Genomic analysis of Oryctes rhinoceros virus reveals genetic relatedness to Heliothis zea virus I

Article *in* Archives of Virology · February 2007

DOI: 10.1007/s00705-006-0872-2 · Source: PubMed

5 authors, including:

Yongjie Wang
Shanghai Ocean University
54 PUBLICATIONS **1,177** CITATIONS

SEE PROFILE

Monique M Van Oers
Wageningen University & Research
246 PUBLICATIONS **2,807** CITATIONS

SEE PROFILE

Allan Muirhead Crawford
AgResearch
158 PUBLICATIONS **4,752** CITATIONS

SEE PROFILE

Just M Vlak
Wageningen University & Research
880 PUBLICATIONS **14,844** CITATIONS

SEE PROFILE

Arch Virol (2007) 152: 519–531
DOI 10.1007/s00705-006-0872-2
Printed in The Netherlands

Archives of Virology

Genomic analysis of Oryctes rhinoceros virus reveals genetic relatedness to Heliothis zea virus 1

Y. Wang[1], **M. M. van Oers**[2], **A. M. Crawford**[3], **J. M. Vlak**[2], and **J. A. Jehle**[1]

[1] Laboratory for Biotechnological Crop Protection, Department of Phytopathology, Agricultural Service Centre Palatinate (DLR Rheinpfalz), Neustadt an der Weinstrasse, Germany
[2] Laboratory of Virology, Wageningen University, Wageningen, The Netherlands
[3] AgResearch Invermay Agricultural Centre, Mosgiel, New Zealand

Received July 19, 2006; accepted September 28, 2006; published online November 16, 2006

Summary

Oryctes rhinoceros virus (OrV) is an unassigned invertebrate dsDNA virus with enveloped and rod-shaped virions. Two cloned *Pst*I fragments, C and D, of OrV DNA have been sequenced, consisting of 19,805 and 17,146 bp, respectively, and comprising about 30% of the OrV genome. For each of the two fragments, 20 open reading frames (ORFs) of 150 nucleotides or greater with no or minimal overlap were predicted. Ten of the predicted 40 ORFs revealed significant similarities to Heliothis zea virus 1 (HzV-1) ORFs, of which five, *lef-4, lef-5*, *pif-2*, *dnapol* and *ac81*, are homologues of conserved core genes in the family *Baculoviridae*, and one is homologous to baculovirus *rr1*. A baculovirus *odv-e66* homologue is also present in OrV. Five ORFs encode proteins homologous to cellular thymidylate synthase (TS), patatin-like phospholipase, mitochondrial carrier protein, Ser/Thr protein phosphatase, and serine protease, respectively. TS is phylogenetically related to those of eukarya and nucleo-cytoplasmic large dsDNA viruses. However, the remaining 25 ORFs have poor or no sequence matches with the current databases. Both the gene content of the sequenced fragments and the phylogenetic analyses of the viral DNA polymerase suggest that OrV is most closely related to HzV-1. These findings and the re-evaluation of the relationship of HzV-1 to baculoviruses suggest that a new virus genus, Nudivirus, should be established, containing OrV and HzV-1, which are genetically related to members of the family *Baculoviridae*.

Introduction

The Oryctes rhinoceros virus (OrV) was discovered in the 1960s in Malaysia and has been effectively used to control the rhinoceros beetle (*O. rhinoceros*) in coconut and oil palm in Southeast Asia and the Pacific until the present day [17, 21]. It is a classical example of successful inoculation and long-

The GenBank accession numbers of the sequences reported in this paper are DQ665870 and DQ665871.

Author's address: Dr. Johannes A. Jehle, Laboratory for Biotechnological Crop Protection, Department of Phytopathology, Agricultural Service Center Palatinate (DLR Rheinpfalz), Breitenweg 71, 67435 Neustadt an der Weinstrasse, Germany. e-mail: Johannes.Jehle@dlr.rlp.de

term control of an insect pest, but surprisingly little is known about its genetics.

The virus consists of rod-shaped virions and replicates in the nuclei of infected cells [17, 35, 36]. On the basis of its (ultra) structure, OrV was previously considered to be a so-called non-occluded baculovirus (NOB) [18]. Due to the lack of occlusion bodies, it was later removed from the family *Baculoviridae* [29] and orphaned. There are a number of other rod-shaped non-occluded dsDNA viruses found in invertebrates [18]. The most notable example is Heliothis zea virus 1 (HzV-1), which has been identified in cultured insect cells but has not yet been observed in insects [2]. The genome of HzV-1 has been sequenced completely, but only limited genetic similarities with baculoviruses were found [5]. Taxonomically, HzV-1 remains an unassigned virus [30]. Another formerly considered NOB, the white spot syndrome virus of shrimp (WSSV), also has rod-shaped nucleocapsids, which are wrapped into an envelope giving it an ovoid appearance. It was recently classified as belonging to the novel family *Nimaviridae* [45]. The most characteristic feature of WSSV is the appearance of a tail-like appendage. A similar appendage is also present on OrV particles [18]. Thus, the question remained, what is the taxonomic position of OrV.

OrV contains a double-stranded DNA genome of about 130 kilobase pairs [7], fitting in the size range of baculoviruses (80–180 kbp). Genetic variation exists, as evidenced by restriction fragment length polymorphism of OrV field isolates [40]. A physical map of OrV has been constructed for various restriction enzymes, showing that the DNA is circular and contains regions with reiterated sequences [7, 32], but sequence information on the viral DNA is very limited. A segment of 4 kb was sequenced to design PCR primers for the detection of OrV [41]. It bears no significant similarity to other sequences in the published databases. Therefore, we decided to sequence portions of the OrV genome until sufficient sequence information was obtained to allow gene comparisons and phylogenetic analyses. For this purpose, the *Pst*I-C and -D fragments of OrV DNA were cloned and sequenced.

