Virus Genes (2011) 42:444–456
DOI 10.1007/s11262-011-0589-5

The genome of *Oryctes rhinoceros* nudivirus provides novel insight into the evolution of nuclear arthropod-specific large circular double-stranded DNA viruses

Yongjie Wang · Olaf R. P. Bininda-Emonds · Monique M. van Oers · Just M. Vlak · Johannes A. Jehle

Received: 28 November 2010 / Accepted: 21 February 2011 / Published online: 6 March 2011

Abstract The *Oryctes rhinoceros* nudivirus (OrNV) is a dsDNA virus with enveloped, rod-shaped virions. Its genome is 127,615 bp in size and contains 139 predicted protein-coding open reading frames (ORFs). In-depth genome sequence comparisons revealed a varying number of shared gene homologues, not only with other nudiviruses (NVs) and baculoviruses, but also with other arthropod-specific large dsDNA viruses, including the so-called Monodon baculovirus (MBV), the salivary gland hypertrophy viruses (SGHVs) and white spot syndrome virus (WSSV). Nudivirus genomes contain 20 baculovirus core gene homologues associated with transcription (*p47*, *lef-8*, *lef-9*, *lef-4*, *vlf-1*, and *lef-5*), replication (*dnapol* and *helicase*), virus structure (*p74*, *pif-1*, *pif-2*, *pif-3*, *19kda/pif-4*, *odv-e56/pif-5*, *vp91*, *vp39*, and *38K*), and unknown functions (*ac68*, *ac81*, and *p33*). Most strikingly, a set of homologous genes involved in peroral infection (*p74*, *pif-1*, *pif-2*, and *pif-3*) are common to baculoviruses, nudiviruses, SGHVs, and WSSV indicating an ancestral mode of infection in these highly diverged viruses. A gene similar to *polyhedrin/granulin* encoding the baculovirus occlusion body protein was identified in non-occluded NVs and in *Musca domestica* SGHV evoking the question of the evolutionary origin of the baculovirus *polyhedrin/granulin* gene. Based on gene homologies, we further propose that the shrimp MBV is an occluded member of the nudiviruses. We conclude that baculoviruses, NVs and the shrimp MBV, the SGHVs and WSSV share the significant number of conserved genetic functions, which may point to a common ancestry of these viruses.

Keywords *Oryctes rhinoceros* nudivirus · Genome · Monodon baculovirus · Salivary gland hypertrophy viruses · White spot syndrome virus · Peroral infection · Ancestry

Electronic supplementary material The online version of this article (doi:10.1007/s11262-011-0589-5) contains supplementary material, which is available to authorized users.

Y. Wang · J. A. Jehle
Laboratory for Biotechnological Crop Protection, Department of Phytopathology, Agricultural Service Center Palatinate (DLR Rheinpfalz), 67435 Neustadt a. d. Weinstrasse, Germany

Y. Wang
Laboratory of Marine and Food Microbiology, College of Food Science and Technology, Shanghai Ocean University, Shanghai, China

O. R. P. Bininda-Emonds
AG Systematics and Evolutionary Biology, Faculty V, Institute for Biology and Environmental Sciences (IBU), Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany

M. M. van Oers · J. M. Vlak
Laboratory of Virology, Wageningen University, 6708 PB Wageningen, The Netherlands

J. A. Jehle (✉)
Institute for Biological Control, Julius Kuehn Institute, Heinrichstr. 243, 64287 Darmstadt, Germany
e-mail: Johannes.Jehle@jki.bund.de

Introduction

Among the dsDNA viruses known to date, many rod-shaped, nuclear-replicating viruses with circular dsDNA genomes have been identified with host ranges solely restricted to arthropods. The largest and best studied group of insect viruses is the *Baculoviridae*, which have been isolated as occlusion bodies (OBs) from several hundred species of Lepidoptera, Hymenoptera, and Diptera. The genomes of about 50 baculoviruses have been completely

sequenced with genome sizes ranging from 86 to 179 kbp and encoding 90 to 180 ORFs [1]. All known baculoviruses share 30 gene orthologues (considered as the baculovirus core genes) associated with gene transcription, genome replication, oral infectivity, genome packaging, and virion assembly or hitherto unknown functions [2, 3]. Recently, phylogenetic relationship was inferred for three other large, non-occluded DNA viruses infecting *Helicoverpa zea* (Lepidoptera), *Gryllus bimaculatus* (Orthoptera), and *Oryctes rhinoceros* (Coleoptera). The genus name *Nudivirus* (NV) has been proposed to accommodate this group. The fact that nudiviruses share up to 20 baculovirus core genes strongly supports their common ancestry with baculoviruses [4–7]. Other arthropod-specific large dsDNA viruses include an occluded "baculovirus-like" virus (Monodon baculovirus, MBV) that has been reported for the tiger prawn *Penaeus monodon* (Crustacea) [8]. However, this virus has not yet been characterized in detail and only a very few genome fragments have been sequenced [9]. The White Spot Syndrome Virus (WSSV) is another penaeid (Crustacea) large DNA virus and causes severe economic damage to shrimp cultures all over the world [10]. With a non-occluded, tailed virion and a genome of about 300 kbp, WSSV has been classified in the family *Nimaviridae* [11]. Finally, two large DNA viruses causing salivary gland hypertrophy of the tsetse fly *Glossina pallidipes* (GpSGHV) and the housefly *Musca domestica* (MdSGHV) were shown to share at least eight homologues of baculovirus core genes [12, 13]. Despite some structural, biological, and genomic similarities with baculoviruses and nudiviruses, single gene trees based on DNA polymerase sequences as well as alignment-free proteome phylogenies implicated an association of SGHVs with vertebrate herpesviruses [12–15].

To obtain a better picture of the evolution and diversity of arthropod large dsDNA viruses, we sequenced the complete genome of the *O. rhinoceros* NV (OrNV) [16]. As this was the first Coleopteran specific nudivirus to be sequenced, it should help to clarify the emerging picture of nudivirus genome plasticity. OrNV was originally discovered in 1963 in Malaysia and has achieved great success as effective bio-control agents of the rhinoceros beetle, *O. rhinoceros* (Coleoptera: Scarabaeidae), a severe pest in coconut and oil palm cultivations in South-East Asia and the Pacific Islands [17]. Starting with six OrNV infected, field-collected beetles, which had provided a few nanograms of DNA, the OrNV DNA was then successfully amplified using multiple displacement amplification (MDA) to obtain enough DNA for sequencing the complete genome of OrNV [16]. Here, we report the in-depth analysis of the OrNV genome, including a comprehensive genome comparison with other large invertebrate DNA viruses. Our results show that the OrNV genome serves as an important missing link allowing us to identify novel gene homologues in a wide array of insect and crustacean viruses. These findings support our hypothesis of an evolutionary relationship of several families and unassigned species of nuclear arthropod-specific large DNA viruses.

Materials and methods

Virus, DNA purification, cloning and sequencing

OrNV isolation, DNA purification, DNA production using multiple displacement amplification, and DNA sequencing were described in [16].

