Journal of General Virology (2006), **87**, 537–551 **DOI** 10.1099/vir.0.81461-0

Genome sequence of an enhancin gene-rich nucleopolyhedrovirus (NPV) from *Agrotis segetum*: collinearity with *Spodoptera exigua* multiple NPV

Agata K. Jakubowska,[1,2] Sander A. Peters,[3] Jadwiga Ziemnicka,[1] Just M. Vlak[2] and Monique M. van Oers[2]

[1]Department of Biological Control and Quarantine, Institute of Plant Protection, Miczurina 20, Poznan 60-318, Poland

[2]Laboratory of Virology, Wageningen University, Binnenhaven 11, 6709 PD Wageningen, The Netherlands

[3]Greenomics, Plant Research International, PO Box 16, 6700 AA Wageningen, The Netherlands

Correspondence
Monique M. van Oers
monique.vanoers@wur.nl

Received 25 August 2005
Accepted 22 November 2005

The genome sequence of a Polish isolate of *Agrotis segetum* nucleopolyhedrovirus (AgseNPV-A) was determined and analysed. The circular genome is composed of 147 544 bp and has a G+C content of 45·7 mol%. It contains 153 putative, non-overlapping open reading frames (ORFs) encoding predicted proteins of more than 50 aa, together making up 89·8 % of the genome. The remaining 10·2 % of the DNA constitutes non-coding regions and homologous-repeat regions. One hundred and forty-three AgseNPV-A ORFs are homologues of previously reported baculovirus gene sequences. There are ten unique ORFs and they account for 3 % of the genome in total. All 62 lepidopteran baculovirus genes, including the 29 core baculovirus genes, were found in the AgseNPV-A genome. The gene content and gene order of AgseNPV-A are most similar to those of *Spodoptera exigua* (Se) multiple NPV and their shared homologous genes are 100 % collinear. Three putative enhancin genes were identified in the AgseNPV-A genome. In phylogenetic analysis, the AgseNPV-A enhancins form a cluster separated from enhancins of the *Mamestra* species NPVs.

INTRODUCTION

The turnip moth, *Agrotis segetum* Dennis & Schiffenmüller (Lepidoptera, Noctuidae), is an important pest of many crops in Europe, Asia and Africa. The larvae feed on many vegetable and field crops, including corn, rape, beet, potatoes, cabbage, cereals, tobacco, vine and many others (Ignoffo & Garcia, 1979). *A. segetum* is often part of a cutworm complex, further consisting of *Agrotis ipsilon*, *Agrotis exclamationis*, *Agrotis subterranea*, *Peridroma saucia* and other noctuid species, that destroys plants near the soil surface (Bourner & Cory, 2004). Viruses with biocontrol potential in the field, belonging to the family *Baculoviridae*, have been isolated from various *Agrotis* species (Caballero *et al.*, 1991; Bourner *et al.*, 1992; Boughton *et al.*, 1999). Baculoviruses form a family of large, circular, double-stranded DNA viruses that occur widely in invertebrates, in particular insect species of the orders Lepidoptera, Diptera and Hymenoptera. To date, over 700 baculoviruses have been isolated and several of these have been investigated as bioinsecticides of phytophagous insects (Federici, 1999) and showed potential against field and forest pests all over the world.

The family *Baculoviridae* is divided into two genera, *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV) (Theilmann *et al.*, 2005). Lepidopteran NPVs show a further division into group I and group II NPVs (Bulach *et al.*, 1999; Herniou *et al.*, 2001). Group I NPVs appear to be much more conserved than those of group II (Hyink *et al.*, 2002; Lange & Jehle, 2003). NPVs are designated single (S) or multiple (M) depending on the number of nucleocapsids packaged in a virion, although this feature has no taxonomic value (Volkman *et al.*, 1995). Twenty-eight completely sequenced baculovirus genomes have been released in GenBank to date (25 from lepidopteran viruses, two from hymenopteran viruses and one from a dipteran virus), including a GV isolated from *A. segetum* (AgseGV) (GenBank accession no. NC_005839). Comparison of these baculovirus genomes revealed 29 core genes, shared by all baculoviruses, and 62 genes common to lepidopteran baculoviruses (Herniou *et al.*, 2003; Lange & Jehle, 2003). Apart from those conserved genes, baculoviruses contain genes shared by a variable number of related virus species or even contain unique genes.

The GenBank/EMBL/DDBJ accession number for the sequence described in this paper is DQ123841.

Two different NPVs have been isolated from the turnip moth *A. segetum*, an English/French and a Polish isolate, both of the multiple nucleocapsid (M) type (Allaway & Payne, 1983). Restriction-enzyme profiling and phylogenetic analysis based on three conserved baculovirus genes – polyhedrin (*polh*), late expression factor 8 (*lef-8*) and *per os* infectivity factor 2 (*pif-2*) – revealed that both isolates are relatively distant and probably represent different virus species (Jakubowska *et al.*, 2005).

In this paper, the genome sequence of the Polish isolate of *A. segetum* NPV is analysed. To avoid nomenclatural confusion, we propose to name this Polish isolate AgseNPV-A, as it is the first AgseNPV isolate whose genome has been sequenced, and hence propose the name AgseNPV-B for the English/French isolate. The Polish isolate was indicated before as AgseNPV-P (Jakubowska *et al.*, 2005). AgseNPV-A is compared in genome sequence and organization with *Spodoptera exigua* (Se) MNPV (IJkel *et al.*, 1999), *Autographa californica* (Ac) MNPV (Ayres *et al.*, 1994), *Mamestra configurata* (Maco) NPV-B (Li *et al.*, 2002a) and the recently sequenced GV from *A. segetum* (AgseGV) (GenBank accession no. NC_005839). The gene order of AgseNPV-A was found to be highly collinear with that of SeMNPV, despite relatively low amino acid identity values (60 % on average).

METHODS

Virus and insects. AgseNPV-A was isolated in 1975 from *A. segetum* larvae and stored at −20 °C. The virus was freshly amplified in the laboratory in second-instar larvae of *A. segetum*. Larvae were infected individually with diet discs contaminated with 10 μl virus suspension at a concentration of 10^5 occlusion bodies (OBs) ml^{-1}.

DNA isolation. Polyhedra were isolated from infected insects. Dead larvae were homogenized by using a glass homogenizer and filtered through four layers of cheesecloth. The filtrate was layered onto a 30 % (w/w) sucrose cushion and centrifuged for 15 min at 5300 r.p.m. at 4 °C. The pellet containing polyhedra was resuspended and the sucrose purification was repeated two times. The final pellet was washed three times in distilled water and finally resuspended in 1 ml water. Polyhedra were dissolved by incubation for 30 min at 37 °C in 0·1 M sodium carbonate (final pH, approx. 11). Large debris was removed by 5 min centrifugation at 1000 r.p.m. and the supernatant was centrifuged for 30 min at 14 000 r.p.m. to pellet the occluded virions. DNA was isolated according to Reed *et al.* (2003) and dialysed for 24 h at 4 °C against 0·1 × TE buffer (10 mM Tris/HCl; 1 mM EDTA; pH 7·5). Quantity and quality of isolated DNA were determined spectrophotometrically and by electrophoresis in 0·7 % agarose.

Nucleotide sequence determination. The full genome sequence was determined by shotgun cloning of 10 μg sheared DNA of 1–1·5 kbp. The DNA fragments were cloned into pBluescript II SK(+) (Stratagene) by using *Escherichia coli* XL2 Blue ultracompetent cells (Stratagene). End-in sequencing was performed on a 3730xL DNA analyser (Applied Biosystems) and a 3100 Genetic Analyser (Applied Biosystems); sequences were assembled with Gap4 from the Staden-Solaris-1-5-3 software package and then checked in detail manually (van Oers *et al.*, 2005). In total, 1 517 059 nt were determined, with a mean redundancy of 9·88.

Sequence analysis. Genes were located with GeneMark software (Borodovsky & McIninch, 1993) and ORF Finder (NCBI). All open reading frames (ORFs) with a minimal size of 150 nt (50 aa), which did not overlap for major parts with other ORFs, were analysed. In addition, the genome was checked in detail for the presence of any ORFs identified for SeMNPV (IJkel *et al.*, 1999) or any other baculovirus in GenBank. Homology searches were performed by using BLAST (Altschul *et al.*, 1990).

To easily compare sequence information from different baculovirus genomes, the GECCO program was exploited. GECCO is a gene content-comparison tool able to align large numbers of sequences quickly by using the standard NCBI BLAST (van Oers *et al.*, 2005). The percentages of identity of all AgseNPV-A ORFs with their homologues in selected genomes were calculated for complete ORFs by using CLUSTAL_X (Thompson *et al.*, 1997). To detect homologous regions, DOTPLOT analysis (DNAstar) and the EMBOSS program (http://bioweb.pasteur.fr/seqanal/EMBOSS/) were applied under various stringency conditions. Pairwise gene-order analysis was performed by making gene-parity plots as described previously (Hu *et al.*, 1998). In this analysis, both shared and non-shared genes were included.

Phylogenetic analysis. For phylogenetic analysis of enhancin genes, protein sequences were aligned in CLUSTAL_X. Maximum-parsimony (MP) analysis was performed by using PAUP* (Swofford, 2003). Bootstrap analyses were performed to evaluate the robustness of the phylogenies using 1000 replicates. Branch lengths were calculated by using the neighbour-joining (NJ) method. The enhancin sequences used for this analysis were found by using the AgseNPV-A enhancin sequences as queries for the BLAST link at NCBI.