Materials and methods

Virus, DNA cloning and sequencing

The O. rhinoceros virus (OrV), strain PV505, was used. Its DNA purification and cloning of *Pst*I fragments into the pBR328 vector were described previously [7]. For sequencing, the plasmids containing *Pst*I-C and -D were subjected to EZ::TN transposon mutagenesis (Epicentre). Positive clones with the antibiotic resistance from both the EZ::TN transposon (Kan^r) and the vector pBR328 (Tet^r and Chl^r) were selected and grown on LB medium according to the manufacturer's instructions. Purified plasmids (Qiagen plasmid Mini Kit) were analysed using restriction endonucleases (RENs) to map the transposon insertion site within the cloned OrV *Pst*I fragments. Sequencing was done by using the primers binding sites at nt positions 52–75 and 1907–1931 in the EZ::TN transposon. To close sequencing gaps, primer walking was performed using OrV sequence-specific oligonucleotides. Sequencing was performed using an ABI 3730 automated 48-capillary sequencer (Applied Biosystems) with the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit version 3.1 by GENterprise (Mainz, Germany) (http://www.genterprise.de/).

DNA sequence analysis

Sequences were assembled using SeqMan (Lasergene 5.0 software, Dnastar, Inc.). The trace files were checked by eye, and minor mistakes were corrected if necessary. The simulated restriction digestion of the sequences was done with the GeneQuest program integrated in Lasergene software. Methionine-initiated ORFs encoding 50 amino acids or more and showing minimum overlap (<40 bp) were identified using the ORF Finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and GeneQuest program. ORFs with less than 50 amino acids were only considered as putative genes in cases of clear homology to ORFs in other dsDNA viruses. Homology comparison of all of the possible ORFs to public databases was done using the BLASTP, TBLASTN and PSI-BLAST programs [1, 42]. The following criteria were considered to assign a putative homologue to the OrV ORFs: (i) A BLASTP search showed a dsDNA virus match with an E value of 0.1 or less; (ii) Amino acid identity to dsDNA virus homologue was 20% or greater based on MegAlign clustalW analysis of entire ORFs; (iii) A conserved domain was found. The putative coding regions were numbered as OrV ORFs.

The sequences within 300 bp upstream of the start codon of each ORF were analysed for baculovirus-like early and late transcription promoters. Early promoters were defined as either TATAWAW or TATAWTW (W = A or T) alone (designed "E" in Table 1) or "E" with initiator motif CAKT (K = T or G), 20–40 nucleotide downstream ("E+" in Table 1). Late promoters contained the sequence DTAAG (D = A, G or T). All ORFs were investigated for character-

Table 1. ORFs predicted in *Pst*I-C and -D fragments of the Oryctes rhinoceros virus

ORF	Strand[1]	Position		Length		Intergenic distance (bp)[2]	% A+T	Promoter[3]		Blast best match					Signature[4]
		Start	End	nt	aa			E/E+	L	Name	Species	BlastP score	Length (aa % identity)	E value	
OrV *Pst*I-C															
C1	+	14	469	456	151	n.d.	60.3								
C2	+	516	1640	1125	374	46	56.4	−58	−172	*per-os* infectivity factor 2	Heliothis zea virus 1	155.0	101/356 (28)	4e-36	SP; TM
C3	−	3018	1672	1347	448	31	56.2	−17		HzV-1_Orf68	Heliothis zea virus 1	55.1	39/139 (28)	1e-05	PIN domain-like
C4	+	3086	3838	753	250	67	56.5	−35; −142 −247	−166						
C5	+	3971	4573	603	200	132	58.9	−93							Tubulin chaperone cofactor A
C6	+	4625	6406	1782	593	51	57.6	−20; −38		occlusion derived virus envelope protein 66	Cryptophlebia leucotreta granulovirus	57.8	73/336 (21)	2e-06	SP; TM
C7	−	7223	6435	789	262	28	59.7	−174; −294	−131	mitochondrial carrier protein	*Ashbya gossypii*	94.4	85/286 (29)	7e-18	Mito_carr
C8	+	7360	7560	201	66	136	65.1		−175						SP
C9	−	8254	7589	666	221	28	60.2	−113 (−76)	−236						
C10	+	8253	8420	168	55	−2	63.1		−262						TM
C11	+	8512	9345	834	277	91	55.4	−92 (−80) −243		ser/thr protein phosphatase family protein	*Croceibacter atlanticus*	65.1	49/192 (25)	5e-09	Metallophos
C12	+	9396	10361	966	321	50	55.7	−9, −11, −13, −15, −17, −19 (+18)		thymidylate synthase	*Drosophila melanogaster*	306.0	155/293 (52)	2e-81	Thymidylat_synt
C13	+	10407	10979	573	190	45	57.7	−46; −86							SP
C14	−	11493	10996	498	165	16	61.8			Ac81	Heliothis zea virus 1	63.2	43/155 (27)	6e-09	TM
C15	−	12632	11517	1116	371	23	58.3	−56 (−13)							ASP_RICH; SP; TM
C16	−	13662	12643	1020	339	10	55.7	−248 (−222) −282	−127	serine protease	*Anopheles culicifacies*	95.9	69/244 (28)	4e-18	Pept_Ser_Cys; Trypsin; CHYMOTRYPSIN; SP; TM
C17	−	17559	13717	3843	1280	54	59.0	−17; −77	−23	DNA polymerase B	Heliothis zea virus 1	147.0	257/1218 (21)	4e-33	DNA_POLYMERASE_B; DNA_pol_B_exo; RNaseH_fold; ASP_RICH
C18	+	17717	17956	240	79	157	57.9	−80	−104						
C19	−	18513	18046	468	155	89	56.9	−115 (−70)	−88						
C20	−	19523	18660	864	287	146	59.3			HzV-1_Orf51	Heliothis zea virus 1	42.0	43/189 (22)	5e-02	

(*continued*)

Table 1 (*continued*)