DNA sequence analysis

Sequences were assembled using SeqMan (Lasergene 5.0 software, Dnastar, Inc.). The trace files were carefully checked by eye and minor mistakes corrected as needed. Methionine-initiated open reading frames (ORFs) encoding 50 amino acids or more and showing minimum overlap were predicted using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), GeneQuest and BioEdit software. ORFs with less than 50 amino acids were considered as putative genes in only the case of clear homology to known ORFs in other DNA viruses. The putative coding regions were numbered as OrNV ORFs.

Sequence similarity comparisons of all predicted ORFs against public databases were performed using the BLASTP, PSI-BLAST, and TBLASTN programs of NCBI. All OrNV ORFs were also searched against the COG database for functional annotations [18]. To avoid any missed predictions, the complete genomic sequence of OrNV was searched against the non-redundant protein databases using BLASTX. All ORFs were investigated for characteristic sequence signatures using the conserved domain search tool [19] (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), the conserved domain architecture retrieval tool [20] (http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps), and the InterProScan (http://www.ebi.ac.uk/InterProScan/) and PROSITE programs (http://au.expasy.org/prosite/). To annotate ORFs lacking significant motifs and/or sequence similarities to genes in the databases, protein structure-based predictions were performed using the PSIPRED Protein Structure Prediction Server [21] (http://bioinf.cs.ucl.ac.uk/psipred/) and the PredictProtein server (http://www.predictprotein.org/). Repeated and palindromic sequences were identified using the REPuter (http://bibiserv.techfak.uni-bielefeld.de/reputer/submission.html), Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.submit.options.html), SoftBerry (http://www.softberry.com/berry.phtml), and GeneQuest programs (Lasergene 5.0).

Sequence alignments

Multiple sequence alignments of individual genes were performed using any of T-Coffee [22], MUSCLE [23], ClustalW/X [24], MAFFT [25], and Kalign [26], and were manually refined as needed. Alignment quality was assessed by using MUMSA [26].

Nucleotide sequence accession number

The genomic sequence for OrNV referred to in this paper has been deposited in GenBank under the accession no. EU747721.

Results and discussion

General features of OrNV genome sequence

The genome of OrNV is 127,615 bp in length and has a G+C content of 42%. In total, 139 ORFs were predicted resulting in a coding density of 89% (Table 1). Four distinct types of repeated sequences (designated I, II, III, and IV) of 4.3 kb in total were identified dispersed along the genome (Table S1; Fig. S1). The repeats do not share any obvious sequence similarities to repeated sequences present in other large dsDNA viruses. Their function in OrNV remains to be clarified, but analogous to baculovirus homologous regions they may be involved as *cis*-acting elements in DNA replication (*hrs)* or enhancers of transcription [27].

Eighty-one ORFs (60%) could be assigned to gene homologues based on sequence similarity (and/or protein domain and structure matches) with sequences in GenBank (Table 1). Seventy-two had identities to genes of other large dsDNA viruses, particularly those of *G. bimaculatus* NV (GbNV), *H. zea* NV 1 (HzNV-1) and baculoviruses, whereas nine ORFs are cellular protein homologues. In-depth analysis of the OrNV ORFs now revealed that all three sequenced NVs have homologues with 20 of the 30 baculovirus core genes: those associated with late and very late gene transcription (*p47*, *lef-8*, *lef-9*, *lef-4*, *vlf-1*, and *lef-5*), replication (*dnapol* and *helicase*), virus structure (*p74*, *pif-1*, *pif-2*, *pif-3*, *ac68*, *vp91*, *vp39*, *38K*/*ac98*, *pif-4*/*19kda*/*ac96*, and *pif-5*/*odv-e56*), or those of unknown function (*ac81* and *p33*/*ac92*) (Table 2, Fig. S2).

Transcription apparatus of late and very late genes

In baculoviruses, early and delayed early genes are transcribed by the amanitin-sensitive host RNA polymerase II, whereas the virus encoded RNA polymerase is utilized for the transcription of late and very late genes [28]. Homologues of six gene products involved in late or very late transcription (*lef-4*, *lef-8*, *lef-9*, *p47*, *lef-5*, and *vlf-1*) were identified in OrNV. Careful re-examination of the recently sequenced genomes of HzNV-1 and GbNV also revealed now that all genes for the subunits of the baculovirus-like transcription apparatus were present in these viruses [5, 29]. Surprisingly, the OrNV sequence also facilitated the identification of the homologues of *lef-8*, *lef-9*, *lef-4*, and *lef-5* in SGHVs, as well as of *lef-9*, *vlf-1*, and *lef-5* in the partially sequenced genome of the occluded MBV from shrimp (Table 2). Although the evolutionary origin of the baculovirus RNA polymerase has yet to be elucidated [30], the presence of homologues of the RNA transcription apparatus represents a unifying feature of the baculovirus, NV and SGHV genomes, suggesting that these viruses all use a similar mode of late gene transcription.

DNA helicase

DNA helicase in baculoviruses is one of the 30 conserved core genes. It belongs to superfamily 3 helicase (SF3H) of the AAA^{+} ATPase class and is homologous to the D5R family helicase (SF3H) encoded in the nuclear cytoplasmic large DNA viruses (NCLDVs) [30]. When using OrNV ORF34 as a query sequence, BLAST searches showed that homologues of the baculovirus helicase were present not only in OrNV but also in HzNV-1 and GbNV, in the SGHVs (Tables 1, 2).

To avoid the potential pitfall of false positive annotation, careful sequence analysis of these proteins together with the D5R family helicase proteins from the NCLDVs was performed. Despite a relatively low overall sequence similarity, the conserved SF3H domain as characterised by walker A and two unknown motifs was observed for all NVs and SGHVs (Fig. S3). Nine amino acid residues, G(X)nGK(S/T)(X)nE(X)nK(X)nN(X)nD(X)nR, in these three motifs of the SF3H domain show >80% conservation among all the helicase protein sequences compared (Fig. S3A), indicating their essential role in the function of helicase. All nine amino acids are conserved in the protein sequences of the NVs; only one was substituted in the SGHVs (Fig. S3A). In NVs and SGHVs, two additional conserved motifs specific to the baculovirus helicase protein were observed upstream of the SF3H domain (Fig. S3B). The presence of multiple motifs and domains in the same relative order makes convergent evolution of these proteins in the different viral groups highly unlikely. Consequently, the DNA helicases in the above mentioned large DNA viruses are likely homologues, albeit ones that appear to have diverged very early from a common ancestral helicase protein given the global lack of sequence similarity.

