36
<i>hr4</i>	–	70,464–71,122	–	–	–	–	–	–	–	–
75	Unknown	72,110<72,367	85	–	–	–	70/77	66/98	–	–
76	<i>bro-b</i>	72,366>73,178	270	E	2/23	76/84	71/83	67/99	67/61	–
77	<i>p74</i>	73,217>75,349	710	L	138/35	77/89	72/89	68/99	68/71	60/42
78	Unknown	75,346<75,681	111	L	–	–	73/68	69/99	–	–
79	<i>p47</i>	75,764>76,948	394	E	40/42	78/91	74/91	70/100	70/74	68/57
80	Rep-like	76,992<77,480	162	E/L	–	–	75/65	71/100	–	–
81	Rep-like	77,574<77,780	68	E	–	–	76/57	72/100	–	–
82	Unknown	78,087>78,764	225	L	38/42	79/97	77/97	73/100	71/85	69/64
83	<i>p24 capsid</i>	78,786>79,301	171	L	129/23	80/88	78/88	74/100	72/70	71/51
84	<i>p38.7</i>	79,338<79,916	192	E/L	13/21	81/73	79/76	75/99	73/46	73/42
85	<i>lef-1</i>	79,917<80,633	238	E	14/31	82/87	80/87	76/99	74/78	74/50
86	<i>p10</i>	80,710>81,291	193	L	–	83/79	81/77	77/100	75/51	–
87	<i>pif-1</i>	81,311>82,936	541	L	119/35	84/86	82/87	78/100	76/68	75/48
88	<i>fgf-1</i>	82,962<83,663	233	L	–	85/68	83/67	79/99	77/45	76/36
89	Unknown	83,713<84,060	115	E/L	–	86/71	84/71	–	78/43	–
90	Unknown	84,175>84,669	164	E/L	150/38	87/73	85/76	80/100	79/30, 80/59	79/29
91	<i>lef-6</i>	84,673<84,972	99	–	28/29	88/84	86/83	81/100	81/57	80/42
92	<i>dbp</i>	85,033<85,866	277	E	25/23	89/84	87/85	82/100	82/56	81/25
93	Unknown	85,972<86,184	70	E/L	–	–	88/82	83/100	83/61	–
94	Unknown	86,154<86,897	247	–	–	90/73	89/75	84/100	84/52	82/35
95	<i>p45</i>	86,896>88,014	372	L	103/36	91/95	90/96	85/100	85/82	83/55
96	<i>p12</i>	88,043>88,408	121	L	102/21	92/80	91/78	86/99	86/59	84/39
97	<i>p40</i>	88,460>89,572	370	E/L	101/20	93/91	92/92	87/100	87/76	85/50
98	<i>p6.9</i>	89,629>89,811	60	L	100/–	94/93	93/90	–	88/79	86/56
99	<i>lef-5</i>	89,847<90,647	266	L	99/39	95/90	94/89	88/100	89/74	87/56
100	<i>38 K</i>	90,570>91,481	303	L	98/37	96/84	95/85	89/99	90/67	88/48
101	Unknown	91,502<91,975	157	L	96/32	97/92	96/92	90/100	91/82	89/49
102	<i>Helicase-1</i>	91,974>95,450	1158	L	95/26	98/89	97/88	91/100	92/78	90/37
103	<i>odv-e25</i>	95,536<96,195	219	L	94/36	99/96	98/95	94/100	93/84	91/65
104	Unknown	96,234<96,710	158	L	93/33	100/94	99/96	95/99	94/66	92/40
105	<i>p33</i>	96,821>97,576	251	L	92/36	101/95	100/94	96/100	95/82	93/54
106	<i>ChaB</i>	97,582<97,845	87	L	60/67	102/95	103/82	97/100	96/80	–

Continued

ORF	Name	Position	Length (aa)	Promoter	Homologous ORF#/amino acid identity (%)					
					AcMNPV	XcGV	HearGV	TnGV	MyunGV-B	CpGV
107	Unknown	97,880<98,107	75	E/L	–	–	104/74	98/98	–	–
108	Chitinase	98,275>100,032	585	E/L	126/62	103/87	105/87	99/100	–	10/60
109	Unknown	100,414>100,755	113	E/L	–	106/68	108/70	100/100	–	–
110	gp37	100,824>101,570	248	E/L	64/45	107/82	109/83	101/99	–	13/44
111	Unknown	101,679>102,152	157	E	–	108/78	110/84	102/100	–	–
112	<i>bro-c</i>	102,340>103,416	358	E	2/20	109/76	101/68	103/98	–	–
113	Unknown	103,455>103,679	74	–	–	–	–	–	–	–
114	<i>lef-4</i>	103,733<105,091	452	–	90/31	110/86	112/86	104/99	100/63	95/42
115	p39 capsid	105,143>106,126	327	L	89/28	111/80	113/80	105/99	101/77	96/38
116	<i>odv-cc27</i>	106,272>107,138	288	L	144/29	112/94	114/95	106/100	102/81	97/45
117	Unknown	107,471<108,685	404	E	–	113/81	116/81	107/100	103/54	99/27
<i>hr5</i>	–	108,728–109,096	–	–	–	–	–	–	–	–
118	<i>bro-d</i>	109,118<110,389	423	–	2/23	114/84	117/86	108/100	104/68	–
119	Unknown	110,368>111,543	391	–	–	115/80	118/79	109/100	105/45	–
120	Unknown	111,652>112,023	123	E/L	–	116/92	119/92	110/100	106/66	100/42
121	Unknown	112,066<112,614	182	E	–	117/73	120/73	111/100	107/39	–
122	<i>vp91</i>	112,702<114,957	751	L	83/29	118/79	121/80	112/99	108/58	101/35
<i>hr5a</i>	–	113,797–113,923	–	–	–	–	–	–	–	–
123	<i>tlp20</i>	114,923>115,417	164	L	82/23	119/85	122/86	113/100	109/63	102/25
124	Unknown	115,438>116,001	187	L	81/47	120/96	123/96	114/100	110/80	103/57
125	<i>gp41</i>	116,058>116,930	290	E/L	80/32	121/91	124/91	115/100	111/71	104/51
126	Unknown	117,000>117,323	107	L	78/25	122/84	125/84	116/100	112/52	105/29
127	<i>vlf-1</i>	117,307>118,422	371	E/L	77/32	123/89	126/89	117/100	113/75	106/55
128	Unknown	118,437<118,973	178	E	–	124/84	127/84	118/100	114/64	–
129	Unknown	119,015>119,272	85	L	76/34	125/97	128/98	119/100	115/88	107/58
130	Unknown	119,344>119,781	145	E/L	75/29	126/93	129/93	120/100	116/54	108/35
131	Unknown	119,833>120,150	105	L	150/29	20/33,105/31	130/36	121/99	–	–
132	Unknown	120,192<120,629	145	E	–	128/71	131/71	122/100	–	–
133	Unknown	120,702<121,781	359	–	–	–	–	–	–	–
134	<i>lef-7</i>	121,991>122,959	322	E	–	129/56	132/57	123/95	–	–
135	<i>bro-a/f</i>	123,207>123,932	241	E	2/30	60/56,131/50	133/51,54/47	127/81	21/51	–
136	<i>bro-a/f</i>	124,050>124,505	151	E/L	–	60/30,131/34	54/69,133/30	125/58	–	–
137	<i>bro-a/f</i>	124,555>126,084	509	E/L	2/27	60/70,131/49	54/67	127/82	21/48	–
138	<i>dna pol</i>	126,186<129,470	1094	E	65/34	132/86	134/86	128/99	117/75	111/52
139	Desmoplakin	129,469>131,460	663	–	66/29	133/68	135/68	129/99	118/47	112/50
140	<i>lef-3</i>	131,553<132,590	345	E	67/28	134/64	136/65	130/99	119/48	113/25
141	<i>pif-6</i>	132,559>132,969	136	–	68/34	135/91	137/91	131/100	120/79	114/44
142	Unknown	133,028>133,543	171	–	–	136/71	138/70	132/100	121/43	115/33
143	<i>iap</i>	133,631>134,503	290	–	–	137/82	139/81	133/99	122/53	116/31
144	Unknown	134,625>136,646	673	E/L	–	138/26	–	134/99	–	–
145	<i>lef-9</i>	136,758>138,251	497	–	62/56	139/93	140/93	135/100	123/81	117/62
146	<i>fp</i>	138,304>138,516	70	E/L	61/31	140/91	141/92	136/100	124/83	118/45
147	Unknown	138,550>138,702	50	L	–	–	–	137/100	–	–
148	DNA ligase	138,757<140,415	552	E	–	141/89	142/89	138/100	125/71	120/43
149	Unknown	140,592>140,831	79	E	–	142/81	143/80	139/100	126/62	121/27
150	Unknown	140,892>141,092	66	–	–	143/95	144/95	140/100	127/81	122/56
151	<i>fgf</i>	141,155<142,369	404	E	32/27	144/74	145/73	141/99	128/46	123/28
152	<i>alk-exo</i>	142,339>143,760	473	–	133/34	145/78	146/77	142/100	129/61	125/41
153	Helicase-2	143,830>145,203	457	L	–	146/85	147/87	143/99	130/71	126/49
154	Unknown	145,314>146,315	333	E	112/30,113/40	147/80	148/79	144/99	131/61	–
155	<i>lef-8</i>	146,356<148,938	860	L	50/48	148/93	149/93	145/99	132/79	131/61
156	<i>odv-e66</i>	149,009<151,012	667	L	46/40	149/92	150/92	146/99	133/79	37/44
<i>hr6</i>	–	151,030–151,402	–	–	–	–	–	–	–	–
157	Enhancin-1	151,411<153,897	828	L	–	150/74	151/74	147/99	134/47	–
158	<i>bro-f</i>	154,072<155,493	473	E	–	131/66	133/67	148/94	21/67	–