ORF	Strand[1]	Position		Length		Intergenic distance (bp)[2]	% A+T	Promoter[3]		Blast best match					Signature[4]
		Start	End	nt	aa			E/E+	L	Name	Species	BlastP score	Length (aa % identity)	E value	
OrV *Pst*I-D															
D1	+	57	344	288	95	n.d.	62.9								HEMOPEXIN; SP
D2	−	1491	382	1110	369	37	57.4			patatin-like phospholipase	*Bacillus cereus*	70.1	63/237 (26)	3e-10	Patatin; SP
D3	+	1622	2965	1344	447	130	56.0	−105 (−74) −94 (−59)							SP; TM
D4	+	3144	3341	198	65	178	75.7	−46	−63						SP; TM
D5	−	4644	3325	1320	439	−17	57.0	−85 (−51)	−47						
D6	+	5298	5945	648	215	653	56.6	−5; −50	−32						
D7	+	5988	6224	237	78	42	58.3	−191 (−159)		late expression factor 5	Xestia c-nigrum granulovirus	33.9	14/39 (35)	2.0	Baculo_LEF-5; Zinc beta-ribbon; SP
D8	−	8501	6288	2214	737	63	52.8	−293 (−269)	−173	ribonucleotide reductase large subunit	Agrotis segetum granulovirus	309.0	182/537 (33)	5e-82	Ribonuc_red_lgC
D9	+	8910	9080	171	56	408	67.2								
D10	+	9216	9377	162	53	135	63.0	−103 (−70) −7							
D11	+	9796	10632	837	278	418	55.3	−97 (−56)							ESTERASE; SP; TM
D12	+	10837	12408	1572	523	204	56.7	−103 (−72) −15	−200						
D13	+	12461	13432	972	323	52	56.7	−13 (+24)	−133						Glutamine-rich region; SP
D14	−	14035	13445	591	196	12	61.7		−107						
D15	+	13997	14152	156	51	−39	66.0	−298							
D16	−	15349	14144	1206	401	−9	58.3		−107 −155	late expression factor 4 (HzV-1_Orf98)	Heliothis zea virus 1	66.6	90/414 (21)	3e-09	LEF-4
D17	−	15666	15325	342	113	−25	55.3	−60 (−32) −121 (−108)		HzV-1_Orf124	Heliothis zea virus 1	45.4	30/113 (26)	1e-03	SP; TM
D18	−	16036	15638	399	132	−29	63.9		−178						
D19	+	16010	16579	570	189	−27	54.6	−67; −197							
D20	+	16664	16945	282	93	84	60.3								SP

[1] '+' Indicates clockwise direction of coding; '−' indicates counterclockwise direction of coding.

[2] '−' Indicates overlap between adjacent ORFs; *n.d.* not determined.

[3] *E* Early promoter (TATAWAW or TATAWTW, W = A or T); *E+* early promoter (E with CAKT cap, K = G or T), numbers indicate the locations of motifs in each promoter region, −xxx represents E, −xxx (−xxx) represents E+; *L* late promoter, numbers indicate the locations of motifs in each promoter region.

[4] *SP* Signal peptides; *TM* transmembrane domain.

istic sequence motifs using the InterProScan program (http://www.ebi.ac.uk/InterProScan/) and the Motif Scan program (http://myhits.isb-sib.ch/cgi-bin/motif_scan). To find possible transmembrane domains (TM) and N-terminal signal peptides (SP) in the putative proteins, the TMHMM server 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/) and SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/), respectively, were also used. Repeated and palindromic sequences were identified using the REPuter program (http://bibiserv.techfak.uni-bielefeld.de/reputer/submission.html), Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.submit.options.html), and GeneQuest program (Lasergene 5.0).

Phylogenetic analysis

Neighbour-joining (NJ), minimum evolution (ME) and maximum parsimony (MP) phylogenetic analyses were performed using MEGA 3.1 [27].

Results and discussion

Nucleotide sequence analysis

*Pst*I fragments C (*Pst*I-C) and D (*Pst*I-D) of OrV DNA were sequenced and were 19,805 and 17,146 bp in size, respectively. This is in good agreement with the originally predicted sizes of 20.1 and 17.7 kb, respectively, as determined in agarose gels [7]. Compared to the whole genome (about 130 kb in length), about 30% was successfully sequenced. So far, these are the largest genome sequence fragments known for OrV. The sequence of *Pst*I-D overlaps with a 4-kb fragment that was previously sequenced to provide the basis for a PCR detection method (AF126716) [41].

The AT contents of the *Pst*I-C and -D are 58.6 and 58.0%, respectively. Previously, the AT content of the OrV genome was estimated to be 57% [35], which is in close agreement with our sequence data. The *Pst*I-C and -D sequences were oriented clockwise in the circular OrV genome by comparing the positions of internal restriction endonuclease sites in the sequenced fragments with the restriction map of Crawford et al. [7] (Fig. 1). Given the fact that these are discontinuous sequence fragments, the starts of the *Pst*I-C and -D were arbitrarily chosen to be base #1, respectively, to simplify the reading, and the successive nucleotides were numbered in clockwise direction.

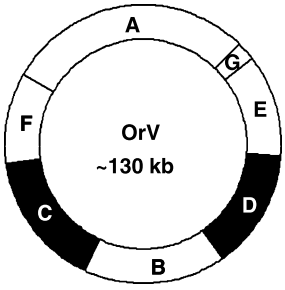

Fig. 1. Physical map (*Pst*I cleavage) of Oryctes rhinoceros virus genome DNA (redrawn according to Crawford et al.). Shadowed *Pst*I-C and -D fragments were sequenced in this study and oriented clockwise in the map

A total of 40 methionine-initiated ORFs with 50 or more amino acids and minimal overlap with adjacent ORFs were detected in the two fragments. Only 15 ORFs showed significant identity (21–51%) to hypothetical or known proteins from other dsDNA viruses or cellular organisms (Table 1, Fig. 2). The orientation of the ORFs was not equally distributed; 24 ORFs (60%) were oriented clockwise and 16 ORFs (40%) counterclockwise (Fig. 2). In total, 6 ORFs overlapped neighbouring ORFs, and the maximal intergenic overlap was found between ORFs D14 and D15, with 39 shared nucleotides. The AT content of the predicted ORFs ranged from 52.8 to 75.7%, with an average of 59.2% (Table 1).