Table 1 ORFs and their proteins predicted in OrNV

ORF	Position		Length		Best blast match				Feature[a]
	Start	End	nt	aa	ORF, protein encoded, or mass	Species	aa identity (% match identity)	E value	
1	*1*	*3843*	*3843*	*1280*	*DNA polymerase B*	*GbNV*	*228/701 (32)*	*1e−89*	*DNA_pol_B; DNA_pol_B_exo*
							103/334 (30)	*1e−33*	
2	3898	4917	1020	339	Trypsin-like serine protease	Anopheles culicifacies	72/245 (29)	9e−21	Tryp_SPc
3	4928	6043	1116	371	GrBNV_gp13	GbNV	75/269 (27)	8e−24	SP; TM
4	*6067*	*6564*	*498*	*165*	*Ac81 (GrBNV_gp14)*	*GbNV*	*68/144 (47)*	*6e−35*	*TM*
5	7152	6580	573	190					SP
6	8163	7198	966	321	Putative thymidylate synthase and pyrimidine hydroxymethylase	Nematostella vectensis	161/292 (55)	3e−87	TS_Pyrimidine_Hmase
7	9043	8210	834	277	Putative calcineurin-like phosphoesterase	Marinobacter algicola	60/234 (25)	3e−10	Metallophos
8	9302	9135	168	55					TM
9	9301	9966	666	221	NeseORF52	NeseNPV	29/114 (25)	7.2	
10	10197	10000	198	65					SP
11	10334	11122	789	262	Mitochondrial carrier protein	Pichia stipitis	73/295 (24)	7e−16	Mito_carr
12	12932	11151	1782	593	ODV-E66	SfMNPV	62/261 (23)	2e−07	Baculo_E66; TM; SP
13	13586	12984	603	200					
14	13798	13589	210	69					
15	*14471*	*13719*	*753*	*250*	*VP39 (GrBNV_gp64)*	*GbNV*	*61/236 (25)*	*3e−13*	
16	14539	15885	1347	448	Polyhedrin (GrBNV_gp65)	GbNV	114/371 (30)	1e−36	
17	*17041*	*15917*	*1125*	*374*	*PIF-2*	*GbNV*	*180/370 (48)*	*3e−100*	*Baculo_44*
18	17182	18285	1104	367	GrBNV_gp67	GbNV	101/402 (25)	3e−14	
19	18307	18618	312	103	Ac146 (TNSV_gp015)	TnSNPV	26/88 (29)	0.076	
20	*18530*	*19504*	*975*	*324*	*P47 (GrBNV_gp69)*	*GbNV*	*110/348 (31)*	*4e−33*	
21	20599	19574	1026	341					
22	20598	21329	732	243	GrBNV_gp72	GbNV	48/187 (25)	2e−06	
23	22214	21456	759	252	Putative guanylate kinase	GbNV	32/146 (21)	2.5	
24	22834	22253	582	193	GrBNV_gp75	GbNV	41/180 (22)	8e−06	
25	22833	24491	1659	552	GrBNV_gp76	GbNV	172/575 (29)	6e−58	
26	24551	25714	1164	387					Myosin_tail_1
27	26953	25748	1206	401	GrBNV_gp78	GbNV	78/289 (26)	1e−16	SP
28	26984	27820	837	278					
29	28611	27832	780	259	GrBNV_gp81	GbNV	37/180 (20)	0.039	SP
30	*28478*	*30598*	*2121*	*706*	*VLF-1*	*GbNV*	*107/373 (28)*	*7e−34*	*INT_phiLC3_C; Smc*
31	31030	30860	171	56					
32	31172	31357	186	61					
33	*32442*	*31675*	*768*	*255*	*19kda/PIF-4*	*GbNV*	*76/157 (48)*	*2e−37*	*TM; SP*
34	*32557*	*36279*	*3723*	*1240*	*DNA Helicase (GrBNV_gp88)*	*GbNV*	*281/1215 (23)*	*3e−71*	
35	36651	36358	294	97					
36	37089	36667	423	140					TM
37	38078	37149	930	309	Viral capsid associated protein (61 K = ORF1629 = PP78/83 = AcMNPV ORF9)	AcMNPV	20/73 (27)	0.25	Actin-binding WH2;
38	38391	38110	282	93					SP
39	39045	38476	570	189	GrBNV_gp93	GbNV	53/187 (28)	7e−12	
40	39019	39417	399	132	GrBNV_gp94	GbNV	27/113 (23)	1.4	

Table 1 continued

ORF	Position		Length		Best blast match				Feature[a]
	Start	End	nt	aa	ORF, protein encoded, or mass	Species	aa identity (% match identity)	E value	
41	39389	39730	342	113	GrBNV_gp95	GbNV	30/112 (26)	7e−07	TM; SP
42	*39706*	*40977*	*1206*	*401*	*LEF-4*	*GbNV*	*128/393 (32)*	*2e−51*	*LEF-4*
43	41058	40903	156	51					
44	41020	41610	591	196	GrBNV_gp97	GbNV	54/168 (32)	1e−13	
45	42594	41623	972	323	GrBNV_gp23	GbNV	86/292 (29)	1e−26	SP
46	44218	42647	1572	523	GrBNV_gp22	GbNV	104/443 (23)	6e−13	
47	45263	44427	837	278	GrBNV_gp19	GbNV	110/275 (40)	1e−45	
48	45849	45688	162	53					
49	46160	45945	216	71					
50	46249	46410	162	53					TM; SP
51	46577	48790	2214	737	RR1	AgseGV	182/537 (33)	4e−82	RNR_1; RNR_1_like; Ribonuc_red_lgC; Protein_Kinase_ATP
52	*49090*	*48854*	*237*	*78*	*LEF-5*	*GbNV*	*15/46 (32)*	*0.008*	*SP*
53	49780	49133	648	215	GrBNV_gp84	GbNV	56/214 (26)	1e−10	
54	50431	51750	1320	439	GrBNV_gp83	GbNV	66/331 (19)	1e−07	
55	51931	51734	198	65					TM; SP
56	53453	52110	1344	447					TM; SP
57	53584	54693	1110	369	Patatin-like phospholipase	Bacillus cereus G9241	63/237 (26)	2e−09	Patatin; SP
58	55360	54731	630	209	HZV_115	HzNV-1	26/104 (25)	1.5	Hemopexin
59	55962	55444	519	172	LEF-3 (GrBNV_gp86)	GbNV	36/150 (24)	0.077	
60	*57541*	*56066*	*1476*	*491*	*PIF-1*	*GbNV*	*187/427 (43)*	*6e−98*	*DUF686; TM; SP*
61	57626	57970	345	114	GrBNV_gp51	GbNV	21/68 (30)	0.032	TM
62	58334	58068	267	88					
63	58288	58509	222	73					TM; SP
64	*61303*	*58535*	*2769*	*922*	*LEF-8*	*GbNV*	*358/876 (40)*	*0*	
65	61771	61589	183	60					
66	62289	61993	297	98					SP
67	62218	62523	306	101					
68	62819	62613	207	68					TM; SP
69	63126	63788	663	220	FIC protein (filamentation induced by cAMP)	Monosiga brevicollis	39/103 (37)	1e−06	Fic
70	63906	64058	153	50					
71	64106	64783	678	225					
72	*65397*	*64987*	*411*	*136*	*AC68 (GrBNV_gp55)*	*GbNV*	*50/98 (51)*	*3e−24*	*TM*
73	65673	66458	786	261					
74	66586	67458	873	290	ORF1	Periplaneta fuliginosa densovirus	35/115 (30)	1e−05	
75	68601	67507	1095	364	DNA integrase/recombinases	GbNV	129/332 (38)	5e−62	INT_REC_C
76	68649	68804	156	51	GrBNV_gp58	Gb NV	20/49 (40)	0.003	TM; SP
77	68898	69266	369	122	Semaphorin-like protein	Tribolium castaneum	44/114 (38)	3e−16	Sema
78	69318	70580	1263	420	Semaphorin-like protein	Tribolium castaneum	195/395 (49)	2e−116	Sema; PSI (Plexin repeat)