Continued

ORF	Name	Position	Length (aa)	Promoter	Homologous ORF#/amino acid identity (%)					
					AcMNPV	XcGV	HearGV	TnGV	MyunGV-B	CpGV
159	Enhancin-3	155,673>158,378	901	L	–	154/80	153/80	149/99	135/35	–
160	Unknown	158,419>159,585	388	L	–	155/89	154/90	150/99	–	–
161	Unknown	159,833>160,012	59	–	–	157/94	155/91	151/96	–	–
162	<i>bro-g</i>	160,230<161,075	281	–	2/27	159/56	158/50	–	–	–
163	Unknown	161,223<161,435	70	E	111/42	160/70	–	152/86	136/63	–
164	Unknown	161,623>163,113	496	E/L	–	161/77	160/76	153/100	137/51	–
165	Unknown	163,143<163,706	187	E	–	162/72	161/72	154/97	138/40	–
<i>hr7</i>	–	163,796–164,430	–	–	–	–	–	–	–	–
166	Unknown	164,387>164,587	66	–	–	–	–	–	–	–
167	<i>bro-g</i>	164,594>165,649	351	–	–	159/64	159/82	155/92	–	–
168	Unknown	165,772<165,939	55	–	–	–	162/83	156/100	139/56	–
169	Unknown	165,865>166,221	118	–	–	165/91	163/94	157/100	–	132/35
170	Enhancin-4	166,267<168,840	857	L	–	166/78	164/79	158/99	140/55	–
171	Unknown	169,109>169,438	109	L	–	167/71	165/71	159/98	141/55	–
<i>hr8</i>	–	169,581–169,956	–	–	–	–	–	–	–	–
172	Unknown	169,958<170,146	62	E	–	170/87	168/87	161/100	142/38	133/53
173	Unknown	170,133>170,552	139	–	53/35	171/88	169/87	162/100	143/56	134/42
174	Unknown	170,556<171,677	373	L	–	172/71	170/72	163/100	144/50	135/43
175	Unknown	171,705<171,908	67	L	–	173/86	171/86	164/100	145/72	–
176	<i>lef-10</i>	171,886>172,098	70	L	53a/31	174/95	172/94	165/100	146/78	137/34
177	<i>vp1054</i>	171,977>172,948	323	–	54/31	175/92	173/91	166/100	147/78	138/49
178	Unknown	173,035>173,217	60	E/L	–	176/91	174/91	167/100	148/72	–
179	Unknown	173,331>173,666	111	E	–	177/73	175/72	168/100	149/42	–
180	<i>fgf-1</i>	173,710>174,633	307	–	–	178/69	176/68	169/98	150/52	140/29
181	Unknown	174,743>175,375	210	E	–	179/75	177/75	170/100	151/45	–
182	<i>me53</i>	175,416>176,330	304	E	139/20	180/88	178/87	171/100	152/63	143/34
183	Unknown	176,336>176,659	107	–	–	181/85	179/84	172/100	153/63	–

Table 1. MyunGV-A (PsunGV-H) open reading frames (ORFs) and homologous repeat regions (*hrs*). Putative MyunGV-A ORFs are listed in column 1 along with the gene homologues designated in column 2. Column 3 indicates ORF location and transcriptional direction on the MyunGV-A genome. Column 4 indicates the number of amino acids. Column 5 indicates the presences of early (E) and/or late (L) promoters located upstream of start codon of each ORF. E indicates a TATA sequence followed by a CAGT or CATT mRNA start site sequence 20–40 nucleotides downstream, with 180 bp upstream of the start codon. L indicates the presence of a (A/T/G)TAAG sequence. Column 6–11 list the homologous ORF and percent of amino acid identity from AcMNPV, XcGV, HearGV, TnGV, MyunGV-B and CpGV respectively.

GeneParityPlot analysis. The gene order of MyunGV-A was compared with that of AcMNPV, XcGV, HearGV, TnGV, MyunGV-B and CpGV by GnenParityPlots analysis (Fig. 2)²⁰. The gene organization of MyunGV-A is distinctly different from that of AcMNPV, except for two reverse collinear gene clusters in which one is a 12-gene group including the core gene cluster of four genes, *lef-5*, 38K(ac98), *ac96*, and helicase, with relative positions that are conserved in baculovirus genomes²⁶. In contrast, the gene order of MyunGV-A exhibits extensive collinearity with XcGV, HearGV, TnGV, MyunGV-B and CpGV, except for several genes in a different order that are almost *bro* or near *bro*, with the highest collinearity to TnGV. Interestingly, the arrangement of the MyunGV-A genome shows lower collinearity to MyunGV-B, a virus from the same host, than to XcGV, HearGV and TnGV.