Gene content

Of the 15 ORFs with significant similarities (21–51%) to hypothetical or known proteins from other dsDNA viruses or cellular organisms, ten showed significant similarities to ORFs of HzV-1. These ORFs include 3 homologues (C3, C20, and D17) of so far unique HzV-1 ORFs with unknown functions as well as 5 homologues of baculovirus core genes, *lef-4* (D16), *lef-5* (D7), *pif-2* (C2), *dnapol* (C17), and *ac81* (C14) (Table 2). DNA polymerase B is essential for viral DNA replication and is commonly found in all nucleo-cytoplasmic large dsDNA viruses infecting eukaryotic organisms. In contrast, *lef-4*, *lef-5*, *pif-2*, and *ac81* homologues have been only identified in HzV-1 and in all baculovirus genomes sequenced so far [23]. *Lef-4* encodes the mRNA capping subunit of the DNA-dependent

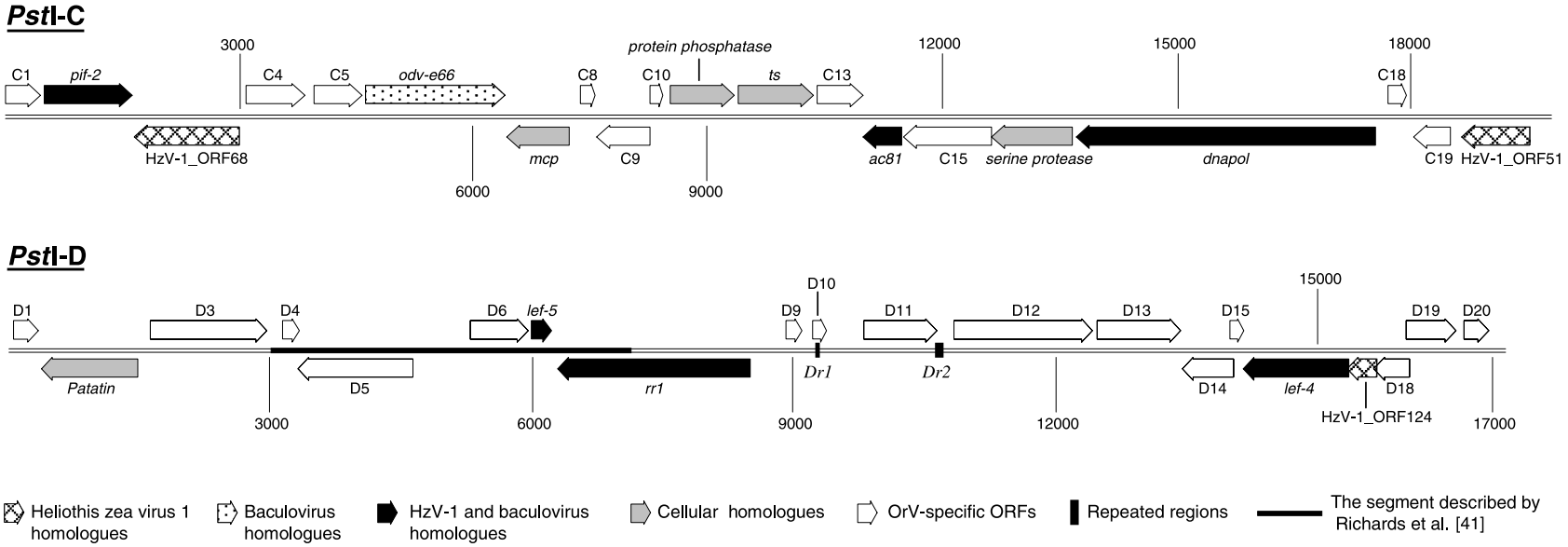

Fig. 2. Linear map of the predicted 40 ORFs for *Pst*I-C and -D fragments of Oryctes rhinoceros virus (OrV). Arrows indicate ORFs and the direction of transcription. A *ts* homologue is also present in HzV-1

Table 2. Baculovirus conserved core gene homologues present in Oryctes rhinoceros virus (OrV) and Heliothis zea virus 1 (HzV-1)

Function	Name[1]	ORF[2]		
		AcMNPV	HzV-1[3]	OrV
Transcription	*p47*	40	–	?
	lef-8	50	90	?
	lef-9	62	75	?
	vlf-1	77	121	?
	lef-4	**90**	**98**	**D16**
	lef-5	**99**	**101**	**D7**
Replication	*lef-2*	6	–	?
	lef-1	14	–	?
	dnapol	**65**	**131**	**C17**
	helicase	95	104	?
Structural proteins	***pif-2***	**22**	**123**	**C2**
	vp1054	54	–	?
	gp41	80	–	?
	vp91/p95	83	46	?
	vp39	89	–	?
	p6.9	100	–	?
	pif-3	115	88	?
	pif-1	119	55	?
	p74	138	11	?
	odv-e27	144	–	?
	odv-e56	148	76	?
Auxiliary function	*alk-exo*	133	–	?
Unknown function		68	–	?
		81	**33**	**C14**
	p33	92	–	?
	19kda	96	103	?
	38K	98	10	?
		109	–	?
	p49	142	–	?
Number of core genes		29	16	5

[1] The five baculovirus core gene homologues (*lef-4*, *lef-5*, *dnapol*, *pif-2*, and *ac81*) present in both OrV and HzV-1 are marked in bold.
[2] '?': Unknown; '+': present; '−': absent.
[3] Six underlined HzV-1 ORFs homologous to baculovirus core genes are detected in this study.

RNA polymerase, which is essential for transcription of viral late and very late genes [24]. *Lef-5* appears to be involved in transcription as an initiation factor and stimulates the transcription of both early and late viral genes [13]. *Pif-2* is located in occlusion-derived virus (ODV) and is essential for the peroral infectivity of baculoviruses [39]. The function of *ac81* is not yet known. Two other ORFs are homologous to non-conserved baculovirus genes. They encode a putative large subunit of ribonucleotide reductase (D8), which is also present in HzV-1, and a protein (C6) with similarity to occlusion-derived virus envelope protein 66 (ODV-E66) [16], which is absent in HzV-1. The other five ORFs showed similarities to cellular proteins (mitochondrial carrier protein, C7; Ser/Thr protein phosphatase family protein, C11; thymidylate synthase, C12; serine protease, C16; and patatin-like phospholipase, D2) (Table 1). A thymidylate synthase (TS) homologue was also found in HzV-1 and WSSV. Moreover, ten of these 15 ORFs contain functionally conserved domain and/or motif signatures, which further corroborate the putative biological functions of these predicted proteins (Table 1). However, the remaining 25 ORFs revealed no homology to any known gene. The three predicted short ORFs (<151 aa), D1, D20, and C1, might encode parts of larger genes extending beyond the cloned restriction fragments. They need to be re-evaluated when the complete virus genome is obtained.