Cross-infectivity. Second-instar *S. exigua* and *A. segetum* larvae were individually fed high doses (10^8 OBs) of AgseNPV-A and SeMNPV, respectively. The larvae were first starved for 16 h and then fed virus-contaminated diet plugs. After consumption of the entire contaminated diet plug, they were given fresh diet and reared separately until death or pupation. The larvae were incubated at 25 °C, in a relative humidity of 70 % and a 16 : 8 h day : night photoperiod. Fifty larvae of each species were tested.

RESULTS AND DISCUSSION

Genome features

The AgseNPV-A genome is a circular, double-stranded DNA molecule containing 147 544 bp, in line with the predicted size of 148 kbp based on restriction-enzyme analysis (Jakubowska *et al.*, 2005). The computationally derived *Hin*dIII restriction map is in agreement with the experimentally constructed map from restriction-enzyme analysis, with the exception of an additional submolar band in the latter, indicating that more than one genetic variant is present in the isolate (Jakubowska *et al.*, 2005). This is confirmed by the presence of several single-base polymorphisms in the sequence. The G+C content of the AgseNPV-A genome is 45·7 mol%, which is slightly higher than that of most sequenced baculoviruses. Only *Choristoneura fumiferana* (Cf) MNPV, *Orgyia pseudotsugata* (Op) MNPV, *Lymantria dispar* (Ld) MNPV and *Culex nigripalpus* (Cuni) NPV have G+C contents higher than that of AgseNPV-A (Ahrens *et al.*, 1997; Kuzio *et al.*, 1999; Afonso *et al.*, 2001; de Jong *et al.*, 2005).

Gene content and genome organization

Using computational analysis, 419 methionine-initiated ORFs of more than 50 aa were initially identified. The

maximal acceptable overlap was set at 20 aa, with the exception of overlapping genes showing significant similarity to ORFs known in other baculoviruses. From those ORFs, 153 ORFs were assigned as AgseNPV-A ORFs, after elimination of ORFs located within larger ORFs and without similarity to baculovirus ORFs (Fig. 1, Table 1). The ORF density is comparable to that of other NPVs, when calculated as no. ORFs/genome size. According to the adopted convention (Vlak & Smith, 1982), polyhedrin was designated gene number 1 (Agse1) and the adenine of the start codon of the polyhedrin gene was assigned as the first nucleotide of the AgseNPV-A circular genome. Of the 153 AgseNPV-A ORFs, 143 have an assigned function or homologues in other baculoviruses. Ten AgseNPV-A ORFs have no homologues in baculoviruses and thus are considered unique to AgseNPV-A until homologues in other baculoviruses are found. The total length of those unique ORFs is 4379 nt (3 % of the genome). Four of these have no consensus promoter and may be non-functional. Predicted proteins are encoded by 89·8 % of the genome; the rest forms non-coding regions and homologous regions (*hr*s). Five *hr*s were identified in the AgseNPV-A genome dispersed around the genome (see below); the maximum distance between two *hr*s is 57 kbp.

All 29 core baculovirus genes and all 62 lepidopteran baculovirus genes (Herniou *et al.*, 2003) were found in the AgseNPV-A genome. There are 21 pairs of overlapping genes along the genome, eight of which have overlaps longer than 20 nt. All of them showed significant similarity to known baculovirus ORFs and thus were assigned as AgseNPV-A ORFs. The maximal overlap of 139 nt exists between *vp1054* and *lef-10* (Agse113/114). Both genes are oriented anticlockwise and both have homologues in SeMNPV. However, a promoter motif was not found upstream of the AgseNPV-A *lef-10* gene. In general, the AgseNPV-A genome is densely packed, with minimal intergenic distances. There are three gene clusters packed very tightly: ORFs 66–70, 95–100 and 109–115. Fifty-five per cent of the ORFs are directed clockwise and 45 % anticlockwise, with respect to the orientation of transcription of the polyhedrin gene (Vlak & Smith, 1982).

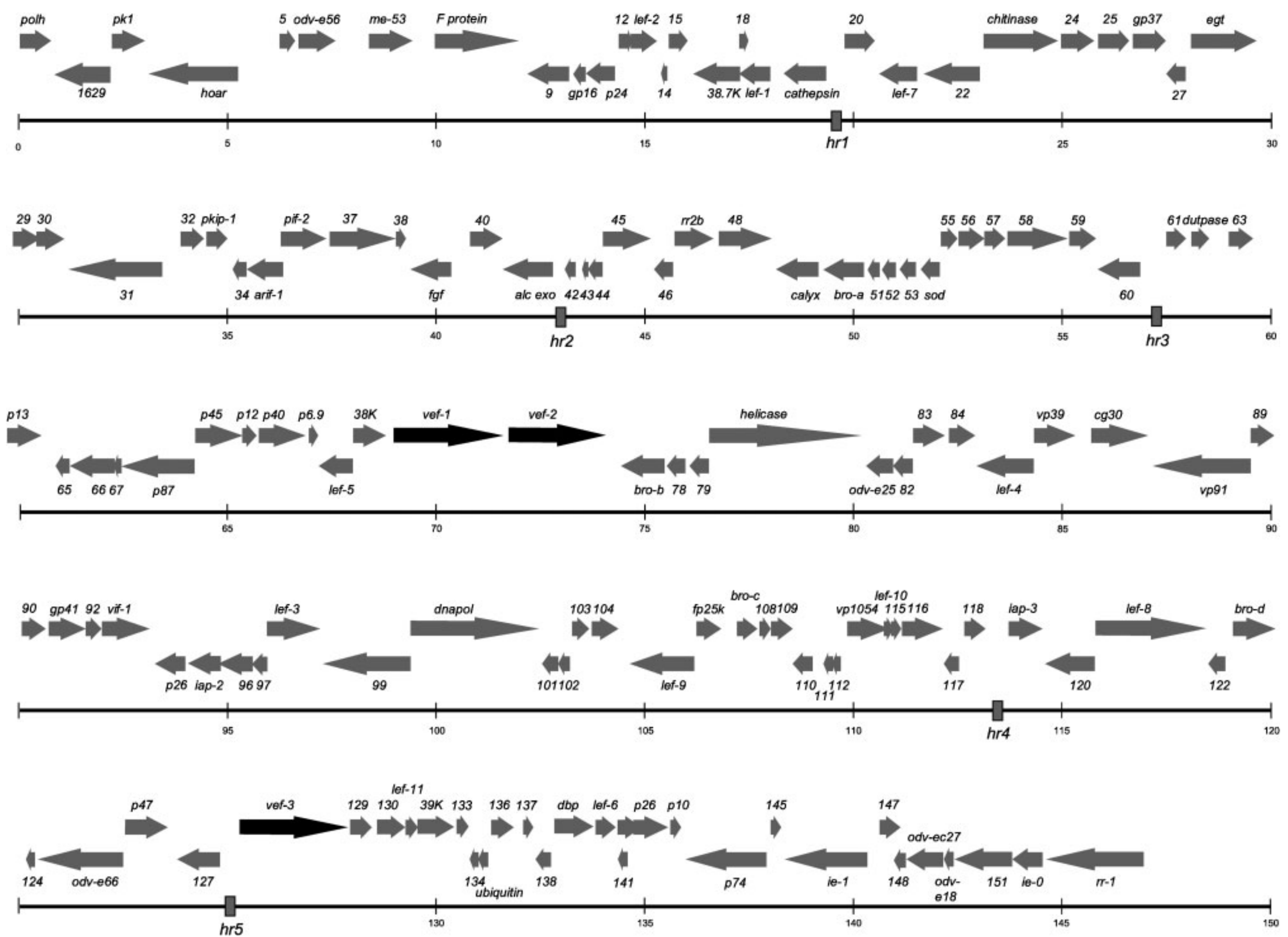


Fig. 1. Linear presentation of the circular AgseNPV-A genome with annotated genes. ORFs are numbered 1–153 relative to polyhedrin (ORF1). Arrows indicate the orientation of the predicted genes. Numbers refer to the nucleotide position in kbp relative to the start codon of polyhedrin. Arrows representing enhancin genes are in black.