Homologous regions (*hrs*). A typical feature of most baculovirus genomes is the presence of homologous regions (*hrs*) interspersed throughout the genome. The numbers of *hrs* in 82 complete baculovirus genomes range from none to 17, with 12 baculovirus genomes lacking typical *hrs* sequences (Table S1). In general, *hrs* are characterized by AT-rich and imperfect, reiterated palindromic sequences that may be replaced with direct repeats.

Eight major *hr* sequences (*hr1-8*) and one short *hr* sequence (*hr5a*) were identified in the MyunGV-A genome (Table 1). *hr1-8* contains two to five direct imperfect repeats, each of approximately 120 bp, whereas *hr5a* does not contain multiple repeated sequences. It is interesting to note that *hr5a* is located in ORF122 (*vp91*), and the same situation exists in the XcGV and HearGV genomes. Six *hrs* were identified in the MyunGV-B genome lacking sequences corresponding to *hr1* and *hr5/5a* of MyunGV-A²⁷. No *hrs* were found in the TnGV genome deposited in 2018 (NC_038375.1), and there is no publication on the analysis of the sequence.

Although the nucleotide sequences of repeats vary between each *hr*, even in the same *hr*, two highly conserved 10 bp core sequences (TTAAT (G/A) TCGA) were found at the roughly same positions (approximately 35 bp)

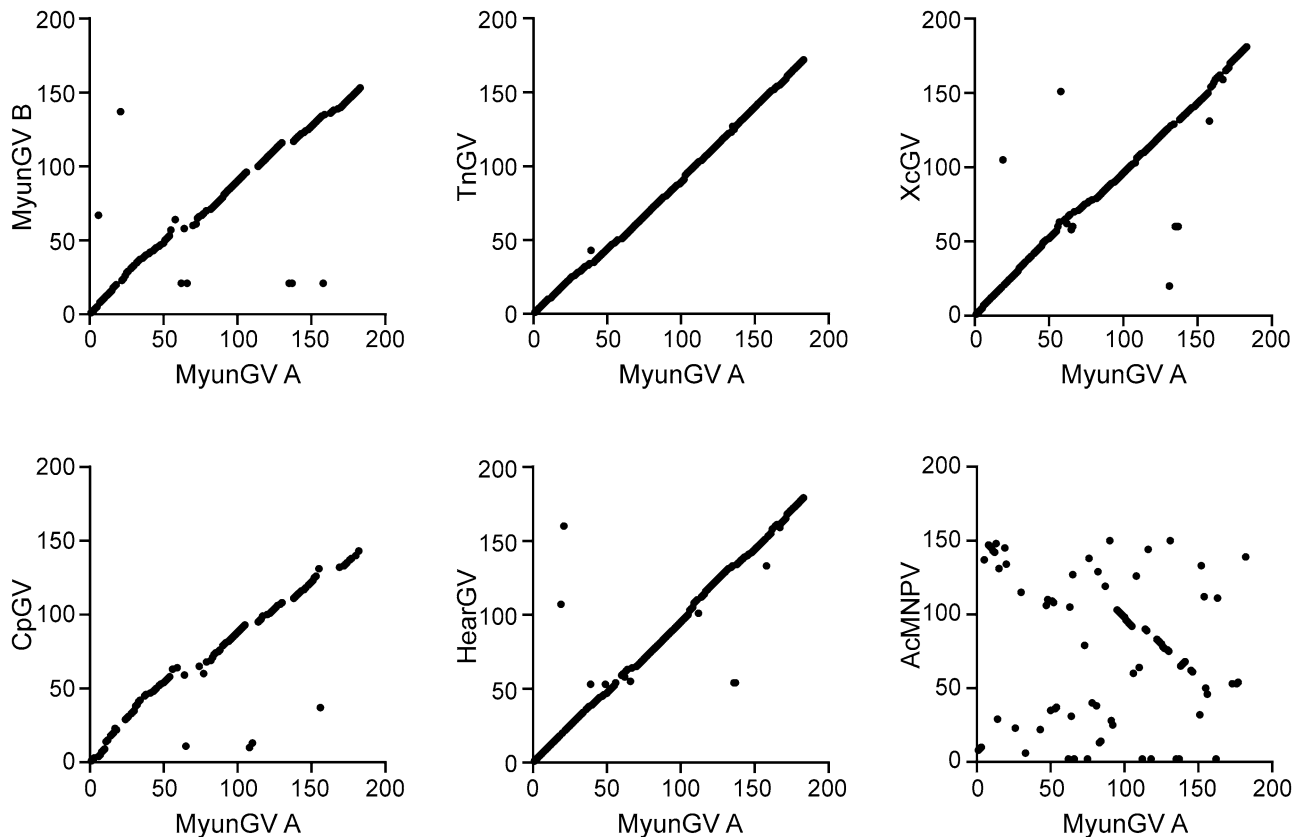


Figure 2. GeneParityPlots analysis.

of each repeat⁶. In the MyunGV-A genome, the core sequences in each repeat of *hr1*, -2, -4, -7 and -8 are in the same directions, while those of *hr3*, -5, -5a and -6 are in opposite directions (Fig. 3).

Hrs have been reported to function in replication origins^{28,29} and serve as enhancers of transcription of early genes³⁰. In addition, the number of *hrs* is connected to the replication efficiency or pathogenicity of a baculovirus. Deletion of one to five *hrs* of AcMNPV had little or no effect on virus infection, while deleting six or seven *hrs* resulted in 90% BV reduction. Deletion of all eight *hrs* caused 99.9% BV reduction and delay of early and late gene expression but did not completely inhibit virus production³¹.

Baculovirus repeated ORFs (bro genes). *Bro* genes have been identified in most baculovirus genomes sequenced to date. The number of *bro* genes in different baculovirus genomes varies considerably. Thirteen of 82 complete baculovirus genomes have only one *bro* gene, though *Lymantria dispar* MNPV (LdMNPV) has 16 *bro* genes. *Bro* genes are entirely absent from 19 baculovirus genomes sequenced to date (Table S1).

In MyunGV-A, 12 *bro* genes were identified, of which 3 adjacent *bro* genes (ORF135, -136, -137) were found. BLAST results of amino acid sequences of these 3 *bro* genes in NCBI showed that ORF135 best matches with TnGV ORF127 (81%), ORF136 with HearGV ORF54 (69%), and ORF137 with TnGV ORF127 (82%). The TnGV genome has 4 adjacent *bro* genes (ORF124, -125, -126, -127), and the HearGV genome has 3 pairs of adjacent *bro* genes (hear54 and -55, hear101 and -102, hear158 and -159), but no adjacent *bro* genes were found in the XcGV and MyunGV-B genomes.

The exact function of *bro* genes is not yet clear, though their presence is very significant for baculoviruses. Studies on the function of *bro* genes have mostly focused on BmNPV and have found that BRO-A and C proteins can bind to DNA in infected cells³²; BRO-A may be involved in influencing host DNA replication, similar to a laminin-binding protein³³.

In addition, BmNPV BRO proteins act as nucleocytoplasmic shuttling proteins via the CRM1-mediated nuclear export pathway³⁴. Recently, BmNPV BRO-B and E proteins associated with host T-cell intracellular antigen 1 homologue (BmTRN-1) were shown to be involved in the inhibitory regulation of certain mRNAs at the post-transcriptional level during infection³⁵. The function of other baculovirus BRO proteins has seldom been reported.