The putative protein encoded by ORF D7 showed low global BlastP similarities to LEF-5 of baculoviruses and HzV-1 (Table 1), which results from its small size of only 78 amino acids (aa) in comparison with 230–315 aa for baculovirus LEF-5. PSI- and PHI-Blast as well as BlastP against a local database containing all of the currently sequenced baculovirus genomes and HzV-1 revealed that ORF D7 is significantly similar to the C-terminal end of baculovirus LEF-5. Furthermore, a C-terminal zinc ribbon domain required for the maximal late transcription activity [14] and characteristic for LEF-5 was also detected for the ORF D7 (Table 1).

ORF D6 gave a highly significant Blast hit to ODV-E66 of Cryptophlebia leucotreta granulovirus (Table 1). ODV-E66 is associated with the envelope of occluded virions of lepidopteran-specific nucleopolyhedroviruses and granuloviruses, where it is highly conserved. It is transported to and localized in the intranuclear microvesicles, which is a prerequisite for the accurate morphogenesis of preoccluded virions [16]. The presence of *odv-e66* and *pif-2* in OrV suggests that the envelope of OrV may

have related structural features to that of HzV-1 and the ODVs of baculoviruses.

Many studies have stated that there were no occlusion bodies in OrV-infected cells or tissues [8, 17, 18, 33, 35, 47]. Under certain circumstances, however, polyhedral virus occlusion bodies (OBs) with a distinct paracrystalline substructure comparable to that of OBs of nucleopolyhedroviruses (NPVs) and granuloviruses (GVs) have been observed in the nuclei of OrV-infected midgut epithelium cells of *O. rhinoceros* larvae [18]. However, a baculovirus-like OB protein gene could not be identified by DNA–DNA hybridization in a previous study [7]. The complete genome sequence and further studies of the protein composition as well as the pathogenesis of OrV will give insight

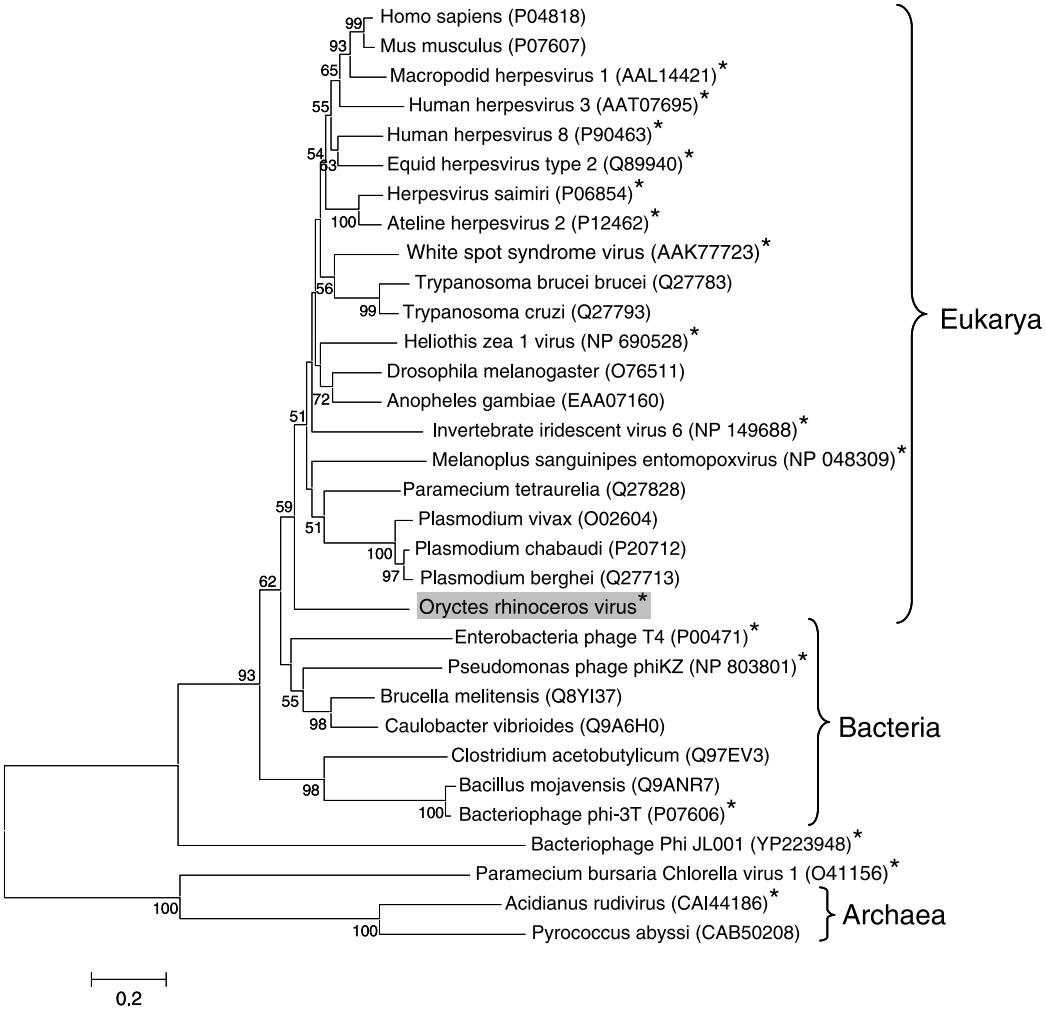

Fig. 3. Mid-point rooted Neighbour-joining phylogenetic tree based on members of 32 species and 184 sites showing the phylogenetic affiliation of Oryctes rhinoceros virus (OrV) thymidylate synthase (TS). The complete amino acid sequences of TS from dsDNA viruses are marked by asterisks, prokaryotic and eukaryotic organisms were retrieved from NCBI database (http://www.ncbi.nlm.nih.gov/). Distances were calculated using Poisson correction. Homogeneous substitution pattern among lineages and uniform rates among sites were employed for reconstruction of the tree. Gaps and missing data were excluded from the analyses. The robustness of the tree was tested using bootstrap analyses (1000 replicates), and the values above 50% are given next to the nodes. The OrV TS is indicated on a grey background. Names of species are followed by their GenBank accession numbers. The scale bar represents a distance of 20%

into whether there are any homologous or analogous structures of the OrV virion to capsid, envelope, and possibly OBs of baculoviruses.