Table 1. Potentially expressed ORFs in the genome of AgseNPV-A in comparison to other baculovirus genomes

ORF	+/−	Position		Name*	Length (aa)	E/L	Homologue				Identity (%)†			
							SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§	SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§
1	+	1	741	*polyhedrin*	246	L	Se1	Ac8	Mc1	As1	92	84	89	52
2	−	2179	833	*orf 1629*	448		Se2	Ac9	Mc2	–	44	14	30	–
3	+	2202	2996	*pk1*	264	L	Se3	Ac10	Mc3	As3	67	36	56	24
4	−	5211	3097	*hoar*	704		Se4	–	Mc4	–	33	–	20	–
5	+	6226	6600		124	E, L	–	–	–	–	–	–	–	–
6	+	6687	7796	*odv-e56*	369		Se6	Ac148	Mc6	As15	61	48	58	39
7	+	8037	9086	*me 53*	349	L	Se7	Ac139	Mc7	As131	53	21	46	17
8	+	9959	11971	*F protein*	670	E, L	Se8	Ac23	Mc8	As25	60	14	40	20
9	−	13181	12174		335	E	–	–	Mc9	–	–	–	44	–
10	−	13568	13278	*gp16*	96	L	Se9	Ac130	Mc10	–	62	31	63	–
11	−	14283	13585	*p24*	232	L	Se10	Ac129	Mc11	As62	63	33	62	18
12	+	14365	14685		106	L	Se11	–	Mc12	–	43	–	36	–
13	+	14645	15277	*lef-2*	210		Se12	Ac6	Mc13	As35	62	37	53	19
14	−	15530	15378		50		–	–	–	–	–	–	–	–
15	+	15551	16012		153	E	–	–	–	–	–	–	–	–
16	−	17251	16148	*38.7K*	367	L	Se13	Ac13	Mc31	As63	58	25	59	10
17	−	17900	17253	*lef-1*	215	L	Se14	Ac14	Mc30	As64	65	33	62	30
18	+	18100	18327		75		Se15	–	Mc29	–	38	–	47	–
19	−	19355	18324	*cath*	343		Se16	Ac127	Mc28	As31	84	51	80	44
*hr*1						–	–	–	–	–	–	–	–	
20	+	19775	20512		245		–	Ac105	Mc27	As132	–	13	22	19
21	−	21541	20615	*lef-7*	308	E	17+18	Ac125	–	–	12	14	–	–
22	−	23022	21673		449	E	22–24	–	–	As39	36	–	–	46
23	+	23142	24890	*chitinase*	582	L	Se19	Ac126	Mc19	As32	79	62	79	53
24	+	24968	25705		246	E	–	–	–	–	–	–	–	–
25	+	25863	26606		247	E	–	–	Mc16	–	–	–	25	–
26	+	26698	27480	*gp37*	260	L	Se25	Ac64	Mc32	–	72	47	70	–
27	−	28008	27496	*ptp-2*	170		Se26	–	Mc33	–	56	–	44	–
28	+	28088	29659	*egt*	523	L	Se27	Ac15	Mc34	As129	77	44	73	34
29	+	29843	30403		186	E	Se28	–	Mc35	–	47	–	50	–
30	+	30405	31049		214		Se29	Ac17	Mc36	–	44	13	40	–
31	−	33740	31179		853		Se30	–	Mc37	–	45	–	40	–
32	+	33862	34422		186	L	Se31	Ac4	Mc39	As58	32	15	33	12
33	+	34485	34985	*pkip-1*	166	L	Se32	Ac24	Mc40	–	61	19	52	–
34	−	35440	35102		112		Se33	–	Mc41	–	48	–	51	–
35	−	36314	35442	*arif-1*	290		Se34	Ac21	Mc42	–	42	15	35	–
36	+	36205	37407	*pif-2*	400		Se35	Ac22	Mc43	As43	80	56	69	42
37	+	37436	39019	*pif-1*	527		Se36	Ac119	Mc44	As65	56	46	60	30

Table 1. cont.

ORF	+/−	Position		Name*	Length (aa)	E/L	Homologue				Identity (%)†			
							SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§	SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§
38	+	39024	39266		80		Se37	Ac120	Mc45	–	53	28	43	–
39	−	40494	39373	*fgf*	373	E, L	Se38	Ac32	Mc46	As113	38	11	34	8
40	+	40802	41575		257		Se 40	–	Mc 47	–	45	–	33	–
41	−	42799	41591	*alc exo*	402	L	Se41	Ac133	Mc48	As115	54	36	47	30
*hr*2						–	–	–	–	–	–	–	–	
42	−	43348	43079		89		–	–	–	–	–	–	–	–
43	−	43641	43483		52	L	–	–	–	–	–	–	–	–
44	−	43978	43622		118	L	Se42	Ac19	Mc49	–	41	15	57	–
45	+	43980	45122		380	L	Se43	Ac18	Mc50	–	64	61	22	–
46	−	45673	45215		152	L	Se44	–	Mc51	–	35	–	40	–
47	+	45706	46647	*rr2b*	313		Se45	–	Mc52	–	77	–	71	–
48	+	46761	48026		421	L	–	–	–	–	–	–	–	–
49	−	49162	48137	*calyx/pep*	341	L	Se46	Ac131	Mc61	–	68	23	63	–
50	−	50244	49270	*bro-a*	324	E, L	–	–	–	–	–	–	–	–
51	−	50629	50336		97	L	Se47	Ac117	Mc62	–	58	28	56	–
52	−	51012	50674		112		–	–	Mc63	–	–	–	48	–
53	−	51488	51090		132	E, L	–	–	–	–	–	–	–	–
54	−	52058	51603	*sod*	151	L	Se48	Ac31	Mc65	As54	80	71	80	53
55	+	52080	52484		134		Se49	–	Mc66	–	15	–	39	–
56	+	52509	53132	*pif-3*	207	L	Se50	Ac115	Mc67	As30	66	44	68	30
57	+	53137	53643		168	L	Se51	–	Mc68	–	39	–	43	–
58	+	53689	55131		480	L	Se52	–	Mc69	–	54	–	34	–
59	+	55162	55812	*ac106*	216	L	Se53	Ac106	Mc70	As46	77	15	77	38
60	−	56940	55852		362	E	Se54	Ac33	Mc71	–	57	13	58	–
*hr*3						–	–	–	–	–	–	–	–	
61	+	57494	57964		156		–	–	–	–	–	–	–	–
62	+	58093	58521	*dutpase*	142	L	Se55	–	Mc73	–	72	–	69	–
63	+	58987	59574		195	E, L	–	–	–	–	–	–	–	–
64	+	59694	60518	*p13*	274	L	Se56	–	Mc75	As42	70	–	67	44
65	−	61188	60859	*hp Se 58*	109	L	Se58	Ac108	Mc78	–	60	24	55	–
66	−	62271	61201	*ac109*	356		Se59	Ac109	Mc79	As48	89	43	77	31
67	−	62434	62255	*ac110*	59		Se60	Ac110	Mc80	As92	74	29	71	20
68	−	64179	62431	*p87*	582	L	Se61	Ac104	Mc81	–	56	13	34	–
69	+	64212	65345	*p45*	377	L	Se62	Ac103	Mc82	As72	79	52	79	33
70	+	65335	65676	*p12*	113	L	Se63	Ac102	Mc83	As73	63	26	47	18
71	+	65739	66848	*p40*	369	L	Se64	Ac101	Mc84	As74	75	40	67	19
72	+	66932	67174	*p6·9*	80	L	Se65	Ac100	Mc85	As75	54	41	64	28
73	−	67989	67171	*lef-5*	272	L	Se66	Ac99	Mc86	As76	77	48	77	36

Table 1. cont.

ORF	+/−	Position		Name*	Length (aa)	E/L	Homologue				Identity (%)†			
							SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§	SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§
74	+	67996	68790	*38K*	264	L	Se67	Ac98	Mc87	As77	65	38	63	36
75	+	68965	71598	*vef-1*	877	L	–	–	Mc88	As55	–	–	36	14
76	+	71721	74372	*vef-2*	883	L	–	–	Mc88	As55	–	–	37	15
77	−	75470	74418	*bro-b*	350		–	–	–	–	–	–	–	–
78	−	75969	75526		147		–	–	Mc90	–	–	–	68	–
79	−	76582	76060	*ac 96*	173	L	Se69	Ac96	Mc91	As78	78	48	79	30
80	+	76538	80179	*helicase*	1213	L	Se70	Ac95	Mc92	As79	73	40	78	25
81	−	80943	80293	*odv-e25*	216	L	Se71	Ac94	Mc93	As81	84	43	83	46
82	−	81416	80940	*ac 93*	158	L	Se72	Ac93	Mc94	As82	77	49	86	32
83	+	81424	82182	*ac 92*	252	L	Se73	Ac92	Mc95	As83	87	53	82	32
84	+	82279	82902		207		–	–	Mc96	–	–	–	28	–
85	−	84319	82946	*lef-4*	457		Se74	Ac90	Mc97	As85	64	46	66	29
86	+	84318	85310	*vp39*	330	L	Se75	Ac89	Mc98	As85	77	40	54	10
87	+	85692	87053	*cg30*	453		Se76	Ac88	Mc99	–	17	10	14	–
88	−	89514	87178	*vp91*	778	L	Se77	Ac83	Mc100	As91	58	37	56	16
89	+	89519	90082	*ac 82*	187	L	Se78	Ac82	Mc101	As93	62	25	60	19
90	+	90054	90629	*ac 81*	191	L	Se79	Ac81	Mc102	As94	58	42	52	44
91	+	90688	91581	*gp41*	297	L	Se80	Ac80	Mc103	As95	81	40	71	27
92	+	91581	91955	*ac 78*	124	L	Se81	Ac78	Mc104	As96	58	24	41	16
93	+	91957	93105	*vlf-1*	382	L	Se82	Ac77	Mc105	As97	88	63	90	31
94	−	93973	93239	*p26*	244	E	Se87	Ac136	Mc108	–	62	19	65	–
95	−	94807	94028	*iap-2*	259	L	Se88	Ac71	Mc109	–	50	30	55	–
96	−	95597	94755	*hp Se 89*	280	L	Se89	Ac69	Mc110	–	62	40	60	–
97	−	95933	95566	*ac 68*	122		Se90	Ac68	Mc111	As104	69	31	74	28
98	+	95933	97201	*lef-3*	422	L	Se91	Ac67	Mc112	As103	52	21	47	11
99	−	99363	97276	*ac 66*	695	L	Se92	Ac66	Mc113	As102	45	12	26	10
100	+	99362	102439	*DNApol*	1025		Se93	Ac65	Mc114	As101	74	42	72	29
101	−	102909	102520	*ac 75*	129	L	Se94	Ac75	Mc115	As100	79	21	84	15
102	−	103173	102916	*ac 76*	85	L	Se95	Ac76	Mc116	As99	85	40	87	32
103	+	103250	103651		133	L	Se96	Ac150	–	–	25	16	–	–
104	+	103691	104355		221		–	–	Mc120	As89	–	–	46	7
105	−	106171	104627	*lef-9*	514		Se97	Ac62	Mc123	As107	85	61	82	50
106	+	106225	106812	*fp25K*	195	L	Se98	Ac61	Mc124	As108	90	55	88	27
107	+	107188	107679	*bro-c*	163		–	–	–	–	–	–	–	–
108	+	107738	107998		86	E, L	Se100	Ac60	Mc127	As84	77	37	59	26
109	+	108012	108538		174	L	Se101	Ac59	Mc128	–	48	16	41	–
110	−	109013	108531		160		Se102	Ac57	Mc129	–	61	37	57	–
111	−	109500	109267		77	L	Se103	Ac56	Mc130	–	39	20	41	–