Two repeat genes in MyunGV-A. Two repeat genes (ORF39 and ORF49), with amino acid sequence identities of 100%, were found in the MyunGV-A genome; the former is in the granulinsense orientation and the latter in the opposite orientation.

There is no homologous gene with these two genes in the XcGV, MyunGV-B and CpGV genomes. Indeed, only one gene, ORF43, of the TnGV genome matches with them, and the amino acid sequence identity is 97%.

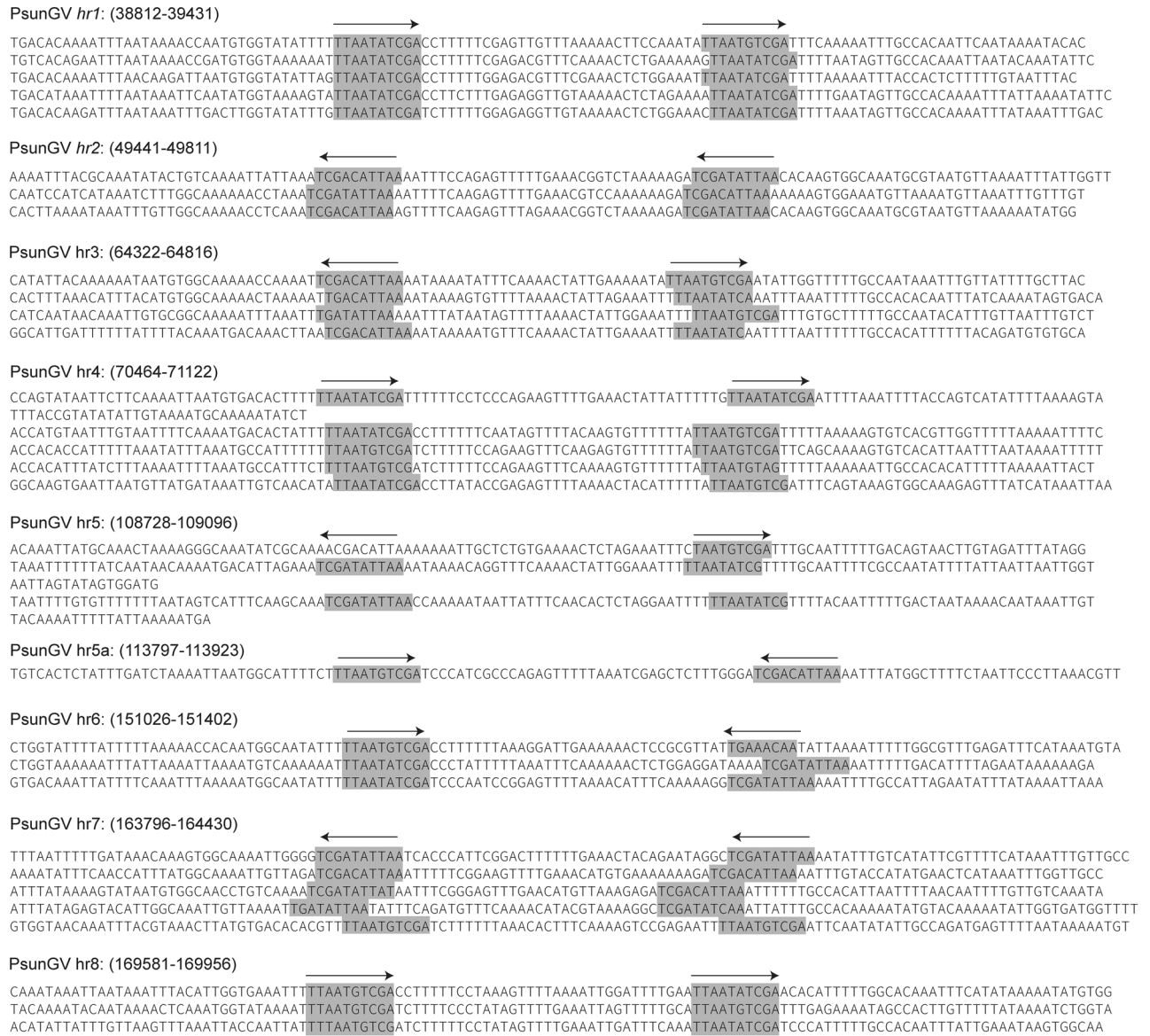


Figure 3. Alignment of homologous regions in the MyunGV-A genome. The conserved 10 bp core sequences (TTAATG/ATCGA) are indicated by shaded boxed. The arrows indicate direction of core sequences.

Two genes, ORF53 and ORF157, in the HearGV genome are homologous, with amino acid sequence identities of 86% and 85%, respectively, and the amino acid sequence identity of ORF53 and ORF157 in the HearGV genome is 99%. One gene with two copies in one baculovirus genome was found in other baculovirus genomes, such as odv-e66, p26 and dbp of EcobNPV³⁶ and odv-e66 and p26 of SfMNPV³⁷.

BLAST results of amino acid sequences of these two homologous genes in MyunGV-A in NCBI suggested they match *hr3* and *hr4* of *Heliothis virescens* ascovirus 3e (amino acid sequence identities both 49%). In addition, they match the 70.4-kDa C-terminal Zn-finger DNA-binding domain of *Spodoptera frugiperda* ascovirus 1a (amino acid sequence identities of 48%), which suggests that their function may be associated with DNA binding.

ORFs with no homologues in other baculoviruses. Three ORFs, including ORF113, -133 and -166, were identified as having no homologues in other baculoviruses (Table 1). These three unique ORFs have no recognizable promoter. Protein homology analysis using HHpred showed that GP133 (aa 50–359) is a likely homologue of Mannan-binding lectin serine peptidase 1 (probability, 99.97%; E value, 1.1e-28). Mannan-binding lectin serine peptidase 1 plays a central role in the initiation of the complement lectin pathway³⁸. This homology indicates that ORF133 might be related to the complement lectin pathway, which deserves further research. ORF113 encodes an 8.5-kDa protein with one transmembrane domain (aa 5–27, analysed by TMHMM server v2.0) at the N terminus of the protein with no similarity to any proteins in the nonredundant protein database. ORF166 encodes a 7.7-kDa protein with no similarity to any proteins in the nonredundant protein database.

The large gene in MyunGV-A. In most cases, *helicase* is the largest gene in baculovirus genomes; however, in the MyunGV-A genome, ORF45 encoding 1213 amino acids (longer than *helicase-1*, 1158) is the largest gene. Similar situations are present in the HearGV (ORF44, 1279 aa), TnGV (ORF39, 1213 aa) and MyunGV-B (ORF45, 1507 aa) genomes, though it is divided into two genes, ORF47 and ORF48, in XcGV⁶. Compared with XcGV, the MyunGV-A genome has an additional adenosine (A) at position 40315, resulting in a reading frame shift. Protein homology analysis using HHpred and SWISS-MODEL showed no significant similarity to any other known sequences for Myun45.