A further striking feature of the partial OrV genome sequence is the presence of a putative thymidylate synthase gene (*ts*). TS plays an important role in the DNA replication of a cell or a DNA virus [38]. Although its distribution is ubiquitous in nature, a dsDNA viral TS homologue is found only in bacteriophages, chlorovirus, entomopoxvirus, herpesvirus, iridovirus, rudivirus, WSSV and HzV-1, but not in baculoviruses. Phylogenetic analyses of various TS genes originating from various viruses and cellular organisms clustered the OrV TS together with the TSs of eukaryotes and large eukaryotic dsDNA viruses (Fig. 3). The presence of putative genes involved in nucleotide metabolism, such as *ts* and *rr1* (ribonucleotide reductase large subunit), suggests that the OrV possibly contributes to the synthesis of the required nucleotides, independent of the host cell metabolism.

Notably, an unusual feature of HzV-1 is the presence of a histone fold protein [12]. This protein is speculated to be involved in mediating protein-DNA interactions, and consequently to serve as a DNA-binding domain [12]. However, its homologue was not detected in the partial genomic sequence of OrV.

Taken together, the observation that 10 OrV ORFs have homologues in HzV-1 indicates a close evolutionary relatedness of OrV to HzV-1. This hypothesis is further supported by the phylogenetic analyses of the DNA polymerase from dsDNA viruses including OrV ORF C17 (see below).

A hemopexin domain was detected in OrV ORF D1. Hemopexin-like domains have been found in vitronectin [46], a cell adhesion and spreading factor in plasma and tissues, and in matrixins, members of the matrix metalloproteinase (MMP) family that cleave extracellular matrix constituents [9]. A baculovirus MMP was identified in Xestia c-nigrum granulovirus (XecnGV) and other granuloviruses, which may degrade the basement membranes that help to hold tissues together [25]. It has been suggested that this domain facilitates binding to a variety of molecules and proteins. Accordingly, the putatively expressed D1 protein may potentially help OrV to interact with host cell membranes in order to trigger the virus infection or to assist in dissemination of the viruses. ORF D11 contains an esterase catalytic domain, suggesting hydrolytic activity of the putative protein. A PIN domain (PilT N terminus) was identified in ORF C3. While its function remains unknown, a role in signalling appears to be possible [6]. Since PIN domains are often present in nucleic-acid-binding proteins, ORF C3 may be involved in viral DNA replication or transcription. ORF C5 contains a tubulin chaperone cofactor A signature. Cofactor A is a chaperone involved in tubulin folding [31] and essential for cell viability [34], which hints that C5 may be involved in controlling cellular processes associated with the cytoskeleton during virus infection.

Promoter motifs

Baculovirus promoter motifs involved in early and late transcriptional regulation [11, 28] were identified in HzV-1 [5]. Since the gene content analysis of OrV indicated a close evolutionary relationship with HzV-1, it was important to know whether these promoter motifs are present in the OrV genome.

Twenty-eight ORFs, accounting for 70% of the 40 predicted ORFs, have an early promoter motif (Table 1). Of these, fifteen ORFs also possessed an initiator (INR) motif. In ORFs C11, D8, and D17, the INR motif was found at 19, 8, and 7 nt downstream of the TATA element, respectively. In ORFs C12 and D13, the INR motif was found downstream of the start codon. A baculovirus late gene promoter motif was detected in 19 ORFs (Table 1). Both early and late promoter motifs were observed in 14 ORFs, which may allow transcription of these genes during both early and late stages of infection. Of the OrV ORFs, six possessed neither early nor late promoter motifs. It is possible that they may be transcribed from hitherto unknown promoter motifs. The presence of baculovirus promoter motifs in OrV, as is the case for HzV-1, supports the view that OrV is evolutionarily related to baculoviruses.

Repeated regions of OrV sequence

Homologous region (*hr*) sequences, which are characterized by direct repeats containing an imperfect

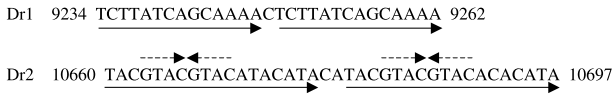

Fig. 4. Repeated regions of the *Pst*I-D fragment of Oryctes rhinoceros virus genome DNA. The numbers are their positons in the fragment. The straight-line arrows indicate direct repeat regions. The dashed-line arrows indicate palindromes

palindromic core and located at different positions in the genome, are a common feature of large, circular, invertebrate dsDNA viruses. They have been shown to function as both origins of DNA replication and transcriptional enhancers [26, 37]. *Hrs* have been found in most baculovirus genomes, and are also present in WSSV with no similarity to that of baculovirus [44]. Instead of *hrs*, abundant tandem repeated sequences were identified in HzV-1 [5]. Evidence for the presence of repeated regions within the genome of OrV was provided by Southern hybridization [7]. Within the *Pst*I-D fragment of OrV DNA, a *hr*-like double repeat of 18 bp with a palindromic core was located at nt position 10660–10697 (Fig. 4). In addition, a short direct repeat sequence (14 bp) was also detected at nt position of 9234–9262 (Fig. 4). No *hrs* and/or tandem repeat regions were found in fragment *Pst*I-C. No sequence homology between these OrV repeats and those of baculoviruses and other dsDNA viruses was observed.