Table 1 continued

ORF	Position		Length		Best blast match				Feature[a]
	Start	End	nt	aa	ORF, protein encoded, or mass	Species	aa identity (% match identity)	E value	
79	70591	70842	252	83	GrBNV_gp59	GbNV	19/49 (38)	3e−04	SERPIN (protease inhibitor I4); TM
80	71570	70980	591	196	GrBNV_gp60	GbNV	36/114 (31)	2e−06	
81	71703	71969	267	88					Glycosyl_Hydrol_F1_1 (Glycosyl hydrolases family 1 active site); SP
82	72179	72015	165	54					
83	72993	72400	594	197					Tachykinin family signature
84	73700	73197	504	167					TM; SP
85	73963	73787	177	58					
86	74529	73960	570	189	GrBNV_gp61	GbNV	71/190 (37)	8e−27	
87	*74492*	*75331*	*840*	*279*	*38K*	*GbNV*	*94/287 (32)*	*2e−27*	
88	75335	75565	231	76					TM; SP
89	75996	75709	288	95					
90	79528	76277	3252	1083	GrBNV_gp28	GbNV	177/841 (21)	3e−18	
91	80271	80092	180	59					
92	80946	81149	204	67					TM
93	81323	81126	198	65					
94	81568	81882	315	104					
95	81946	83199	1254	417	GrBNV_gp09	GbNV	66/204 (32)	3e−23	
96	*85038*	*83368*	*1671*	*556*	*LEF-9*	*GbNV*	*168/484 (34)*	*3e−70*	
97	84920	85708	789	262	Putative mRNA decapping enzyme 2	Drosophila melanogaster	88/289 (30)	6e−26	Dcp2p
98	86277	85816	462	153					
99	86519	86271	249	82	Ac120	AcMNPV	20/71 (28)	1.1	
100	86766	86927	162	53					
101	87351	87070	282	93					
102	87372	88574	1203	400	RR2	GbNV	86/289 (29)	1e−24	RNRR2
103	88760	89839	1080	359					TM; SP
104	90123	89836	288	95	GrBNV_gp62	GbNV	21/62 (33)	5e−04	
105	91245	90178	1068	355	GrBNV_gp43	GbNV	40/180 (22)	3.0	SP
106	*91338*	*93317*	*1980*	*659*	*VP91*	*GbNV*	*204/633 (32)*	*6e−79*	*CBM_14 (Chitin binding Peritrophin-A domain); TM; SP*
107	*93360*	*93974*	*615*	*204*	*PIF-3*	*GbNV*	*78/191 (40)*	*7e−39*	*DUF666; TM; SP*
108	96499	94037	2463	820	Putative helicase 2	GbNV	185/632 (29)	2e−73	
109	97076	97342	267	88					TM
110	97364	97573	210	69					
111	97830	97982	153	50					
112	98133	98285	153	50					TM; SP
113	*99664*	*98390*	*1275*	*424*	*P33 = Ac92, GrBNV_gp07*	*GbNV*	*80/268 (29)*	*3e−21*	*ERV_ALR (*ERV/ALR sulfhydryl oxidase domain)
114	99829	101244	1416	471	GrBNV_gp06	GbNV	78/334 (23)	2e−08	

Table 1 continued

ORF	Position		Length		Best blast match				Feature[a]
	Start	End	nt	aa	ORF, protein encoded, or mass	Species	aa identity (% match identity)	E value	
115	*101455*	*102693*	*1239*	*412*	*ODV-E56/PIF-5*	*GbNV*	*155/392 (39)*	*3e−74*	*TM*
116	103751	102690	1062	353	GrBNV_gp33	GbNV	75/262 (28)	1e−14	SP
117	104472	103804	669	222	GrBNV_gp34	GbNV	68/188 (36)	5e−20	
118	104818	105609	792	263	GrBNV_gp35	GbNV	81/275 (29)	4e−21	
119	105670	106800	1131	376	GrBNV_gp36	GbNV	90/325 (27)	7e−25	
120	108907	106862	2046	681	GrBNV_gp37	GbNV	114/538 (21)	2e−14	
121	110077	109004	1074	357	DNA ligase	HzNV-1	71/269 (26)	6e−13	DNA_Ligase_A_M; DNA_Ligase_A1 (ATP-dependent DNA ligase AMP-binding site)
122	110108	110590	483	160	GrBNV_gp39	GbNV	32/128 (25)	0.011	
123	111023	110595	429	142	GrBNV_gp41	GbNV	31/105 (29)	2e−06	SP
124	112251	111004	1248	415					
125	113333	112314	1020	339	GrBNV_gp44	GbNV	90/251 (35)	2e−28	
126	*113447*	*115660*	*2214*	*737*	*P74*	*GbNV*	*283/733 (38)*	*1e−141*	*Baculo_p74_N; Baculo_p74; TM*
127	116009	115842	168	55					
128	116920	116708	213	70					TM
129	117012	118100	1089	362					
130	118616	118407	210	69					
131	119191	118643	549	182					
132	119243	121021	1779	592	GrBNV_gp48	GbNV	139/590 (23)	1e−37	
133	121959	121102	858	285					TM
134	123225	122125	1101	366	IAP-3	CpGV	57/177 (32)	4e−29	BIR; ZF_RING_2
135	123392	123601	210	69					
136	124993	123908	1086	361					
137	125047	126510	1464	487	GrBNV_gp17	GbNV	128/456 (28)	4e−37	
138	126657	127124	468	155					
139	127458	127219	240	79					

Homologues to baculovirus core genes are in italics

SP signal peptides, *TM* transmembrane regions

[a] Protein families, domains and functional sites as well as associated patterns and profiles

Viral structural proteins

Baculoviruses produce two forms of enveloped virions: budded viruses (BVs) and occlusion derived viruses (ODVs) [27]. Infection of host insects is initiated in the midgut after the viral OBs are dissolved in the alkaline midgut lumen and the ODVs have been released. The ODVs infect the midgut columnar cells, which subsequently produce BVs to spread the infection to other tissues and cell types. All baculoviruses express six highly conserved "*per os* infectivity factors" (P74/PIF-0, PIF-1, PIF-2, PIF-3, PIF-4, ODV-E56/PIF-5) that are essential for ODV infectivity of midgut cells but dispensable for BV infectivity [31–33]. Most likely, P74, PIF-1, and PIF-2 function as attachment/fusion proteins for ODVs, whereas PIF-3 mediates another unidentified but essential event during primary infection [32]. Recently, it was shown that baculovirus PIF-1, PIF-2, and PIF-3 form a stable complex present on the ODV envelope to which P74 is associated [34]. PIF-4 may also act as the specific binding of ODV to midgut cells or as the cofactor for mediating ODV entry [31]. ODV-E56 (PIF-5) was