Enhancins in MyunGV-A. It was first observed in *Mythimna* (formerly *Pseudaletia*) *unipuncta* that GV can increase the rate of infection and fatality of NPV and decrease the larval survival time when GV and NPV coinfect larvae¹⁰. Subsequent studies found that the factor responsible for synergistic interaction is a GV protein that shows a synergistic effect only when larvae are infected with NPV; it was identified as a synergistic factor (SF)³⁹. The synergistic effect of viral enhancing factor (VEF) was also observed in TnGV⁴⁰. The location and sequence of the VEF gene of TnGV have been identified⁴¹. This enhancing protein (enhancin) can disrupt the midgut peritrophic membrane (PM), thereby resulting in the more efficient passage of virions to host midgut cells¹². Enhancin was identified as a metalloprotease via the discovery of a zinc-binding site as well as by inhibition with a metal chelator and reactivation with divalent ions⁴².

The MyunGV-A genome has three enhancin genes (Myun157, -159 and -170). Similarly, three enhancin genes were found in MyunGV-B and TnGV, but they show large diversity in amino acid sequence identity compared to MyunGV-A. MyunGV-B enhancins are only 35% to 55% identical to that of MyunGV-A but are as high as 99% identical to that of TnGV. Four enhancin genes were found in the XcGV and HearGV genomes, of which enhancin-1, -3, and -4 have high homology (amino acid sequence identities all above 74%) to three enhancin genes of MyunGV-A. The MyunGV-A enhancin gene (*enhancin-3*) encoding 901 amino acids has been sequenced and characterized¹². The canonical sequence HEXXH, the zinc-binding site in most metalloproteases, was found in enhancing-3 but not in the other two enhancins. It is not clear why three enhancins are present in MyunGV-A, and the roles of these three enhancins in promoting NPV infection remain unclear.

Enhancins are found mainly in GVs and a few NPVs. They are localized within the granulin matrix in granuloviruses and released to increase virus pathogenicity by acting in the midgut. In contrast, LdMNPV enhancins are located within ODV envelopes and facilitate ODVs to pass the host defence barrier by acting directly on the peritrophic membrane as the nucleocapsids move through the barrier⁴³. However, subsequent studies have indicated that LdMNPV enhancins have a function that may assist virus-host cell fusion beyond peritrophic membrane degradation⁴⁴.

Protein analysis of OBs of MyunGV-A. To date, nine baculovirus proteomic studies have been performed with the intent of revealing infectious mechanisms and virus-host interactions, as follows: six alphabaculoviruses—AcMNPV^{45,46}, BmNPV⁴⁷, HearSNPV⁴⁸, HearNPV-G4⁴⁹, AgMNPV⁵⁰ and ChchNPV⁵¹; two betabaculoviruses, ClanGV⁵² and PrGV⁵³; and one deltabaculovirus, CuniNPV⁵⁴. In this study, we performed an analysis of MyunGV-A OB proteins. For 29 samples, 24 proteins were identified from the putative protein database of MyunGV-A (NC_013772.1) (Table 2). Among the 24 proteins, 20 were detected with two or more peptides, and the other four were detected with one matching peptide. In addition, 15 of 24 identified proteins were detected in more than one sample. Granulin was found in 28 of the 29 samples (Table S2). The same situations were found for CuniNPV⁵⁴, HearsNPV⁴⁸ and AgMNPV⁵⁰. A noticeable phenomenon was also observed, whereby the identified proteins were not distributed according to their molecular mass in SDS-PAGE gels. The reason was postulated to be incomplete denaturation of OBs and the breakdown of protein complexes or protein processing⁵⁴.

Of the 24 identified proteins, eight are encoded by core genes, including ORF11 (ODV-e18), ORF12 (VP49), ORF14 (ODV-e56), ORF48 (PIF-7), ORF51 (ODV-ec43), ORF103 (ODV-e25), ORF115 (VP39) and ORF125 (GP41); among them, VP39 is the major capsid protein, GP41 is a tegument protein only found in ODVs and is present in the nucleocapsid and the viral envelope as a structural protein of ODVs, and four proteins, including ODV-e18, ODV-e56, ODV-e25 and ODV-ec43, are ODV envelope proteins (Table 2). An ODV envelope protein, ODV-e66, was also identified.

For the 24 identified proteins, six are encoded by additional genes conserved in GVs, including ORF16, ORF17, ORF18, ORF120, ORF174 and ORF175⁵⁵. Among them, proteins encoded by two contiguous ORFs (ORF16 and 17) belong to the CpGV ORF16 L family⁵⁶, and the protein encoded by ORF18 is similar to P10, containing a baculovirus polyhedron envelope protein (PEP) C domain (pfam04513). In addition to structural proteins or those implicated in DNA replication and transcription, four important auxiliary proteins were identified, including SOD, cathepsin and two enhancins. Enhancin-1 and enhancin-3 were detected in our proteomic studies; enhancin-3 was present in 16 samples, while enhancin-1 was present in only 1 sample. Most baculovirus enhancins, including MyunGV-A, are located in the OB matrix, whereas LdMNPV enhancins were found to be associated with ODV envelopes^{43,57}. In this study, we did not attempt to determine the specific location of enhancins.

Moreover, four proteins (Myun29, Myun32, Myun44 and Myun67) with unknown functions were detected (Table 2). An increasing number of baculovirus proteomic studies can provide valuable insight into baculovirus structure, infectious mechanisms and interactions with their hosts.

ORF	Protein	ODV of GV		ODV of NPV					Characteristics/function
		PrGV	ClanGV	AcMNPV	HearNPV	AgMNPV	ChchNPV	CuniNPV	
1 (Ac8)	Granulin	1	1	8	1	1	1	1	Occlusion bodies (OBs) matrix protein
11 (Ac143)	ODV-e18	14	13	143	10	139	12	NA	Core gene; Structural protein of ODV envelope
12 (Ac142)	VP49	15	14	142	9	138	11	30	Core gene; a caspase inhibitor; Inhibiting diverse apoptotic stimuli
14 (Ac148)	ODV-e56	16	15	148	15	144	7	102	Core gene; Structural protein of ODV envelope
16 (Cp20)	PEP-1	20	19	NA	120	127	NA	NA	Additional ORFs conserved in GVs ; ORF16L family
17 (Cp23)	PEP-2	22	36	NA	NA	NA	-	NA	Additional ORFs conserved in GVs ; ORF16L family
18 (Cp22)	PEP/P10	21	35	NA	21	133	-	NA	Additional ORFs conserved in GVs ; Similarities to P10 containing a baculovirus PEP C domain
29 (Cp34)	Unknown	29	NA	NA	NA	NA	NA	NA	Unknown
32 (Cp39)	Unknown	-	39	NA	NA	NA	NA	NA	Additional ORFs conserved in GVs
44 (Cp49)	Unknown	NA	NA	NA	NA	NA	NA	NA	Unknown
48 (Ac110)	PIF-7	-	20	-	NA	NA	NA	NA	Unknown
51 (Ac109)	ODV-ec43	46	44	109	94	107	99	69	Core gene; Structural protein of ODV envelope
64 (Ac31)	SOD	-	48	-	106	-	115	NA	Metalloenzyme; Protecting the virus against superoxide radical induced in the environment by sunlight
65 (Ac127)	Cathepsin	-	11	127	-	NA	-	NA	Associated with liquefaction of insects at the end of infection; Promoting the release and spread of progeny virus
67 (Xc70)	Unknown	NA	NA	NA	NA	NA	NA	NA	Unknown
103 (Ac94)	ODV-e25	-	76	94	82	91	86	NA	Core gene; Structural protein of ODV envelope
115 (Ac89)	VP39	81	81	89	78	86	82	24	Core gene; Major capsid protein
120 (Cp100)	Unknown	84	84	NA	NA	NA	NA	NA	Additional ORFs conserved in GVs
125 (Ac80)	GP41	88	89	80	73	79	78	33	Core gene; O-linked glycosylated ODV protein; BV production
156 (Ac40)	ODV-e66	39/44	NA	46	96	114	101	NA	Structural protein of ODV envelope
157 (Xc150)	Enhancin-1	NA	NA	NA	NA	NA	NA	NA	Metalloproteinases; Enhancing the oral infectivity of NPVs
159 (Xc154)	Enhancin-3	NA	NA	NA	NA	NA	NA	NA	Metalloproteinases; Enhancing the oral infectivity of NPVs
174 (Cp135)	Unknown	114	116	NA	NA	NA	NA	NA	Additional ORFs conserved in GVs
175 (Cp136)	Unknown	-	117	NA	NA	NA	NA	NA	Additional ORFs conserved in GVs