Phylogeny

The DNA polymerase (type B family) was used to reconstruct the phylogeny of OrV. The DNA polymerase sequence is commonly employed to infer the evolutionary relationships of dsDNA viruses [3, 4, 19, 20, 22, 43]. It is also one of the 29 conserved core genes found in all baculovirus genomes sequenced so far. In total, 23 DNA polymerase sequences from different families of eukaryotic dsDNA viruses including nuclear and cytoplasmic large dsDNA viruses and one from an unclassified dsDNA virus group (HzV-1) were retrieved from GenBank, aligned and subjected to phylogenetic analysis. As shown in Fig. 5, the DNA polymerase tree clustered OrV together with HzV-1 with strong bootstrap support (71%), suggesting that both viruses shared a common ancestor. They formed a monophyletic group and represented a sister group to baculoviruses (Fig. 5). Coincidentally, these viruses were previously considered to be related based on their rod-shaped virion/nucleocapsid morphology, replication in the nucleus of infected cells and occurrence in arthropods. The phylogenetic relatedness between OrV and HzV-1 as revealed by the DNA polymerase tree is strongly supported by the observed gene content in these taxa. Of the 40 ORFs identified for OrV, ten ORFs have homologues in HzV-1, and seven homologues (*lef-4*, *lef-5*, *pif-2*, *dnapol*, *ac81*, *rr1*, *odv-e66)* are present in baculoviruses. Since only 30% of the OrV genome has been sequenced, the number of common genes shared with HzV-1 and baculoviruses is likely to increase. So far, 29 core baculovirus genes are present in all sequenced baculoviruses from Lepidoptera, Hymenoptera and Diptera [15, 23]. PSI-Blast analyses showed that HzV-1 shares 16 ORFs with baculovirus core genes (Table 2). Due to the low number of available baculovirus genome sequences at the time, six of these homologues were not identified when the genome of HzV-1 was sequenced [5] (Table 2). Baculovirus core genes confer basic biological functions in the infection process *in vivo*, in RNA synthesis and in DNA replication. Given the crucial functions of their encoded proteins during the viral infection and replication cycles and given the fact that many of these proteins are a part of multi-factorial and multi-functional protein complexes, their horizontal transferability seems to be less likely than the transfer of genes encoding proteins of less central importance. Hence, most, if not all, of these genes can be considered as phylogenetically ancestral and may have evolved under stringent driving forces. It is thus very likely that the homologues found in HzV-1 and OrV genomes are the footprints of a common ancestor of baculoviruses and HzV-1/OrV rather than the result of horizontal gene transfer. These findings indicate that both OrV and HzV-1 are distantly related to baculoviruses and share a common ancestor with them.

In conclusion, based on sequence and phylogenetic analyses, OrV can not be readily assigned to any currently known dsDNA virus family. Given the fact that OrV has high genetic similarity to

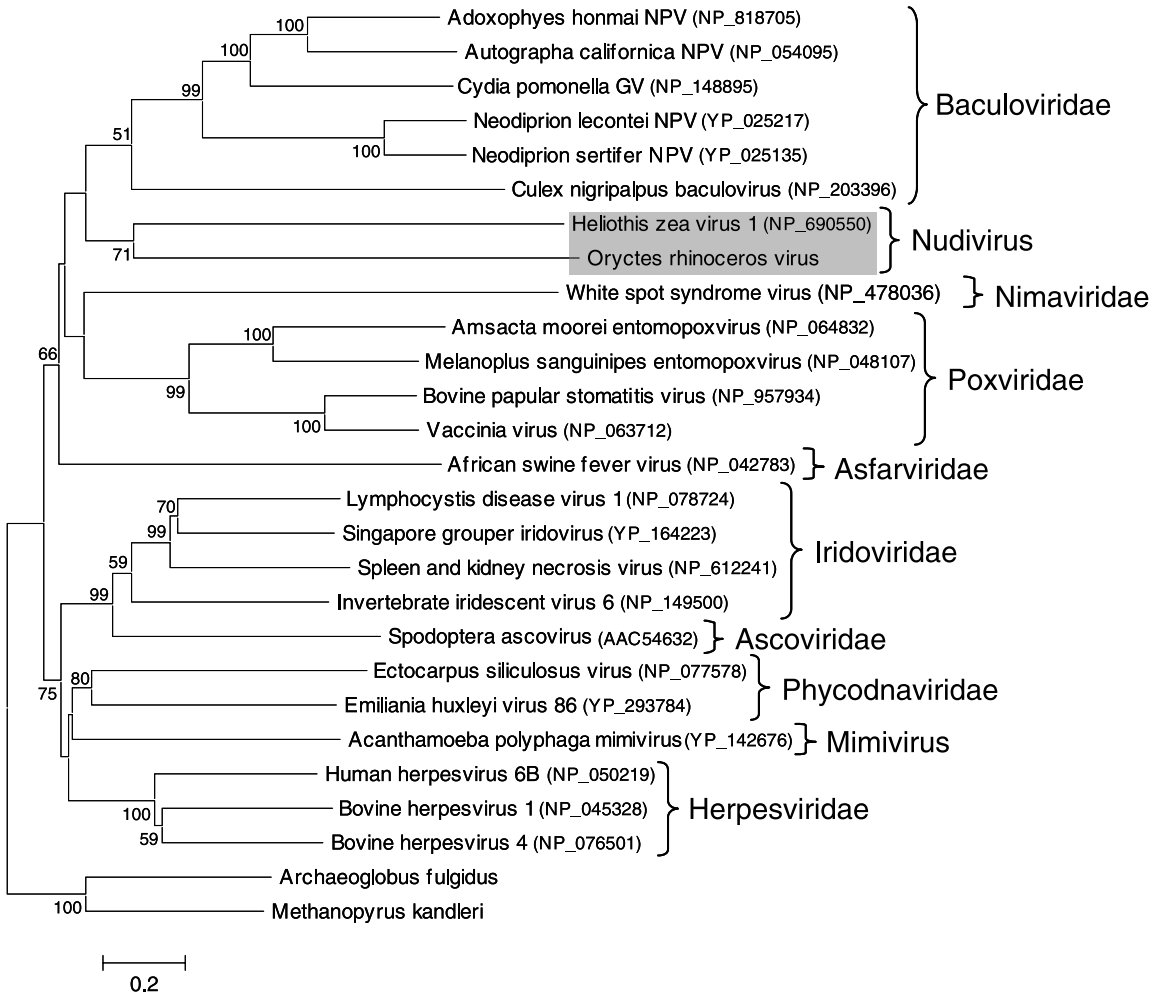

Fig. 5. Minimal evolution (ME) phylogenetic tree, close-neighbour-interchange (CNI) level 1, constructed using DNA polymerase (583 sites) shows the phylogenetic affiliations of Oryctes rhinoceros virus (OrV) to other dsDNA viruses. Distances were calculated using Poisson correction. Homogeneous substitution patterns among lineages with uniform rates among sites are employed for reconstruction of the tree. Gaps and missing data were excluded from the analyses. The robustness of the tree was tested using bootstrap analyses (1000 replicates), and the values above 50% are given next to the nodes. Neighbour-joining (NJ) and maximum parsimony (MP) analyses revealed similar tree topology. The names of the selected virus families are marked on the tree. Both OrV and Heliothis zea virus 1 (HzV-1) are indicated on a grey background. Names of species are followed by their GenBank accession numbers. The tree is rooted using two archael species: *Archaeoglobus fulgidus* (AE001070) and *Methanopyrus kandleri* (AE010392). The scale bar represents a distance of 20%