Table 1. cont.

ORF	+/−	Position		Name*	Length (aa)	E/L	Homologue				Identity (%)†			
							SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§	SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§
112	−	109687	109478		69		Se104	Ac55	Mc131	–	66	30	70	–
113	−	110856	109843	*vp1054*	337	L	Se105	Ac54	Mc132	As127	72	39	67	30
114	−	110717	110947	*lef-10*	76		Se106	Ac53a	Mc133	As126	70	42	57	27
115	+	110892	111134		80	L	–	–	Mc134	–	–	–	45	–
116	+	111150	112130		326	L	Se107	–	Mc135	–	49	–	30	–
117	−	112517	112152	*ac 53*	121	E, L	Se108	Ac53	Mc136	As122	59	43	51	18
118	+	112642	113151		169	E	Se109	Ac52	Mc137	–	26	13	46	–
hr4						–	–	–	–	–	–	–	–	
119	+	113702	114517	*iap-3*	271		Se110		Mc138	As53	42	–	40	30
120	−	115789	114596		397	L	Se111	Ac51	Mc139	–	42	9	32	–
121	+	115795	118443	*lef-8*	882		Se112	Ac50	Mc140	As118	77	61	78	48
122	−	118907	118506		133	E	–	–	Mc141	–	–	–	23	–
123	+	119095	120105	*bro-d*	336	E	–	–	–	–	–	–	–	–
124	−	120360	120154		68	L	Se113	Ac43	Mc142	–	47	30	50	–
125	−	122479	120443	*odv-e66*	678	L	Se114	Ac46	Mc143	As33	51	22	47	20
126	+	122528	123733	*p47*	401		Se115	Ac40	Mc144	As60	77	50	75	40
127	−	124794	123778		338	E	–	–	Mc145	–	–	–	31	–
hr5						–	–	–	–	–	–	–	–	
128	+	125278	127866	*vef-3*	862	L	–	–	Mc88	As55	–	–	35	15
129	+	127923	128444		173		Se117	–	Mc146	As60	35	–	27	6
130	+	128559	129236	*ac 38*	225	E, L	Se118	Ac38	Mc147	As61	69	51	82	33
131	+	129168	129544	*lef-11*	125		Se119	Ac37	Mc148	As52	52	30	60	26
132	+	129534	130412	*39K/pp31*	292	E, L	Se120	Ac36	Mc149	As51	50	27	45	11
133	+	130485	130769		94		Se121	–	–	–	28	–	–	–
134	−	131004	130795		69		Se122	–	Mc150	–	42	–	32	–
135	−	131234	130998	*ubiquitin*	78	L	Se123	Ac35	Mc151	As47	91	73	74	59
136	+	131300	131836		178	L	Se124	Ac34	Mc152	–	70	27	60	–
137	+	132071	132316		81		–	–	–	–	–	–	–	–
138	−	132733	132362		123	L	Se125	Ac26	Mc153	–	51	25	56	–
139	+	132816	133754	*dbp*	312	L	Se126	Ac25	Mc154	As69	58	23	59	12
140	+	133790	134295	*lef-6*	168	L	Se127	Ac28	Mc155	As68	40	23	47	6
141	−	134578	134345	*ac 29*	77	E, L	Se128	Ac29	Mc156	As16	42	32	75	16
142	+	134714	135541	*p26*	275	E, L	Se129	Ac136	Mc157	–	58	30	62	–
143	+	135601	135864	*p10*	87	L	Se130	Ac137	Mc158	As114	72	20	60	19
144	−	137902	135968	*p74*	644	E, L	Se131	Ac138	Mc159	As56	72	53	64	38
145	+	138006	138257		83	E, L	–	–	Mc160	–	–	–	50	–
146	−	140327	138342	*ie1*	661	L	Se132	Ac147	Mc161	As8	46	21	43	8
147	+	140336	140911	*ac 146*	191	L	Se133	Ac146	Mc162	As9	54	31	63	19

Table 1. cont.

ORF	+/–	Position		Name*	Length (aa)	E/L	Homologue				Identity (%)†			
							SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§	SeMNPV	AcMNPV	MacoNPV-B‡	AgseGV§
148	–	141239	140961	*ac 145*	92	L	Se134	Ac145	Mc163	As10	81	33	70	36
149	–	142103	141264	*odv-ec27*	279	L	Se135	Ac144	Mc164	As87	87	53	90	25
150	–	142399	142166	*odv-e18*	77	L	Se136	Ac143	Mc165	As11	67	35	62	36
151	–	143790	142408	*ac 142*	460	L	Se137	Ac142	Mc166	As12	88	48	77	33
152	–	144524	143805	*ie-0*	239	L	Se138	Ac141	Mc167	–	64	24	61	–
153	–	146949	144610	*rr-1*	779	E	Se139	–	Mc168	As37	58	–	60	16

*Genes present in all lepidopteran NPVs and GVs are underlined.
†Percentage identity values were calculated for complete ORFs by using CLUSTAL_X.
‡MacoNPV isolate 96B (Li *et al.*, 2002a).
§AgseGV (GenBank accession no. NC_005839).

The AgseNPV-A genome content and organization were compared with those of four other baculoviruses: SeMNPV, AcMNPV, MacoNPV-B and AgseGV (Table 2). AgseNPV-A shares 127 ORFs with SeMNPV, 134 with MacoNPV-B, 103 with AcMNPV and 81 with AgseGV, with mean amino acid identities of 60·0, 55·9, 34·2 and 26·9 %, respectively. The highest mean amino acid identity was found between AgseNPV-A and SeMNPV, but the largest number of ORFs is shared with MacoNPV-B (134), underscoring a close relationship with MacoNPV-B as well. This is further evidenced by phylogenetic (Jakubowska *et al.*, 2005) and gene-parity plot (Fig. 2) analyses. Seventy-eight ORFs show the highest percentage of identity with SeMNPV, 53 with MacoNPV-B, one with AcMNPV (Agse21) and one with AgseGV (Agse22).

The most-conserved genes between AgseNPV-A, SeMNPV, AcMNPV and MacoNPV-B are polyhedrin, superoxide dismutase (*sod*), ubiquitin, chitinase, *vfl-1*, *lef-8*, *lef-9*, *ac92*, *fp25K* and *ac38*. One hundred AgseNPV-A ORFs have homologues in all three NPVs with which they were compared. The presence of an F protein (Agse8) and the absence of *gp64* classify AgseNPV-A as a group II NPV and this is in agreement with previous phylogenetic analyses based on polyhedrin, *lef-8* and *pif-2* sequences (Jakubowska *et al.*, 2005).

The identity of the 81 genes shared by AsgeNPV-A and AgseGV, despite being isolated from the same host (*A. segetum*), is very low. None of the 19 shared genes that do not belong to the 62 common lepidopteran baculovirus genes (Herniou *et al.*, 2003) are exclusively present in AgseNPV and AgseGV.

To examine the genome organization, the order of homologous ORFs of AgseNPV-A, SeMNPV, MacoNPV-B and AcMNPV was compared by using gene-parity plot analysis as described previously (Hu *et al.*, 1998) (Fig. 2). Genes of AcMNPV were renumbered manually, starting with the polyhedrin gene as number 1. The gene arrangement of AgseNPV-A was completely collinear with that of SeMNPV. A high degree of collinearity was also observed with MacoNPV-B, with exception of a short gene cluster including *38.7K*, *lef-1* and cathepsin. When compared with AcMNPV, a major part of the AgseNPV-A genome (around 70 000–140 000 bp) is inversely oriented relative to the orientation of the polyhedrin gene, but the gene order in this region is similar in both viruses. The other parts of the genome differ considerably. Parity analysis of AgseNPV-A and AgseGV ORFs only provided a scattered, non-informative distribution (not shown).