Table 2. Analysis of proteins identified from MyunGV-A. Dash (-), the protein was not detected. NA, ORF was not found in the baculovirus genome by BLASTP.

Phylogenetic analysis of MyunGV-A. A phylogenetic tree based on the combined amino acid sequences of 38 core genes from 82 complete baculovirus genomes (Table S1) classified MyunGV-A into clade “a” of *Betabaculovirus*, which clusters infecting the larvae of the Lepidopteran family Noctuidae. Within this clade, MyunGV-A is present into a subcluster together with TnGV, the closest neighbour, sharing a common hypothetical ancestor. XcGV and HearGV form another subcluster next to the MyunGV-A and TnGV subclusters. However, MyunGV-B, another granulovirus from the same host, groups into a subcluster with SpfrGV and slightly away from MyunGV-A across MolaGV (Fig. 4). This is consistent with the above comparison results of gene organization in which MyunGV-A is similar to TnGV, XcGV and HearGV, regardless of genome size, ORF number or gene order.

Conclusion

The purified OBs of MyunGV-A show typical GV morphological characteristics under EM. The complete MyunGV-A (NC_013772.1) genome is 176,677 bases, with a G+C content of 39.79%, the second largest baculovirus genome to date. It contains 183 ORFs with a minimal size of 50 codons. The genome of MyunGV-A exhibits extensive sequence similarity and collinearity with TnGV, XcGV and HearGV. Three unique genes, 12 bro, 2 helicase and 3 enhancin genes, were identified. In particular, two repeated genes (ORF39 and 49) are present in the genome in reverse and complementarily orientations. Twenty-four OB proteins were identified from the putative protein database of MyunGV-A. According to our phylogenetic tree, MyunGV-A belongs to the *Betabaculovirus* group and is most closely related to TnGV.

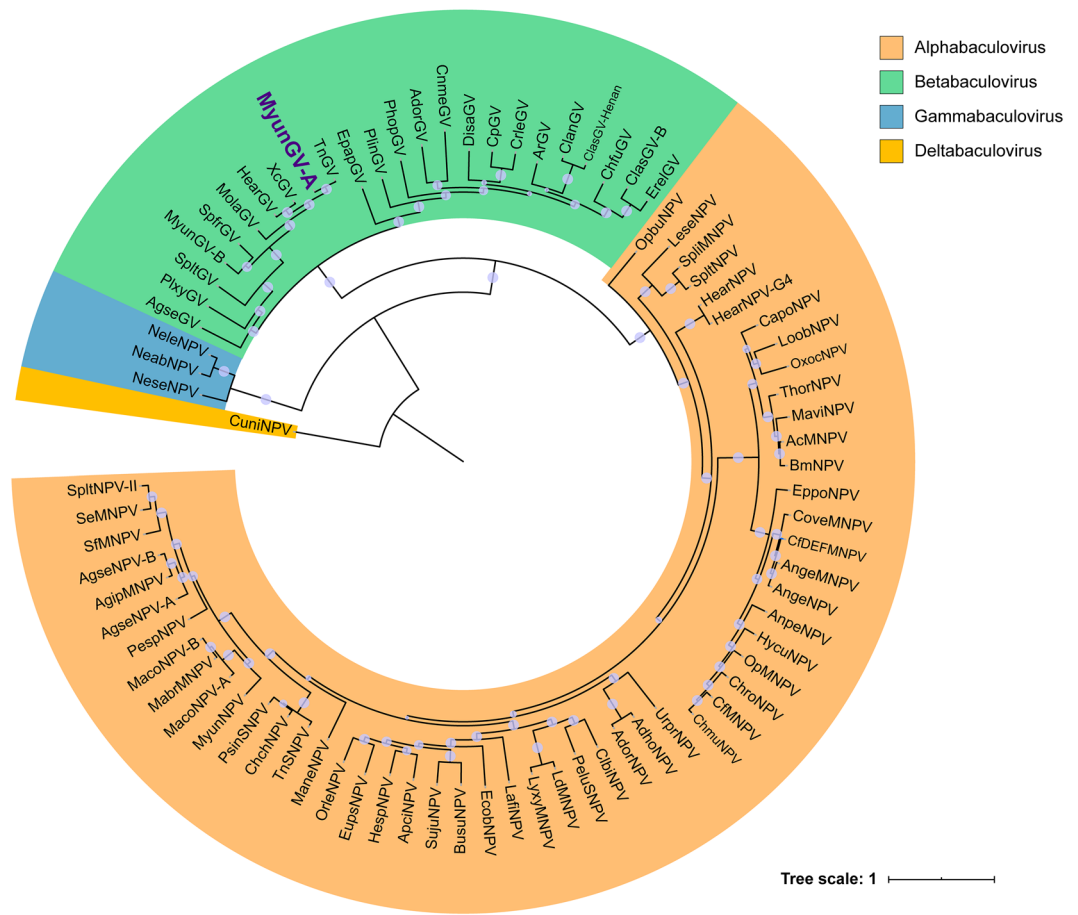


Figure 4. Phylogenetic tree of 82 baculoviruses with complete sequences. The phylogenetic tree was generated using MEGA X⁵⁸ software and performed with the maximum likelihood method and JTT matrix-based model⁵⁹. The result was visualized using iTOL⁶⁰.