Table 2 Homologues of baculovirus core genes present in nudiviruses, SGHVs and other related viruses

Function	Name	Baculovirus (53)	Nudivirus (3)				SGHV (2)		Nimavirus (1)	Bracovirus (2)[a]	
		AcMNPV	OrNV	GbNV	HzNV-1	MBV	GpSGHV	MdSGHV	WSSV	CcBV	CiBV
Transcription	*p47*	40	20	69	75	n.d.	–	–	–	*Ccp47*	–
	lef-8	50	64	49	90	n.d.	40	70	–	*Cclef-8*	–
	lef-9	62	96	24	75	ABX44704	33 + 32	74	–	–	–
	vlf-1	77	30	80	121	ABX44703	–	–	–	–	*Civlf-1*
	lef-4	90	42	96	98	n.d.	51	87	–	–	*Cilef-4*
	lef-5	99	52	85	101	ABX44700	87	61	–	*Cclef-5*	–
Replication	*dnapol*	65	1	12	131	n.d.	79	1	514	–	–
	helicase	95	34	88	104	n.d.	45	83		–	–
Structural proteins											
Oral infectivity	*pif-2*	22	17	66	123	n.d.	53	89	35	–	*Cipif-2*
	pif-4/19kda	96	33	87	103	n.d.	–	–	–	*Cc19kda*	*Ci19kda*
	pif-3	115	107	3	88	n.d.	76	106	306	*Ccpif-3*	–
	pif-1	119	60	52	55	n.d.	102	29	209	–	*Cipif-1*
	p74	138	126	45	11	n.d.	1	39	115	*Ccp74*	*Cip74*
	pif-5/odv-e56	148	115	5	76	n.d.	–	–	–	*Ccodv-e56*	*Ciodv-e56*
Packaging, assembly, and morphogenesis	*ac68*	68	72	55	74	n.d.	–	–	–	–	–
	38K	98	87	1	10	ABX44705	–	–	–	*Cc38K*	*Ci38K*
Unknown function	*vp91/p95*	83	106	2	46	n.d.	–	–	–	–	*Civp91*
	vp39	89	15	64	89	n.d.	–	–	–	*Ccvp39*	*Civp39*
Unknown function	*ac81*	81	4	14	33	n.d.	78	108	–	–	
	p33	92	113	7	13	n.d.	72	102	327	–	–
Number of baculovirus core gene homologues			20				12		6	14	

Numbers represent the ORF number according to GenBank. MBV was partially sequenced and protein IDs are given. For each group of viruses, the number in parentheses represents the total number of viral genome sequences currently available in GenBank

[a] Genes are present in the genome of bracovirus host

recently identified to be required for midgut infection of AcMNPV [33].

A striking feature of all the NV genomes including OrNV is that they encode homologues of all six baculovirus *pif* genes (Table 2, Fig. S4). Additionally, four *pif* homologues, *pif-0*, *pif-1*, *pif-2*, and *pif-3*, have been recently detected in the GpSGHV and MdSGHV genomes [12, 13]. Using the predicted amino acid sequences of the OrNV *pifs* in PSI-BLAST searches, we now also identified previously undetected *pif* homologues in WSSV (wsv35:VP110, wsv115:VP53B, wsv209:VP187, wsv306:VP39A) [35] (Table 2; Fig. S4). Baculovirus P74/PIF-0, PIF-1, PIF-2, and PIF-4 are present in the ODV envelope [31, 36, 37]. Proteomic analyses of viral structural proteins confirmed the presence of the PIFs in the envelopes of several baculoviruses and SGHVs [13, 38–41]. Their homologues in WSSV, albeit not previously recognized, were also localized in the virion envelope [42, 43]. Additionally, both VP187 (PIF-1 homologue) and VP110 (PIF-2 homologue) appear to attach to and interact with host cells in WSSV primary infection, akin to baculovirus PIF-1 and PIF-2 [44, 45].

The presence of these highly conserved envelope proteins suggests the usage of a similar and evolutionarily highly conserved cellular entry mechanism by baculovirus ODVs, nudiviruses, SGHVs, and WSSV and, thus, an ancient origin of the *pif* genes. However, it should be noted that NVs and SGHVs invade the host in addition to the major oral transmission route also by using vertical modes such as mating. Whether the *pif*s are involved in this alternative infection process remains unclear. Interestingly, homologues of the *pif* genes have also been detected recently in the proviral genome of the parasitic braconid wasp, housing symbiotic bracoviruses for its larvae to survive and develop within its lepidopteran host [46]. Transcripts of the *pif* genes and their expression products were detected in specific wasp tissue and on the envelope of bracovirus particles, respectively [46], indicating that the PIFs are critical components of the bracovirus virion structure and transmission cycle. However, although bracoviruses are likely derived from an ancestral NV, they are transmitted to the wasp's host by their coating on the egg surface rather than orally [47]. The cellular functions of the bracovirus PIF proteins remain unknown; presumably, they are involved in delivering immune suppressive bracovirus genes into Lepidopteran hemocytes. Taken together, deciphering the biological functions of all known *pif* homologues will shed light on the evolution of the *pif* genes as well as on host-virus interaction.

Among the other structural virion proteins, the baculovirus ODV-E66, albeit of unknown function, is conserved across lepidopteran-specific baculoviruses (both granuloviruses (GVs) and nucleopolyhedroviruses (NPVs)), but is not found in the dipteran- and hymenopteran-specific NPVs [48, 49]. Interestingly, OrNV ORF12 is homologous to *odv-e66* and its counterparts were also predicted in SGHVs [12, 13] but not in the other nudiviruses. To better understand the evolution and function of odv-e66, it is necessary to sample more Gamma- and Deltabaculoviruses, nudiviruses and SGHVs.

The putative *polyhedrin/granulin* (*polh/gran*) gene

In baculoviruses, the *polh/gran* genes encode the main matrix protein of the OB in which the ODVs are embedded. Analogous to bacterial spores, viral OBs provide remarkable environmental stability such that ODV viability can be maintained for years outside of the host larvae [27]. This protein, however, is encoded by two different *polh/gran* genes in baculoviruses: all Alpha-, Beta-, and Gammabaculoviruses use a similar 28–30 kDa protein (type A), whereas the OB protein of the dipteran-specific Deltabaculovirus is a 90 kDa protein (type B) [50].