Comparison of AgseNPV-A and SeMNPV ORFs

Having observed the high level of genome collinearity of AgseNPV-A and SeMNPV, we compared the gene content of both NPVs. Among the 127 homologues between AgseNPV-A and SeMNPV, the most conserved genes are polyhedrin, with amino acid identity of 92 %, and *sod*,

Table 2. Characteristics of some baculovirus genomes

Characteristic	AgseNPV-A	SeMNPV	MacoNPV-B	AcMNPV	AgseGV
Size (bp)	147 544	135 611	158 482	133 894	131 680
G+C content (mol%)	46	44	40	41	37
Total no. ORFs	153	139	168	156	132
No. *hrs*	5	6	4	8	–
Mean amino acid identity (%) with AgseNPV-A	–	60·0	55·9	34·2	26·9
No. homologues in AgseNPV-A	–	127	134	103	81
Reference	This paper	IJkel *et al.* (1999)	Li *et al.* (2002a)	Ayres *et al.* (1994)	GenBank no. NC_005839

ubiquitin, chitinase, *vfl-1*, *lef-8* and *lef-9*, with >60 % amino acid identity. Two putative SeMNPV ORFs, Se17 and Se18, constitute one ORF in AgseNPV-A (Agse21), which was assigned as *lef-7* according to homology with Ac125. The best BLAST match for Agse21 was found with *Xestia c-nigrum* (Xecn) GV ORF129 (38 % identity) (Hayakawa *et al.*, 1999). Agse22 was homologous to three SeMNPV ORFs (Se22, Se23 and Se24) and assigned as a single ORF in HearNPV (Chen *et al.*, 2002). The best BLAST match for Agse22 was found with AgseGV ORF39 (46 % identity), but its function is unknown.

There are 12 ORFs present in SeMNPV that are absent in AgseNPV-A: Se5, Se20, Se21, Se39, Se57, Se68, Se83, Se84, Se85, Se86, Se99 and Se116. All except Se99, which was assigned as *p94*, are genes with unknown functions (IJkel *et al.*, 1999). The *p94* gene is not essential for virus replication in cell culture, but may be involved in inhibition of apoptosis (Friesen & Miller, 1987; Clem & Miller, 1994). Having seen the different location of *p94* in AcMNPV compared with AgseNPV, MacoNPV-A and B and SeMNPV, it may have been acquired by the ancestor of these viruses in an independent insertion and from a source different from the *p94* of AcMNPV (as suggested for SeMNPV *p94*; IJkel *et al.*, 2001). No homologue was found in AgseGV.

Between Agse64 and Agse65, a 182 nt fragment was detected with nucleotide similarity to *odv-e66*. This gene encodes an occlusion-derived virion (ODV) protein and is present in SeMNPV in two copies, Se57 and Se114, which are both >2 kbp long. The identity between the two SeMNPV ORFs is only 32 % (IJkel *et al.*, 2001). An intact *odv-e66* gene appears to be present in the AgseNPV-A genome in another position (Agse125) and has a length of 2036 nt, which is comparable in size with the complete Se114 ORF (2057 nt). In AgseNPV-A, the *odv-e66* gene is located next to the *p47* gene, as is the case in SeMNPV. The identity between Agse125 and Se114 is 52 %, which is higher than that between the two SeMNPV *odv-e66* genes. It is probable that Se57 was acquired independently of Se114 from a source related more closely to LdMNPV (Kuzio *et al.*, 1999) or *Leucania separata* (Ls) NPV (Wang *et al.*, 1995). In AgseNPV-A, an Se57 homologue may have been lost during evolution, as there is a small part of this gene left in the AgseNPV-A genome between Agse64 and Agse65. There are also two copies of *odv-e66* in AcMNPV and MacoNPV-B. All sequenced GVs, except for *Adoxophyes orana* (Ador) GV (Wormleaton *et al.*, 2003), contain a single copy of an *odv-e66* gene.

With an increasing number of complete baculovirus genome sequences, the number of unique ORFs decreases. Twenty unique ORFs were assigned to SeMNPV (IJkel *et al.*, 1999). To the best of our knowledge, seven ORFs were still unique to SeMNPV with 28 baculovirus genomes having

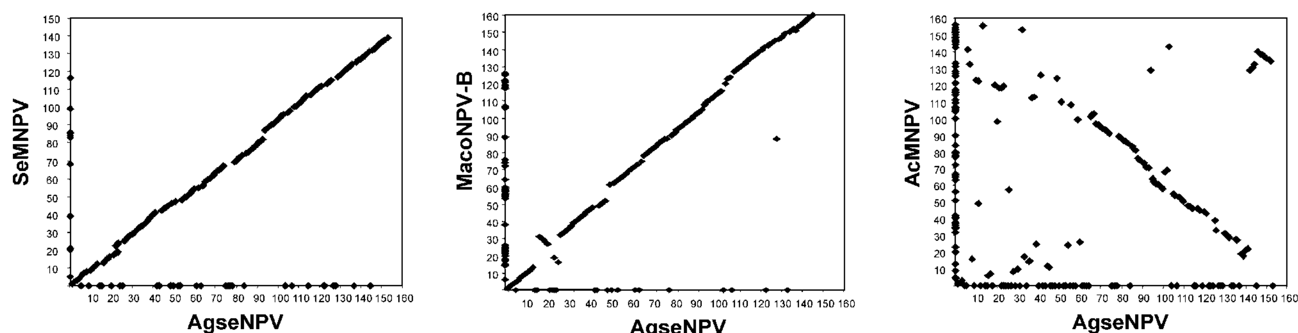


Fig. 2. Pairwise comparison of gene content and position of AgseNPV-A with SeMNPV, AcMNPV and MacoNPV-B using gene-parity plot analysis. Genes present in only one of the two viruses in the pairwise comparison appear on the *x* or *y* axis.

been sequenced: Se5, Se20, Se39, Se83, Se85, Se86 and Se121. A homologue of Se121 (Agse133) is now identified in the AgseNPV-A genome, thus reducing the number of unique SeMNPV ORFs to six. The function of Agse133 is so far unknown. Similarly, a homologue of *Trichoplusia ni* SNPV ORF62 (Willis *et al.*, 2005) (Tn62) was identified in the AgseNPV-A genome. This ORF is so far present exclusively in TnSNPV and AgseNPV-A.

Unique genes in AgseNPV-A

There are ten unique ORFs in AgseNPV-A (Agse5, Agse14, Agse15, Agse42, Agse43, Agse48, Agse53, Agse61, Agse63 and Agse137) and they account for 3 % of the genome in total. Their length ranges from 50 to 421 aa. For six unique ORFs, baculovirus early or late promoter motifs were found upstream of the ATG codons, as well as putative polyadenylation signals, indicating that these may represent expressed genes. Five of these (Agse5, Agse15, Agse48, Agse53 and Agse63) are larger than 100 aa. Agse5 exhibits homology to an *Arabidopsis thaliana* helicase domain-containing protein, with a BLAST *e* value of 0·049.

bro genes

A common feature of baculovirus genomes is the presence of repeated ORFs, named *bro* (baculovirus repeated ORFs) genes by Kuzio *et al.* (1999). The highest number of *bro* genes was identified in LdMNPV (Kuzio *et al.*, 1999), which contains 16 ORFs related to the AcMNPV *bro* gene Ac2. The role of the *bro* gene family has not yet been defined. The *bro* genes were subdivided into four groups, depending on the percentage of amino acid identity in both termini and in the central region, as well as their length (Kuzio *et al.*, 1999). We have identified four *bro* genes dispersed along the genome of AgseNPV-A (Agse50, Agse77, Agse107 and Agse123) and named them *bro-a* to *bro-d*, according to the order of appearance on the linearized genome. *bro-a* shows the highest amino acid identity (53 %) to *bro-c* of *Bombyx mori* NPV, *bro-b* to *bro-d* of MacoNPV-B (63 %), *bro-c* to *bro-h* of MacoNPV-A (51 %) and *bro-d* to Ac2 (58 %). The size of each *bro* gene corresponds well to that of their homologues in other baculoviruses. *bro-a*, *bro-b* and *bro-d* belong to group I *bro* genes, whilst *bro-c* shows the highest similarity to group II *bro* genes according to the classification of Kuzio *et al.* (1999). AgseNPV-A *bro-b* is located directly downstream of the second enhancin gene (Agse76), like *bro-d* in the MacoNPV-B genome. We have not found any relatedness of the location of *bro* genes to that of *hr*s, which was suggested for the LdMNPV genome (Kuzio *et al.*, 1999). Alignment of all AgseNPV-A *bro* genes revealed similarities ranging from 7 % (*bro-b*/*bro-d* against *bro-c*) to 53 % between *bro-b* and *bro-d* (analysis not shown).