Received: 11 February 2020; Accepted: 17 December 2020

Published online: 11 January 2021

References

- Murphy, F. *et al.* Virus taxonomy: Sixth report of the international committee on taxonomy of viruses. *Arch. Virol.* **10**, 350–354 (1995).
- Jehle, J. A. *et al.* On the classification and nomenclature of baculoviruses: A proposal for revision. *Arch. Virol.* **151**, 1257–1266. <https://doi.org/10.1007/s00705-006-0763-6> (2006).
- Zanotto, P. M., Kessing, B. D. & Maruniak, J. E. Phylogenetic interrelationships among baculoviruses: Evolutionary rates and host associations. *J. Invertebr. Pathol.* **62**, 147–164. <https://doi.org/10.1006/jipa.1993.1090> (1993).
- Pearson, M. N. & Rohrmann, G. F. Transfer, incorporation, and substitution of envelope fusion proteins among members of the Baculoviridae, Orthomyxoviridae, and Metaviridae (insect retrovirus) families. *J. Virol.* **76**, 5301–5304. <https://doi.org/10.1128/jvi.76.11.5301-5304.2002> (2002).
- Federici, B. A. *The Baculoviruses*. *The Viruses* 33–59 (Springer, New York, 1997).
- Hayakawa, T. *et al.* Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* **262**, 277–297. <https://doi.org/10.1006/viro.1999.9894> (1999).
- Luque, T., Finch, R., Crook, N., O'Reilly, D. R. & Winstanley, D. The complete sequence of the *Cydia pomonella* granulovirus genome. *J. Gen. Virol.* **82**, 2531–2547. <https://doi.org/10.1099/0022-1317-82-10-2531> (2001).
- Federici, B. A. & Stern, V. M. Replication and occlusion of a granulosis virus in larval and adult midgut epithelium of the western grapeleaf skeletonizer, *Harrisina brillians*. *J. Invertebr. Pathol.* **56**, 401–414 (1990).
- Jehle, J. A. *et al.* Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology* **346**, 180–193. <https://doi.org/10.1016/j.virol.2005.10.032> (2006).
- Tanada, Y. Synergism between two viruses of the armyworm, *Pseudaletia unipuncta* (Haworth) (Lepidoptera, Noctuidae). *J. Insect Pathol.* **1**, 215–231 (1959).
- Tanada, Y. A synopsis of studies on the synergistic property of an insect baculovirus: A tribute to Edward A. Steinhaus. *J. Invertebr. Pathol.* **45**, 125–138 (1985).
- Roelvink, P. W., Corsaro, B. G. & Granados, R. R. Characterization of the *Helicoverpa armigera* and *Pseudaletia unipuncta* granulovirus enhancin genes. *J. Gen. Virol.* **76**(Pt 11), 2693–2705. <https://doi.org/10.1099/0022-1317-76-11-2693> (1995).
- Tsuey, D., Ta, D. & Xiuyu, T. Enhancement of baculovirus infection by the synergistic factor of a granulosis virus of the armyworm. *Acta Entomol. Sin.* **4**, 1–10 (1995).

14. O'Reilly, D. R., Miller, L. K. & Luckow, V. A. *Baculovirus Expression Vectors: A Laboratory Manual* (W.H. Freeman & Company, New York, 1992).
15. Tang, P. *et al.* Genomic sequencing and analyses of HearMNPV: A new Multinucleocapsid nucleopolyhedrovirus isolated from *Helicoverpa armigera*. *Virology* **9**, 168. <https://doi.org/10.1186/1743-422X-9-168> (2012).
16. Chen, Y., Lin, X., Yi, Y., Lu, Y. & Zhang, Z. Construction and application of a baculovirus genomic library. *Z. Naturforsch. C* **64**, 574–580 (2009).
17. Zimmermann, L. *et al.* A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007> (2018).
18. Soding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–248. <https://doi.org/10.1093/nar/gki408> (2005).
19. Kool, M. & Vlaskovits, J. M. The structural and functional organization of the *Autographa californica* nuclear polyhedrosis virus genome. *Arch. Virol.* **130**, 1–16 (1993).
20. Hu, Z. H. *et al.* Distinct gene arrangement in the Buzura suppressaria single-nucleocapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **79**(Pt 11), 2841–2851. <https://doi.org/10.1099/0022-1317-79-11-2841> (1998).
21. Shi, X. F., Li, Y. N., Yi, Y. Z., Xiao, X. G. & Zhang, Z. F. Identification and characterization of 30 K protein genes found in *Bombyx mori* (Lepidoptera: Bombycidae) transcriptome. *J. Insect Sci.* **15**, 71. <https://doi.org/10.1093/jisesa/iev057> (2015).
22. Javed, M. A. *et al.* *Autographa californica* multiple nucleopolyhedrovirus AC83 is a per os infectivity factor (PIF) protein required for occlusion-derived virus (ODV) and budded virus nucleocapsid assembly as well as assembly of the PIF complex in ODV envelopes. *J. Virol.* **91**, 5. <https://doi.org/10.1128/JVI.02115-16> (2017).
23. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054> (2016).
24. Blissard, G. W., Kogan, P. H., Wei, R. & Rohrmann, G. F. A synthetic early promoter from a baculovirus: Roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. *Virology* **190**, 783–793 (1992).
25. Kogan, P. H., Chen, X. & Blissard, G. W. Overlapping TATA-dependent and TATA-independent early promoter activities in the baculovirus gp64 envelope fusion protein gene. *J. Virol.* **69**, 1452–1461 (1995).
26. Herniou, E. A., Olszewski, J. A., Cory, J. S. & O'Reilly, D. R. The genome sequence and evolution of baculoviruses. *Annu. Rev. Entomol.* **48**, 211–234. <https://doi.org/10.1146/annurev.ento.48.091801.112756> (2003).
27. Harrison, R. L. *et al.* The complete genome sequence of a second distinct betabaculovirus from the true armyworm, *Mythimna unipuncta*. *PLoS ONE* **12**, e0170510. <https://doi.org/10.1371/journal.pone.0170510> (2017).
28. Hayakawa, T., Rohrmann, G. F. & Hashimoto, Y. Patterns of genome organization and content in lepidopteran baculoviruses. *Virology* **278**, 1–12. <https://doi.org/10.1006/viro.2000.0668> (2000).
29. Hilton, S. & Winstanley, D. Identification and functional analysis of the origins of DNA replication in the *Cydia pomonella* granulovirus genome. *J. Gen. Virol.* **88**, 1496–1504. <https://doi.org/10.1099/vir.0.82760-0> (2007).
30. Theilmann, D. A. & Stewart, S. Tandemly repeated sequence at the 3' end of the IE-2 gene of the baculovirus *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus is an enhancer element. *Virology* **187**, 97–106 (1992).
31. Bossert, M. & Carstens, E. B. Sequential deletion of AcMNPV homologous regions leads to reductions in budded virus production and late protein expression. *Virus Res.* **256**, 125–133. <https://doi.org/10.1016/j.virusres.2018.08.010> (2018).
32. Zemskov, E. A., Kang, W. & Maeda, S. Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. *J. Virol.* **74**, 6784–6789 (2000).
33. Kang, W. K. *et al.* Interaction of *Bombyx mori* nucleopolyhedrovirus BRO-A and host cell protein laminin. *Arch. Virol.* **148**, 99–113. <https://doi.org/10.1007/s00705-002-0902-7> (2003).
34. Kang, W., Kurihara, M. & Matsumoto, S. The BRO proteins of *Bombyx mori* nucleopolyhedrovirus are nucleocytoplasmic shuttling proteins that utilize the CRM1-mediated nuclear export pathway. *Virology* **350**, 184–191. <https://doi.org/10.1016/j.viro.2006.01.008> (2006).
35. Kotani, E., Muto, S., Ijiri, H. & Mori, H. *Bombyx mori* nucleopolyhedrovirus nucleic acid binding proteins BRO-B and BRO-E associate with host T-cell intracellular antigen 1 homologue BmTRN-1 to influence protein synthesis during infection. *J. Gen. Virol.* **96**, 1947–1956. <https://doi.org/10.1099/vir.0.000136> (2015).
36. Ma, X. C. *et al.* Genome sequence and organization of a nucleopolyhedrovirus that infects the tea looper caterpillar, *Ectropis oblique*. *Virology* **360**, 235–246. <https://doi.org/10.1016/j.viro.2006.10.024> (2007).
37. Harrison, R. L., Puttler, B. & Popham, H. J. Genomic sequence analysis of a fast-killing isolate of *Spodoptera frugiperda* multiple nucleopolyhedrovirus. *J. Gen. Virol.* **89**, 775–790. <https://doi.org/10.1099/vir.0.83566-0> (2008).
38. Megyeri, M. *et al.* Quantitative characterization of the activation steps of mannan-binding lectin (MBL)-associated serine proteases (MASPs) points to the central role of MASP-1 in the initiation of the complement lectin pathway. *J. Biol. Chem.* **288**, 8922–8934. <https://doi.org/10.1074/jbc.M112.446500> (2013).
39. Tanada, Y., Himeno, M. & Omi, E. M. Isolation of a factor, from the capsule of a granulosis virus, synergistic for a nuclear-polyhedrosis virus of the armyworm. *J. Invertebr. Pathol.* **21**, 31–40 (1973).
40. Derksen, A. C. & Granados, R. R. Alteration of a lepidopteran peritrophic membrane by baculoviruses and enhancement of viral infectivity. *Virology* **167**, 242–250 (1988).
41. Hashimoto, Y., Corsaro, B. G. & Granados, R. R. Location and nucleotide sequence of the gene encoding the viral enhancing factor of the *Trichoplusia ni* granulosis virus. *J. Gen. Virol.* **72**(Pt 11), 2645–2651. <https://doi.org/10.1099/0022-1317-72-11-2645> (1991).
42. Lepore, L. S., Roelink, P. R. & Granados, R. R. Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. *J. Invertebr. Pathol.* **68**, 131–140. <https://doi.org/10.1006/jipa.1996.0070> (1996).
43. Slavicek, J. M. & Popham, H. J. The *Lymantria dispar* nucleopolyhedrovirus enhancins are components of occlusion-derived virus. *J. Virol.* **79**, 10578–10588. <https://doi.org/10.1128/JVI.79.16.10578-10588.2005> (2005).
44. Hoover, K., Humphries, M. A., Gendron, A. R. & Slavicek, J. M. Impact of viral enhancin genes on potency of *Lymantria dispar* multiple nucleopolyhedrovirus in *L. dispar* following disruption of the peritrophic matrix. *J. Invertebr. Pathol.* **104**, 150–152. <https://doi.org/10.1016/j.jip.2010.02.008> (2010).
45. Wang, R. *et al.* Proteomics of the *Autographa californica* nucleopolyhedrovirus budded virions. *J. Virol.* **84**, 7233–7242. <https://doi.org/10.1128/JVI.00040-10> (2010).
46. Braunagel, S. C., Russell, W. K., Rosas-Acosta, G., Russell, D. H. & Summers, M. D. Determination of the protein composition of the occlusion-derived virus of *Autographa californica* nucleopolyhedrovirus. *Proc. Natl. Acad. Sci. USA.* **100**, 9797–9802. <https://doi.org/10.1073/pnas.1733972100> (2003).
47. Liu, X., Chen, K., Cai, K. & Yao, Q. Determination of protein composition and host-derived proteins of *Bombyx mori* nucleopolyhedrovirus by 2-dimensional electrophoresis and mass spectrometry. *Intervirology* **51**, 369–376. <https://doi.org/10.1159/000193462> (2008).
48. Deng, F. *et al.* Proteomics analysis of *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus identified two new occlusion-derived virus-associated proteins, HA44 and HA100. *J. Virol.* **81**, 9377–9385. <https://doi.org/10.1128/JVI.00632-07> (2007).
49. Hou, D. *et al.* Comparative proteomics reveal fundamental structural and functional differences between the two progeny phenotypes of a baculovirus. *J. Virol.* **87**, 829–839. <https://doi.org/10.1128/JVI.02329-12> (2013).
50. Braconi, C. T. *et al.* Proteomic analyses of baculovirus *Anticarsia gemmatalis* multiple nucleopolyhedrovirus budded and occluded virus. *J. Gen. Virol.* **95**, 980–989. <https://doi.org/10.1099/vir.0.061127-0> (2014).