Intriguing in several aspects is the identification of a homologue of the baculovirus type A *polh/gran* genes in the OrNV genome (ORF16) as well as in HzNV-1, GbNV, and MdSGHV (Fig. 1). OrNV ORF 16 shares up to 17% global amino acid identity with type A *polh/gran* genes. In particular, the region between amino acids 178 and 229 of the predicted OrNV ORF16 protein revealed the highest identity (up to 32%) to type A *polh/gran* genes. The function of the predicted *polh/gran* gene in NVs and SGHV remains unknown, given that these viruses have not been considered to date to produce OBs. However, incidental observations of atypical OBs in the midguts of infected *Oryctes* larvae and in the *bursa copulatrix* of corn ear worms infected with *H. zea* NV-2 call this assumption into question [51, 52]. As shown below, the crustacean MBV appears to be closely related to HzNV-1 and is also occluded. However, its OB gene encodes a 51 kDa protein and does not reveal any apparent similarity to any genes in GenBank (including baculovirus type A and B *polh/gran* gene) nor to the putative type A homologue present in both NVs and SGHVs, suggesting a distinct OB gene in MBV [53].

Deciphering the nature of possibly facultative OBs in NVs and SGHVs, and elucidating the function of their *polh/gran* homologues will be pivotal for understanding the evolutionary forces that have led to the formation of an obligatory OB in baculoviruses. However, if we assume the ancient presence of a homologous type A *polh/gran* gene in a common ancestor of baculoviruses and NVs, this gene must have been functionally replaced by non-orthologous OB genes in both the dipteran *Culex nigripalpis* (Cuni) NPV and the marine shrimp MBV.

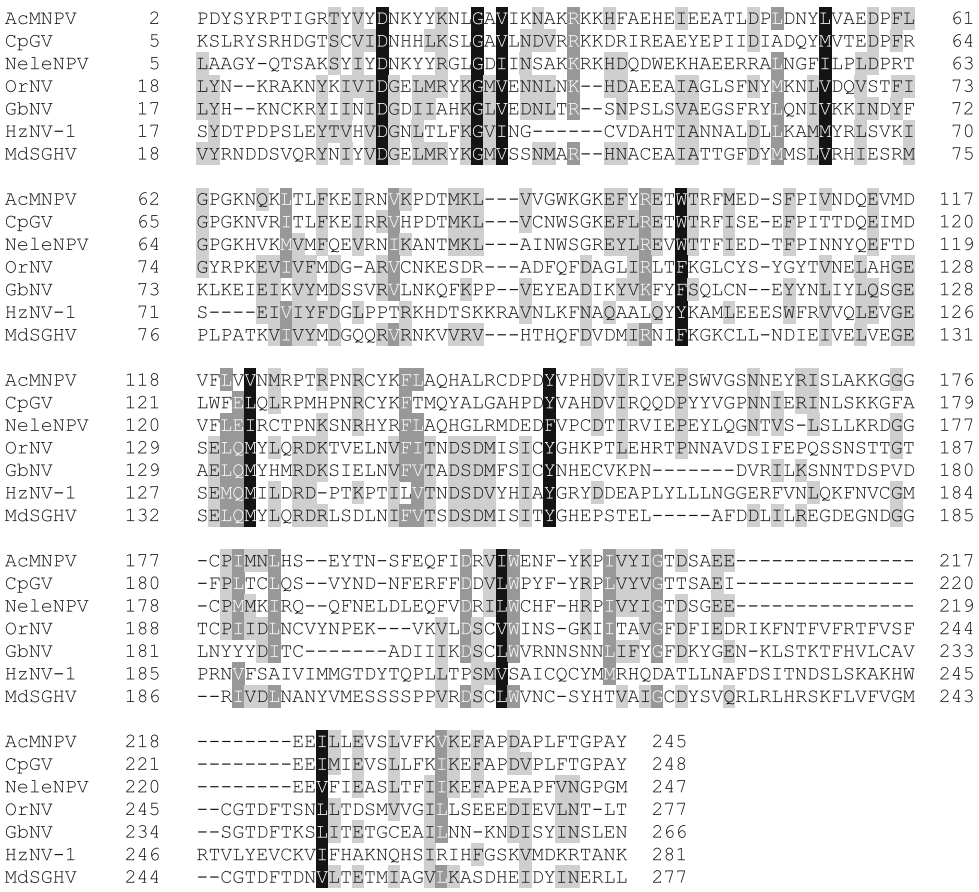

Fig. 1 Sequence alignment of the translated *polyhedrin/granulin* genes in baculoviruses and their putative homologues in nudiviruses and SGHV. The *numbers on left* and *right* sides of the aligned sequences are the positions of the first and last residues of the aligned regions in the corresponding protein sequences. The conservation of the aligned residues using BLOSUM-62 of amino acid substitution matrices is coloured as follows: *white residue against black background* (conservation, 100%), *white residue against gray background* (conservation, 80%), and *black residues against gray background* (conservation, 60%)

Other unknown function genes and gene clusters

A homologue of the baculovirus *p33* gene was also detected in OrNV and other NVs, as well as in SGHVs and WSSV (Fig. S5). Recently, the baculovirus protein P33 was confirmed to be a flavin adenine dinucleotide-linked sulfhydryl oxidase [54]. Our analyses also revealed a notable degree of sequence similarity between P33 and the NCLDV thiol oxidoreductase [30]. All these proteins share a common ERV_ALR sulfhydryl oxidase domain with predicted conserved secondary structural elements (Fig. S5). Thiol oxidoreductase mediates the formation of disulfide bridges between conserved cysteines in the structural proteins of NCLDV virions and plays a crucial role in the assembly and/or stability of these particles [55]. OrNV ORF74 is homologous to the non-structural protein NS3 from insect densoviruses (*Parvoviridae: Densovirinae*). Its presence in dsDNA and ssDNA viruses is notable, but it also has homologues in several baculoviruses [56].

Gene order is poorly conserved in all sequenced baculovirus and NV genomes. However, a conserved core gene cluster of the four genes *helicase*, *pif-4/19kda*, *38K*, and *lef-5* is found in all baculoviruses [48, 57] and a similar cluster of *helicase*, *pif-4/19kda*, and/or *lef-5* was detected in all three NV genomes (but not in SGHVs).

The shrimp MBV represents a nudivirus

BLASTP searches using OrNV ORFs further revealed that a number of homologues are present in the partially sequenced genome of shrimp MBV (21,150 bp in total; GenBank accession nos. EU246943, EU246944, EF458632, AY819785). When using the annotated shrimp MBV ORFs best hits were frequently found with HzNV-1 (Table 3). Phylogenetic analyses of the homologues of baculovirus core genes *lef-9*, *vlf-1*, *lef-5*, *38K*, consistently placed MBV together with the non-occluded HzNV-1 (Fig. S2). Given that the sequences of seven other MBV and HzNV-1 ORFs are also highly similar (Table 3), it is strongly suggested to consider MBV as an occluded member of the NVs.