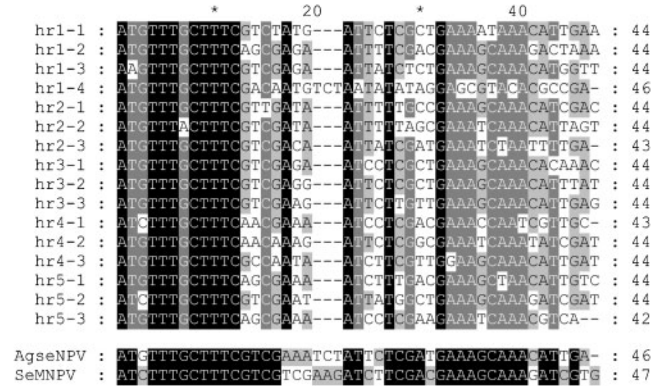
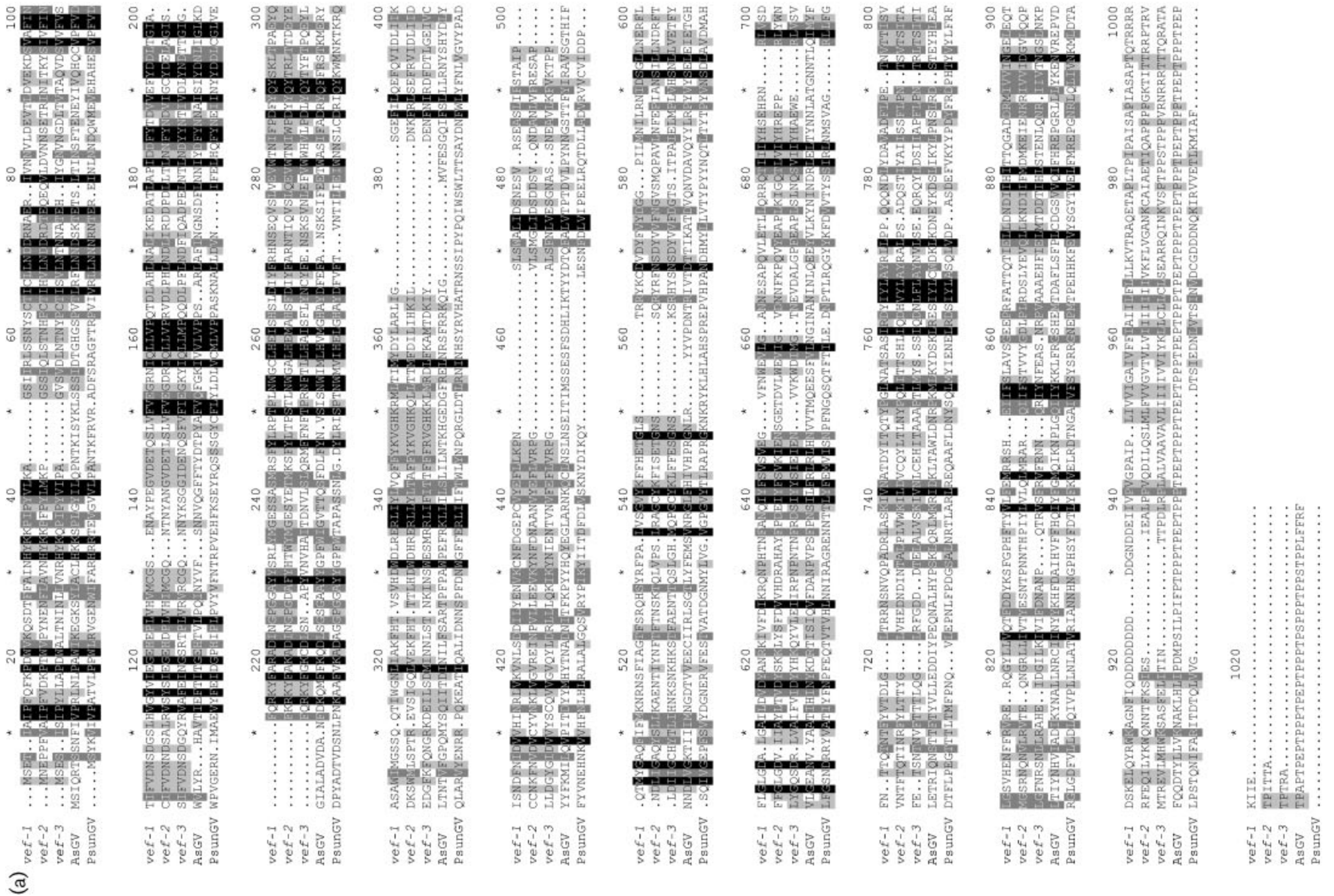


Fig. 3. AgseNPV-A homologous regions in comparison with SeMNPV. The nucleotide sequences were aligned in CLUSTAL_X and displayed in GeneDoc. Black shading indicates 100 % identity, dark-grey shading indicates 80–100 % identity and light-grey shading indicates 60–80 % identity. The last two lines show the AgseNPV-A and SeMNPV *hr* consensus sequence.

*hr*s

*hr*s were first identified in AcMNPV (Cochran & Faulkner, 1983; Pearson *et al.*, 1992). *hr*s are *cis*-acting putative origins of DNA replication (*ori*) (Ahrens & Rohrmann, 1995; Kool *et al.*, 1995) and may also act as enhancers of transcription (Guarino & Summers, 1986; Guarino *et al.*, 1986). AcMNPV *hr*s have two to eight repeats of a 72 bp long sequence, with an internal imperfect palindrome with an *Eco*RI restriction site in the centre. In five of the 28 baculoviruses sequenced until now [*Chrysodeixis chalcites* (Chch) NPV (van Oers *et al.*, 2005), TnSNPV, AdorNPV, AgseGV and *Cydia pomonella* (Cp) GV (Luque *et al.*, 2001)], canonical *hr*s were not identified. The remaining fully sequenced baculoviruses contain between three [*Cryptophlebia leucotreta* (Crle) GV] (Lange & Jehle, 2003) and 17 [*Spodoptera litura* (Splt) NPV] (Pang *et al.*, 2001) *hr*s.

Fig. 4. Alignment and phylogenies of enhancin genes. (a) Alignment of enhancin sequences of AgseNPV-A (*vef1*–*vef3*), AgseGV and PsunGV. The amino acid sequences were aligned in CLUSTAL_X and displayed in GeneDoc. Black shading indicates 100 % identity, dark-grey shading indicates 80–100 % identity and light-grey shading indicates 60–80 % identity. Conserved motif is underlined. (b) Phylogenies of enhancin amino acid sequences obtained by MP analysis. Branch lengths were determined by NJ. Numbers indicate MP bootstrap scores using 1000 replicates. Bacterial enhancin sequences were used as an outgroup. GenBank accession numbers: AgseGV, NC_005839; AgseNPV-A, DQ123841; *Bacillus anthracis* Ames, AE017034; *Bacillus cereus* ZK, NC_006274; *Bacillus thuringiensis* serovar konkukian str. 97-27, AE017355; ChfuGV, AF319939; CfMNPV, AF512031; HearGV, D28558; LdMNPV, AF081810; MacoNPV-A 90/2, AF467808; MacoNPV-A 90/4, AF539999; MacoNPV-B, AY126275; PsunGV, D14871; TnSNPV, D12617; XecnGV, AF162221; *Yersinia pestis* biovar Medievalis, AE017128; *Yersinia pseudotuberculosis* IP32953, BX936398.


(a)
(b)
B. cereus ZK
B. anthracis Ames
B. thuringiensis
Y. pestis
Y. pseudotuberculosis
AgseNPV vef-3
AgseNPV vef-1
AgseNPV vef-2
MacoNPV-B
MacoNPV-A 90/4
MacoNPV-A 90/2
AgseGV
XecnGV vef-4
XecnGV vef-1
XecnGV vef-2
TnSNPV
ChfuGV
PsunGV
HearGV
XecnGV vef-3
LdMNPV vef-1
LdMNPV vef-2
CfMNPV
0·1 substitutions per site

Five *hr*s have been identified in the AgseNPV-A genome (Fig. 3). They are dispersed along the genome in intergenic regions, around map positions 19·5 (*hr*1), 42·8 (*hr*2), 57·0 (*hr*3), 113·2 (*hr*4) and 124·8 (*hr*5) (see Fig. 1, Table 1). The conserved repeat size is around 40 nt. All AgseNPV-A *hr*s contain from one to four imperfect palindromic repeats. Thirteen nucleotides are absolutely conserved in all 16 repeats identified. The striking feature of AgseNPV-A *hr*s is a very high consensus sequence similarity to SeMNPV *hr*s. Two motifs also characteristic for SeMNPV *hr*s, TTTGCTTT and GAAAGCAAAC, are present in almost all AgseNPV-A *hr* repeats. A notable feature of the AgseNPV-A *hr*s also is a low G + C content of around 30 mol%, as in SeMNPV.

Enhancin genes

A most significant feature of AgseNPV-A, in particular relative to SeMNPV, is the presence of three enhancin genes (*vef*) in AgseNPV-A. Enhancins, or synergistic or viral-enhancing factors (VEF), have been found to dramatically increase the infectivity of baculoviruses in non-natural lepidopteran hosts (Derksen & Granados, 1988). They function by enzymic hydrolysis of the peritrophic membrane, the barrier for pathogens in the insect midgut, or by increasing the fusion efficiency with midgut cells through interaction between the viral envelope and the cell plasma membrane (Wang *et al.*, 1994; Bischoff & Slavicek, 1997). Enhancins are metalloproteases, which degrade mucins, the major proteins of the peritrophic membrane (Lepore *et al.*, 1996; Wang *et al.*, 1997). The first enhancin described originated from *Pseudaletia unipuncta* (Psun) GV and increased the infectivity of PsunNPV in NPV/GV mixed infections (Tanada & Hukuhara, 1971; Yamamoto & Tanada, 1980; Zhu *et al.*, 1989). Enhancin protein constitutes up to 5 % of the total protein content in the granules of GVs and is localized in the granule matrix (Tanada, 1985). Recently, it was shown that LdMNPV enhancins are present in ODV envelopes in association with nucleocapsids (Slavicek & Popham, 2005).