51. Xu, F., Ince, I. A., Boeren, S., Vlaskovits, J. M. & van Oers, M. M. Protein composition of the occlusion derived virus of *Chrysodeixis chalcites* nucleopolyhedrovirus. *Virus Res.* **158**, 1–7. <https://doi.org/10.1016/j.virusres.2011.02.014> (2011).
52. Zhang, X., Liang, Z., Yin, X. & Shao, X. Proteomic analysis of the occlusion-derived virus of *Clostera anachoreta* granulovirus. *J. Gen. Virol.* **96**, 2394–2404. <https://doi.org/10.1099/vir.0.000146> (2015).
53. Wang, X. F. *et al.* ODV-associated proteins of the *Pieris rapae* granulovirus. *J. Proteome Res.* **10**, 2817–2827. <https://doi.org/10.1021/pr2000804> (2011).
54. Perera, O., Green, T. B., Stevens, S. M. Jr., White, S. & Becnel, J. J. Proteins associated with *Culex nigripalpus* nucleopolyhedrovirus occluded virions. *J. Virol.* **81**, 4585–4590. <https://doi.org/10.1128/JVI.02391-06> (2007).
55. van Oers, M. M. & Vlaskovits, J. M. Baculovirus genomics. *Curr. Drug Targets* **8**, 1051–1068 (2007).
56. Kang, W., Crook, N. E., Winstanley, D. & O'Reilly, D. R. Complete sequence and transposon mutagenesis of the BamHI J fragment of *Cydia pomonella* granulosis virus. *Virus Genes* **14**, 131–136 (1997).
57. Yamamoto, T. & Tanada, Y. Physicochemical properties and location of capsule components, in particular the synergistic factor, in the occlusion body of a granulosis virus of the armyworm, *Pseudaletia unipuncta*. *Virology* **107**, 434–440. [https://doi.org/10.1016/0042-6822\(80\)90310-4](https://doi.org/10.1016/0042-6822(80)90310-4) (1980).
58. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. <https://doi.org/10.1093/molbev/msy096> (2018).
59. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282. <https://doi.org/10.1093/bioinformatics/8.3.275> (1992).
60. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259. <https://doi.org/10.1093/nar/gkz239> (2019).

Acknowledgements

This work was supported by National Natural Sciences Foundation of China (No. 31872430 and 31670156), The State Key Laboratory of Integrated Management of Pest Insects and Rodents (Grant No. IPM1902) and The Agricultural Science and Technology Innovation Program.

Author contributions

Z.Z. and Q.Q. conceived the idea; Y.L. and Z.Z. designed research; Y.L., X.L., P.T., and H.Z. performed research; Y.L. and X.L. analyzed data; Y.L. and X.L. wrote the main manuscript text; All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80117-3>.

Correspondence and requests for materials should be addressed to Q.Q. or Z.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021