Shared baculovirus core genes suggest a common ancestry

Summarizing the data mining using the OrNV ORFs as query and performed on other dsDNA viruses (Table 2),

Table 3 BlastP similarity match of MBV protein sequences

MBV protein sequence	Best blast match		
	ORF or protein encoded	Species	E value
ABX44711	–	–	–
ABX44710	HZV_51	HzNV-1	1e-28
ABX44709	–	–	–
ABX44708	–	–	–
ABX44707	–	–	–
ABX44706	HZV_93	HzNV-1	0.69
ABX44705	*HZV_10 (38K)*	*HzNV-1*	*3e-15*
ABX44704	*DNA-directed RNA polymerase (LEF-9)*	*HzNV-1*	*8e-60*
ABX44703	*Very late factor 1 (VLF-1)*	*HzNV-1*	*1e-10*
ABX44702	HZV_144 (Integrase)	HzNV-1	2e-29
ABX44701	–	–	–
ABX44700	*Late expression factor 5 (LEF-5)*	*Spodoptera frugiperda MNPV*	*0.017*
ABX44699	–	–	–
ABX44698	–	–	–
ABX44697	HZV_141	HzNV-1	5e-07
ABX44696	–	–	–
ABO65260	HZV_115	HzNV-1	3e-04
ABO65259	HZV_118	HzNV-1	0.29
ABO65258	–	–	–
ABO65257	–	–	–
ABO38811	P51	HzNV-1	2e-10
ABO38810	–	–	–

MBV was partially sequenced and protein IDs are given. The homologues to baculovirus core genes are indicated in italic typeface

twenty baculovirus core gene homologues were identified in nudiviruses and MBV, 12 in SGHVs, and six in WSSV, respectively. In addition, homologues of fourteen baculovirus core genes were present in bracoviruses [*Polydnaviridae*], which are considered to be derived from an ancient nudivirus [46]. Most of these baculovirus core gene homologues are unique to these virus groups considered in this study. Therefore, phylogenetic analyses appeared to be complicated by the lack of sufficient homologous genes in other dsDNA viruses making it impossible to define a suitable outgroup. Sequence conservation observed in the alignments of the identified core gene homologues was poor, albeit significant. It rendered a gobal phylogenetic inference difficult and unrobust, as different phylogenetic methods produced conflicting results, as it is typical for weakly conserved sequences (data not shown). Interestingly, when an alignment-free phylogenetic analysis based on whole virus proteomes was performed, WSSV clustered with SGHVs, which is partially in agreement with the presented hypothesis of a common ancestry of baculoviruses, nudiviruses, SGHVs, and WSSV [15]. However, the same study placed SGHV and WSSV not with baculoviruses and nudiviruses but within the herpesviruses, though there is not any further evidence of a relationship between SGHVs or WSSV and herpesviruses, when considering structural, biological and other genome features. Consequently, we consider the shared gene content of baculoviruses, nudiviruses, SGHVs, and WSSV to be a better witness of a proposed common ancestry of these viruses. The alternative explanation, horizontal gene transfer, seems less likely. Single gene tree phylogenies did not reveal convincing evidence for horizontal gene transfer among these virus groups (Fig. S2). Also, the number of conserved genes among these groups is considerable and the virus lineages in question are highly divergent and infect crustaceans as well as hemi- and holometabolous insects in very different environments and ecological niches. Other insect DNA viruses such as the entomopoxviruses, iridoviruses, and ascoviruses have developed infection modes different from those under study herein, gene convergence driven by the necessity to find the same mode to infect insect and crustacean cells seems highly unlikely. Further shared characters are the rod-shaped enveloped virions, the circular dsDNA genomes with homologous repeat regions, and the virus replication in the host cell nucleus. Considering these shared characters and our genome wide comparisons, we conclude that baculoviruses, nudiviruses, hytrosaviruses, and WSSV are more

closely related to each other than to any other known eukaryotic large dsDNA virus.

Acknowledgements This work was supported by a grant of the Deutsche Forschungsgemeinschaft (Je245-7) to J.A.J. M.v.O. was supported by a MEERVOUD grant from the Research Council of Earth and Life Sciences (ALW) with financial aid from the Netherlands Organization for Scientific Research (NWO) (Project 836.05.070). Dwight Lynn is acknowledged for helpful comments on the manuscript.