To date, enhancins have been described in many GVs, including PsunGV, *T. ni* (Tni) GV, *Helicoverpa armigera* (Hear) GV (Roelvink *et al.*, 1995), AgseGV (GenBank accession no. NC_005839), *C. fumiferana* (Cf) GV (GenBank accession no. AF319939) and XecnGV (Hayakawa *et al.*, 1999), and in four group II NPVs, CfMNPV (de Jong *et al.*, 2005), LdMNPV (Bischoff & Slavicek, 1997), MacoNPV-A isolates 90/2 and 90/4 (Li *et al.*, 2002a, 2005) and MacoNPV-B (Li *et al.*, 2002b). So far, enhancin has not been found in any group I NPV. Most of the group II NPVs and GVs carry a single copy of an enhancin gene. LdMNPV has two enhancin copies, whilst in the XecnGV genome, four enhancins were found.

AgseNPV-A encodes three enhancin genes, *vef-1*, *vef-2* and *vef-3* (Fig. 4), as Agse75, Agse76 and Agse128. For all three AgseNPV-A *vef*s, potential baculovirus consensus late promoter motifs were found, suggesting expression in the late stage of infection. This is compatible with their association with ODVs (Slavicek & Popham, 2005). The first two enhancin genes, *vef-1* and *vef-2*, are located in tandem downstream of the *38K* gene (Agse74). This location corresponds to the position of Se68 in the SeMNPV genome (IJkel *et al.*, 1999). AgseNPV-A lacks an Se68 homologue (function unknown) at this position. Compared with SeMNPV, *vef-3* (Agse128) is in place of Se116, also an ORF with unknown function. All three AgseNPV-A *vef* genes contain large ORFs of 2633, 2652 and 2588 nt, respectively. These sizes are comparable to those of other baculovirus enhancin genes, which range from 2352 nt in LdMNPV to 3015 nt in AgseGV. The identity between the predicted AgseNPV-A VEF proteins ranges from 30 to 40 % and the similarity oscillates around 60 % (Fig. 4a). Overall, AgseNPV-A enhancins have only about 20 % amino acid identity with other baculovirus enhancins. Identities of up to 40 % were only found with the MacoNPV species enhancins.

Proteins in the metalloprotease family are characterized by an HEXXH motif (Bischoff & Slavicek, 1997). AgseNPV-A *vef-1* and *vef-2* encode this conserved zinc-binding domain as HEISH and HEMAH, respectively. AgseNPV-A *vef-3* encodes an HAISF motif at a comparable position, which does not meet the requirement of an HEXXH motif. A similar case was described for two XecnGV enhancin genes (ORF150 and ORF166), with HQIGH and QKIGD motifs aligning with the HEXXH motif of other enhancins (Hayakawa *et al.*, 1999). It is not known whether the aberrant XecnGV enhancin is active. Only a part of known metalloproteases contains this conserved motif, which can be even better defined as abXHEbbHbc, where 'a' is most often valine or threonine, 'b' an uncharged residue and 'c' a hydrophobic residue (Rawlings & Barrett, 1995).

Phylogenetic analysis of enhancins was performed (Fig. 4b) to address the question as to whether the origin of the AgseNPV-A enhancin genes could be found in the granulovirus AgseGV, a virus with the same host. In addition to previously presented phylogenies (Popham *et al.*, 2001; Li *et al.*, 2003), we also included in our analysis the CfMNPV, MacoNPV-A 90/4, MacoNPV-B and AgseGV enhancins, as well as several bacterial enhancin sequences from *Bacillus* spp. and *Yersinia* spp. This resulted in a slightly different clustering of baculovirus enhancin genes, but the previously observed tendency that enhancins of similar size fall in the same clade (Li *et al.*, 2003) is supported by our analysis. Also in this case, AgseNPV and MacoNPVs are closely related.

We obtained two baculovirus enhancin clusters, one consisting of MacoNPVs and AgseNPV-A and a second one including all GV, LdMNPV and CfMNPV enhancins. In the second cluster, AgseGV groups with CfMNPV and LdMNPV enhancins and is apart from other GV enhancins. The bacterial enhancins are grouped together and are separated from the baculovirus enhancins. The analysis is supported by relatively high (mostly > 80 %) bootstrap values for all branches. Phylogenetic analysis also showed that CfMNPV and ChfuGV, as well as AgseNPV-A and AgseGV, enhancins are not closely related, rejecting the

hypothesis, at least for these viruses, that baculoviruses infecting a common host have gained enhancin genes from each other.

Cross-infectivity

The relatively high genetic identity between AgseNPV-A and SeMNPV and their similarity in genome organization may be reflected in their host range. Therefore, AgseNPV-A and SeMNPV infectivity was tested *per os* by using second-instar *S. exigua* and *A. segetum* larvae, respectively. SeMNPV was found to be non-infectious for *A. segetum* at a high dose, whereas AgseNPV-A caused mortality in 23 (46 %) tested *S. exigua* larvae. The identity of the progeny virus was confirmed by PCR with two sets of AgseNPV-A-specific primers (not shown).

Conclusion

In conclusion, the genome of AgseNPV-A was found to be highly collinear with that of SeMNPV in organization. So far, ten ORFs were found to be unique to AgseNPV-A and one ORF, Agse133, is so far only shared with SeMNPV (Se121). The most prominent difference between AgseNPV-A and SeMNPV genomes is the presence of three enhancin gene copies (*vef*) in AgseNPV-A. Sequence information of the AgseNPV genome adds to the knowledge of baculovirus genomes and, in comparison to SeMNPV and AgseGV, may lead to further insight into baculovirus–host interactions.

ACKNOWLEDGEMENTS

This research was supported by the Polish State Committee for Scientific Research, grant no. 2P06R 073 26, and a scholarship from the European Union (Functional Biodiversity and Crop Protection), contract no. HPMT-CT-2000-00199.

REFERENCES

Afonso, C. L., Tulman, E. R., Lu, Z., Balinsky, C. A., Moser, B. A., Becnel, J. J., Rock, D. L. & Kutish, G. F. (2001). Genome sequence of a baculovirus pathogenic for *Culex nigripalpus*. *J Virol* **75**, 11157–11165.

Ahrens, C. H. & Rohrmann, F. (1995). Replication of *Orgyia pseudotsugata* baculovirus DNA: *lef-2* and *ie-1* are essential and *ie-2*, *p34*, and Op-*iap* are stimulatory genes. *Virology* **212**, 650–662.

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

Allaway, G. P. & Payne, C. C. (1983). A biochemical and biological comparison of three European isolates of nuclear polyhedrosis viruses from *Agrotis segetum*. *Arch Virol* **75**, 43–54.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

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

Bischoff, D. S. & Slavicek, J. M. (1997). Molecular analysis of an *enhancin* gene in the *Lymantria dispar* nuclear polyhedrosis virus. *J Virol* **71**, 8133–8140.

Borodovsky, M. & McIninch, J. (1993). GeneMark: parallel gene recognition for both DNA strands. *Comput Chem* **17**, 123–133.

Boughton, A. J., Harrison, R. L., Lewis, L. C. & Bonning, B. C. (1999). Characterization of a nucleopolyhedrovirus from the black cutworm, *Agrotis ipsilon* (Lepidoptera: Noctuidae). *J Invertebr Pathol* **74**, 289–294.

Bourner, T. C. & Cory, J. S. (2004). Host range of an NPV and a GV isolated from the common cutworm, *Agrotis segetum*: pathogenicity within the cutworm complex. *Biol Control* **31**, 372–379.

Bourner, T. C., Vargas-Osuna, E., Williams, T., Santiago Alvarez, C. & Cory, J. S. (1992). A comparison of the efficacy of nuclear polyhedrosis and granulosis viruses in spray and bait formulations for the control of *Agrotis segetum* (Lepidoptera: Noctuidae) in maize. *Biocontrol Sci Technol* **2**, 315–326.

Bulach, D. M., Kumar, C. A., Zaia, A., Liang, B. & Tribe, D. E. (1999). Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerase gene sequences. *J Invertebr Pathol* **73**, 59–73.

Caballero, P., Vargas-Osuna, E. & Santiago-Alvarez, C. (1991). Efficacy of a Spanish strain of *Agrotis segetum* granulosis virus (*Baculoviridae*) against *Agrotis segetum* Schiff (Lep., Noctuidae) on corn. *J Appl Entomol* **112**, 59–64.

Chen, X., Zhang, W. J., Wong, J. & 9 other authors (2002). Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. *J Gen Virol* **83**, 673–684.

Clem, R. J. & Miller, L. K. (1994). Control of programmed cell death by the baculovirus genes *p35* and *iap*. *Mol Cell Biol* **14**, 5212–5222.

Cochran, M. A. & Faulkner, P. (1983). Location of homologous DNA sequences interspersed at five regions in the baculovirus AcMNPV genome. *J Virol* **45**, 961–970.

de Jong, J. G., Lauzon, H. A. M., Dominy, C., Poloumienko, A., Carstens, E. B., Arif, B. M. & Krell, P. J. (2005). Analysis of the *Choristoneura fumiferana* nucleopolyhedrovirus genome. *J Gen Virol* **86**, 929–943.

Derksen, A. C. G. & Granados, R. R. (1988). Alteration of a lepidopteran peritrophic membrane by baculoviruses and enhancement of viral infectivity. *Virology* **167**, 242–250.

Federici, B. A. (1999). Naturally occurring baculoviruses for insect pest control. *Methods Biotechnol* **5**, 301–320.

Friesen, P. D. & Miller, L. K. (1987). Divergent transcription of early 35- and 94-kilodalton protein genes encoded by the *Hin*dIII K genome fragment of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J Virol* **61**, 2264–2272.

Guarino, L. A. & Summers, M. D. (1986). Functional mapping of a *trans*-activating gene required for expression of a baculovirus delayed-early gene. *J Virol* **57**, 563–571.

Guarino, L. A., Gonzales, M. A. & Summers, M. D. (1986). Complete sequence and enhancer function of the homologous DNA regions of *Autographa californica* nuclear polyhedrosis virus. *J Virol* **60**, 224–229.

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

Herniou, E. A., Luque, T., Chen, X., Vlak, J. M., Winstanley, D., Cory, J. S. & O'Reilly, D. R. (2001). Use of whole genome sequence data to infer baculovirus phylogeny. *J Virol* **75**, 8117–8126.

Herniou, E. A., Olszewski, J., Cory, J. S. & O'Reilly, D. R. (2003). The genome sequence and evolution of baculoviruses. *Annu Rev Entomol* **48**, 211–234.

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

Hyink, O., Dellow, R. A., Olsen, M. J., Caradoc-Davies, K. M. B., Drake, K., Herniou, E. A., Cory, J. S., O'Reilly, D. R. & Ward, V. K. (2002). Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus. *J Gen Virol* **83**, 957–971.

Ignoffo, C. M. & Garcia, C. (1979). Susceptibility of larvae of the black cutworm to species of entomopathogenic bacteria, fungi, protozoa, and viruses. *J Econ Entomol* **72**, 767–769.

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

IJkel, W. F. J., Lebbink, R.-J., Op den Brouw, M. L., Goldbach, R. W., Vlak, J. M. & Zuidema, D. (2001). Identification of a novel occlusion derived virus-specific protein in *Spodoptera exigua* multicapsid nucleopolyhedrovirus. *Virology* **284**, 170–181.

Jakubowska, A., van Oers, M. M., Ziemnicka, J., Lipa, J. J. & Vlak, J. M. (2005). Molecular characterization of *Agrotis segetum* nucleopolyhedrovirus from Poland. *J Invertebr Pathol* **90**, 64–68.

Kool, M., Ahrens, C. H., Vlak, J. M. & 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. & Rohrmann, G. F. (1999). Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. *Virology* **253**, 17–34.

Lange, M. & Jehle, J. A. (2003). The genome of the *Cryptophlebia leucotreta* granulovirus. *Virology* **317**, 220–236.

Lepore, L. S., Roelvink, P. R. & Granados, R. R. (1996). Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. *J Invertebr Pathol* **68**, 131–140.

Li, L., Donly, C., Li, Q., Willis, L. G., Keddie, B. A., Erlandson, M. A. & Theilmann, D. A. (2002a). Identification and genomic analysis of a second species of nucleopolyhedrovirus isolated from *Mamestra configurata*. *Virology* **297**, 226–244.

Li, Q., Donly, C., Li, L., Willis, L. G., Theilmann, D. A. & Erlandson, M. (2002b). Sequence and organization of the *Mamestra configurata* nucleopolyhedrovirus genome. *Virology* **294**, 106–121.

Li, Q., Li, L., Moore, K., Donly, C., Theilmann, D. A. & Erlandson, M. (2003). Characterization of *Mamestra configurata* nucleopolyhedrovirus enhancin and its functional analysis via expression in an *Autographa californica* M nucleopolyhedrovirus recombinant. *J Gen Virol* **84**, 123–132.

Li, L., Li, Q., Willis, L. G., Erlandson, M., Theilmann, D. A. & Donly, C. (2005). Complete comparative genomic analysis of two field isolates of *Mamestra configurata* nucleopolyhedrovirus-A. *J Gen Virol* **86**, 91–105.

Luque, T., Finch, R., Crook, N., O'Reilly, D. R. & Winstanley, D. (2001). The complete sequence of the *Cydia pomonella* granulovirus genome. *J Gen Virol* **82**, 2531–2547.

Pang, Y., Yu, J., Wang, L. & 7 other authors (2001). Sequence analysis of the *Spodoptera litura* multicapsid nucleopolyhedrovirus genome. *Virology* **287**, 391–404.

Pearson, M., Bjornson, R., Pearson, G. & Rohrmann, G. (1992). The *Autographa californica* baculovirus genome: evidence for multiple replication origins. *Science* **257**, 1382–1384.

Popham, H. J. R., Bischoff, D. S. & Slavicek, J. M. (2001). Both *Lymantria dispar* nucleopolyhedrovirus *enhancin* genes contribute to viral potency. *J Virol* **75**, 8639–8648.

Rawlings, N. D. & Barrett, A. J. (1995). Evolutionary families of metallopeptidases. *Methods Enzymol* **248**, 183–228.

Reed, C., Otvos, I. S., Reardon, R., Ragenovich, I. & Williams, H. L. (2003). Effects of long-term storage on the stability of *Op*MNPV DNA contained in TM Biocontrol-1. *J Invertebr Pathol* **84**, 104–113.

Roelvink, P. W., Corsaro, B. G. & Granados, R. R. (1995). Characterization of the *Helicoverpa armigera* and *Pseudaletia unipuncta* granulovirus enhancin genes. *J Gen Virol* **76**, 2693–2705.

Slavicek, J. M. & Popham, H. J. R. (2005). The *Lymantria dispar* nucleopolyhedrovirus enhancins are components of occlusion-derived virus. *J Virol* **79**, 10578–10588.

Swofford, D. L. (2003). PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sunderland, MA: Sinauer Associates.

Tanada, Y. (1985). A synopsis of studies on the synergistic property of an insect baculovirus: a tribute to Edward A. Steinhaus. *J Invertebr Pathol* **45**, 125–138.

Tanada, Y. & Hukuhara, T. (1971). Enhanced infection of a nuclear-polyhedrosis virus in larvae of the armyworm, *Pseudaletia unipuncta*, by a factor in the capsule of a granulosis virus. *J Invertebr Pathol* **17**, 116–126.

Theilmann, D. A., Blissard, G. W., Bonning, B., Jehle, J., O'Reilly, D. R., Rohrman, G. F., Thiem, S. & Vlak, J. M. (2005). Family *Baculoviridae*. In *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, pp. 177–185. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. London: Elsevier.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

van Oers, M. M., Abma-Henkens, M. H. C., Herniou, E. A., de Groot, J. C. W., Peters, S. & Vlak, J. M. (2005). Genome sequence of *Chrysodeixis chalcites* nucleopolyhedrovirus, a baculovirus with two DNA photolyase genes. *J Gen Virol* **86**, 2069–2080.

Vlak, J. M. & Smith, G. E. (1982). Orientation of the genome of *Autographa californica* nuclear polyhedrosis virus: a proposal. *J Virol* **41**, 1118–1121.

Volkman, L. E., Blissard, G. W., Friesen, P., Keddie, B. A., Possee, R. & Theilmann, D. A. (1995). Family *Baculoviridae*. In *Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses*, pp. 104–113. Edited by F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo & M. D. Summers. Vienna: Springer.

Wang, P., Hammer, D. A. & Granados, R. R. (1994). Interaction of *Trichoplusia ni* granulosis virus-encoded enhancin with the midgut epithelium and peritrophic membrane of four lepidopteran insects. *J Gen Virol* **75**, 1961–1967.

Wang, J. W., Qi, Y. P., Huang, Y. X. & Li, S. D. (1995). Nucleotide sequence of a 1446 base pair *Sal*I fragment and structure of a novel early gene of *Leucania separata* nuclear polyhedrosis virus. *Arch Virol* **140**, 2283–2291.

Wang, P., Hammer, D. A. & Granados, R. R. (1997). Binding and fusion of *Autographa californica* nucleopolyhedrovirus to cultured insect cells. *J Gen Virol* **78**, 3081–3089.

Willis, L. G., Siepp, R., Stewart, T. M., Erlandson, M. A. & Theilmann, D. A. (2005). Sequence analysis of the complete genome of *Trichoplusia ni* single nucleopolyhedrovirus and the identification of a baculoviral photolyase gene. *Virology* **338**, 209–226.

Wormleaton, S., Kuzio, J. & Winstanley, D. (2003). The complete sequence of the *Adoxophyes orana* granulovirus genome. *Virology* **311**, 350–365.

Yamamoto, T. & Tanada, Y. (1980). Physiochemical 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.

Zhu, Y., Hukuhara, T. & Tamura, K. (1989). Location of a synergistic factor in the capsule of a granulosis virus of the armyworm, *Pseudaletia unipuncta*. *J Invertebr Pathol* **54**, 49–56.