References

1. M.M. van Oers, J.M. Vlak, Curr. Drug Targets **8**, 1051–1068 (2007)
2. E.A. Herniou, T. Luque, X. Chen, J.M. Vlak, D. Winstanley, J.S. Cory, D.R. O'Reilly, J. Virol. **75**, 8117–8126 (2001)
3. C.B. McCarthy, D.A. Theilmann, Virology **375**, 277–291 (2008)
4. Y. Wang, J.P. Burand, J.A. Jehle, Virol. Sinica **22**, 128–136 (2007)
5. Y. Wang, R.G. Kleespies, A.M. Huger, J.A. Jehle, J. Virol. **81**, 5395–5406 (2007)
6. Y. Wang, M.M. van Oers, A.M. Crawford, J.M. Vlak, J.A. Jehle, Arch. Virol. **152**, 519–531 (2007)
7. Y. Wang, J.A. Jehle, J. Invertebr. Pathol. **101**, 187–193 (2009)
8. D.V. Lightner, R.M. Redman, J. Invertebr. Pathol. **38**, 299–302 (1981)
9. J. Mari, J.-R. Bonami, B. Poulos, D. Lightner, Dis. Aquat. Organ. **16**, 207–215 (1993)
10. J.H. Leu, F. Yang, X. Zhang, X. Xu, G.H. Kou, C.F. Lo, Curr. Top. Microbiol. Immunol. **328**, 197–227 (2009)
11. J.M. Vlak, J.R. Bonami, T.W. Flegel, G.H. Kou, D.V. Lightner, C.F. Lo, P.C. Loh, P.W. Walker, in *Virus Taxonomy: The Eighth Report of the International Committee on Taxonomy of Viruses*, ed. by C.M. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger, L.A. Ball (Elsevier Academic Press, New York, 2005), pp. 187–192
12. A.M. Abd-Alla, F. Cousserans, A.G. Parker, J.A. Jehle, N.J. Parker, J.M. Vlak, A.S. Robinson, M. Bergoin, J. Virol. **82**, 4595–4611 (2008)
13. A. Garcia-Maruniak, J.E. Maruniak, W. Farmerie, D.G. Boucias, Virology **377**, 184–196 (2008)
14. A.M. Abd-Alla, J.M. Vlak, M. Bergoin, J.E. Maruniak, A.G. Parker, J.P. Burand, J.A. Jehle, D.G. Boucias, Arch. Virol. **154**, 909–918 (2009)
15. G.A. Wu, J.R. Jun, G.E. Sims, S.H. Kim, Proc. Natl. Acad. Sci. USA **106**, 12826–12831 (2009)
16. Y. Wang, R.G. Kleespies, M.B. Ramle, J.A. Jehle, J. Virol. Methods **152**, 106–108 (2008)
17. T.A. Jackson, A.M. Crawford, T.R. Glare, J. Invertebr. Pathol. **89**, 91–94 (2005)
18. R.L. Tatusov, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, B.S. Rao, S. Smirnov, A.V. Sverdlov, S. Vasudevan, Y.I. Wolf, J.J. Yin, D.A. Natale, BMC Bioinformatics **4**, 41 (2003)
19. A. Marchler-Bauer, J.B. Anderson, M.K. Derbyshire, C. DeWeese-Scott, N.R. Gonzales, M. Gwadz, L. Hao, S. He, D.I. Hurwitz, J.D. Jackson, Z. Ke, D. Krylov, C.J. Lanczycki, C.A. Liebert, C. Liu, F. Lu, S. Lu, G.H. Marchler, M. Mullokandov, J.S. Song, N. Thanki, R.A. Yamashita, J.J. Yin, D. Zhang, S.H. Bryant, Nucleic Acids Res. **35**, D237–D240 (2007)
20. L.Y. Geer, M. Domrachev, D.J. Lipman, S.H. Bryant, Genome Res. **12**, 1619–1623 (2002)
21. L.J. McGuffin, D.T. Jones, Bioinformatics **19**, 874–881 (2003)
22. C. Notredame, D.G. Higgins, J. Heringa, J. Mol. Biol. **302**, 205–217 (2000)
23. R.C. Edgar, BMC Bioinformatics **5**, 113 (2004)
24. R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson, Nucleic Acids Res. **31**, 3497–3500 (2003)
25. K. Katoh, K. Misawa, K. Kuma, T. Miyata, Nucleic Acids Res. **30**, 3059–3066 (2002)
26. T. Lassmann, E.E. Sonnhammer, Nucleic Acids Res. **33**, 7120–7128 (2005)
27. D.A. Theilmann, G.W. Blissard, B. Bonning, J.A. Jehle, D.R. O'Reilly, G.F. Rohrmann, S. Thiem, J.M. Vlak, in *Virus Taxonomy: The Eighth Report of the International Committee on Taxonomy of Viruses*, ed. C.M. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger, L.A. Ball (Elsevier Academic Press, New York 2005), pp. 177–185
28. L.A. Guarino, M.D. Summers, J. Virol. **57**, 563–571 (1986)
29. C.H. Cheng, S.M. Liu, T.Y. Chow, Y.Y. Hsiao, D.P. Wang, J.J. Huang, H.H. Chen, J. Virol. **76**, 9024–9034 (2002)
30. L.M. Iyer, S. Balaji, E.V. Koonin, L. Aravind, Virus Res. **117**, 156–184 (2006)
31. M. Fang, Y. Nie, S. Harris, M.A. Erlandson, D.A. Theilmann, J. Virol. **83**, 12569–12578 (2009)
32. T. Ohkawa, J.O. Washburn, R. Sitapara, E. Sid, L.E. Volkman, J. Virol. **79**, 15258–15264 (2005)
33. W.O. Sparks, R.L. Harrison, B.C. Bonning, Virology **409**, 69–76 (2011)
34. K. Peng, M.M. van Oers, Z. Hu, J.W. van Lent, J.M. Vlak, J. Virol. **84**, 9497–9504 (2010)
35. F. Yang, J. He, X. Lin, Q. Li, D. Pan, X. Zhang, X. Xu, J. Virol. **75**, 11811–11820 (2001)
36. P. Faulkner, J. Kuzio, G.V. Williams, J.A. Wilson, J. Gen. Virol. **78**, 3091–3100 (1997)
37. S. Gutiérrez, I. Kikhno, M. López Ferber, J. Gen. Virol. **85**, 331–341 (2004)
38. S.C. Braunagel, W.K. Russell, G. Rosas-Acosta, D.H. Russell, M.D. Summers, Proc. Natl. Acad. Sci. USA **100**, 9797–9802 (2003)
39. F. Deng, R. Wang, M. Fang, Y. Jiang, X. Xu, H. Wang, X. Chen, B.M. Arif, L. Guo, Z. Hu, J. Virol. **81**, 9377–9385 (2007)
40. H. Kariithi, I.A. Ince, S. Boeren, J. Vervoort, M. Bergoin, M.M. van Oers, A.M. Abd-Alla, J.M. Vlak, J. Gen. Virol. **91**, 3065–3074 (2010)
41. O.P. Perera, T.B. Green, S.M. Stevens Jr, S. White, J.J. Becnel, J. Virol. **81**, 4585–4590 (2007)
42. Z. Li, Q. Lin, J. Chen, J.L. Wu, T.K. Lim, S.S. Loh, X. Tang, C.L. Hew, Mol. Cell Proteomics **6**, 1609–1620 (2007)
43. X. Xie, L. Xu, F. Yang, J. Virol. **80**, 10615–10623 (2006)
44. L. Li, S. Lin, F. Yang, J. Gen. Virol. **87**, 1909–1915 (2006)
45. D.F. Li, M.C. Zhang, H.J. Yang, Y.B. Zhu, X. Xu, Virology **368**, 122–132 (2007)
46. A. Bézier, M. Annaheim, J. Herbinière, C. Wetterwald, G. Gyapay, S. Bernard-Samain, P. Wincker, I. Roditi, M. Heller, M. Belghazi, R. Pfister-Wilhem, G. Periquet, C. Dupuy, E. Huguet, A.N. Volkoff, B. Lanzrein, J.M. Drezen, Science **323**, 926–930 (2009)
47. S. Wyder, F. Blank, B. Lanzrein, J. Insect Physiol. **49**, 491–500 (2003)
48. E.A. Herniou, J.A. Olszewski, J.S. Cory, D.R. O'Reilly, Annu. Rev. Entomol. **48**, 211–234 (2003)
49. H.A. Lauzon, A. Garcia-Maruniak, P.M. Zanotto, J.C. Clemente, E.A. Herniou, C.J. Lucarotti, B.M. Arif, J.E. Maruniak, J. Gen. Virol. **87**, 1477–1489 (2006)
50. O.P. Perera, S.M. Valles, T.B. Green, S. White, C.A. Strong, J.J. Becnel, J. Invertebr. Pathol. **91**, 35–42 (2006)

51. A.M. Huger, A. Krieg, in *Atlas of Invertebrate Viruses*, ed. by J.R. Adams, J.R. Bonami (CRC Press, Inc., Boca Raton, 1991), pp. 287–319
52. A.K. Raina, J.R. Adams, B. Lupiani, D.E. Lynn, W. Kim, J.P. Burand, E.M. Dougherty, J. Invertebr. Pathol. **76**, 6–12 (2000)
53. P. Chaivisuthangkura, C. Tawilert, T. Tejangkura, S. Rukpratanporn, S. Longyant, W. Sithigorngul, P. Sithigorngul, Virology **381**, 261–267 (2008)
54. C.M. Long, G.F. Rohrmann, G.F. Merrill, Virology **388**, 231–235 (2009)
55. L.M. Iyer, L. Aravind, E.V. Koonin, J. Virol. **75**, 11720–11734 (2001)
56. M. Lange, J.A. Jehle, Virology **317**, 220–236 (2003)
57. J.A. Jehle, H. Backhaus, J. Gen. Virol. **75**, 1815–1820 (1994)