HzV-1, we propose to include HzV-1 and OrV as members of two different species in a new virus genus. We further propose the name Nudivirus for this genus. This name refers to the vernacular name and has been used for these viruses for a long time [2]. According to the effective rules of ICTV, a virus may not be named only according to its host, with the suffix virus [10]. OrV was discovered and named more than 40 years ago, at a time when ICTV was not yet even established. Huger proposed the name Rhabdionvirus in order to indicate the extremely destructive effect of the virus to *O. rhinoceros* populations [17]. However, Rhabdionvirus sounds very similar to the acknowledged genus *Rhabdovirus* (family *Rhabdoviridae*), and we suggest not using this historic name for OrV any more. In order to give OrV a name that is in agreement to the ICTV rules, we propose to rename OrV

to Oryctes rhinoceros nudivirus (OrNV). Accordingly, HzV-1 should be renamed to Heliothis zea nudivirus 1 (HzNV-1). In light of the phylogenetic analysis of the DNA polymerase and the observed gene content of OrV and HzV-1, we conclude that the proposed nudiviruses are much more closely related to members of the family *Baculoviridae* than to *Nimaviridae*. The complete genome sequences of OrV and of other non-occluded baculovirus-like viruses should provide indications whether these viruses should be considered as belonging to a separate virus family or to a subfamily within the family *Baculoviridae*.

Acknowledgements

This study was funded by the Deutsche Forschungsgemeinschaft to J.A.J (Je245-7). We thank Alexander Rosisko for technical assistance.

References

1. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402
2. Burand JP (1998) Nudiviruses. In: Miller LK, Ball LA (eds) The insect viruses. Plenum Press, New York, pp 69–90
3. Chen F, Suttle CA (1996) Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. Virology 219: 170–178
4. Chen LL, Wang HC, Huang CJ, Peng SE, Chen YG, Lin SJ, Chen WY, Dai CF, Yu HT, Wang CH, Lo CF, Kou GH (2002) Transcriptional analysis of the DNA polymerase gene of shrimp white spot syndrome virus. Virology 301: 136–147
5. Cheng CH, Liu SM, Chow TY, Hsiao YY, Wang DP, Huang JJ, Chen HH (2002) Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the *Baculoviridae*. J Virol 76: 9024–9034
6. Clissold PM, Ponting CP (2000) PIN domains in nonsense-mediated mRNA decay and RNAi. Curr Biol 10: R888–R890
7. Crawford AM, Ashbridge K, Sheehan C, Faulkner P (1985) A physical map of the *Oryctes* baculovirus genome. J Gen Virol 66: 2649–2658
8. Crawford AM, Sheehan C (1985) Replication of *Oryctes* baculovirus in cell culture: viral morphogenesis, infectivity and protein synthesis. J Gen Virol 66: 529–539
9. Das S, Mandal M, Chakraborti T, Mandal A, Chakraborti S (2003) Structure and evolutionary aspects of matrix metalloproteinases: a brief overview. Mol Cell Biochem 253: 31–40
10. Fauquet CM, Fargette D (2005) International Committee on Taxonomy of Viruses and the 3142 unassigned species. Virol J 2: 64
11. Friesen PD (1997) Regulation of baculovirus early gene expression. In: Miller LK (ed) The baculoviruses. Plenum Press, New York, pp 141–170
12. Greco C, Fantucci P, De Gioia L (2005) *In silico* functional characterization of a double histone fold domain from the *Heliothis zea* virus 1. BMC Bioinformatics 6: S15
13. Guarino LA, Dong W, Jin J (2002) In vitro activity of the baculovirus late expression factor LEF-5. J Virol 76: 12663–12675
14. Harwood SH, Li L, Ho PS, Preston ÅK, Rohrmann GF (1998) AcMNPV late expression factor-5 interacts with itself and contains a zinc ribbon domain that is required for maximal late transcription activity and is homologous to elongation factor TFIIS. Virology 250: 118–134
15. Herniou EA, Olszewski JA, Cory JS, O'Reilly DR (2003) The genome sequence and evolution of baculoviruses. Annu Rev Entomol 48: 211–234
16. Hong T, Summers MD, Braunagel SC (1997) N-terminal sequences from *Autographa californica* nuclear polyhedrosis virus envelope proteins ODV-E66 and ODV-E25 are sufficient to direct reporter proteins to the nuclear envelope, intranuclear microvesicles and the envelope of occlusion derived virus. Proc Natl Acad Sci USA 94: 4050–4055
17. Huger AM (1966) A virus disease of the Indian rhinoceros beetle, *Oryctes rhinoceros* (Linnaeus), caused by a new type of insect virus, *Rhabdionvirus oryctes* gen. n., sp. n. J Invertebr Pathol 8: 38–51
18. Huger AM, Krieg A (1991) Baculoviridae. Nonoccluded baculoviruses. In: Adams JR, Bonami JR (eds) Atlas of invertebrate viruses. CRC Press, Boca Raton, pp 287–319
19. Iyer LM, Aravind L, Koonin EV (2001) Common origin of four diverse families of large eukaryotic DNA viruses. J Virol 75: 11720–11734
20. Iyer LM, Balaji S, Koonin EV, Aravind L (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res 117: 156–184
21. Jackson TA, Crawford AM, Glare TR (2005) *Oryctes* virus – time for a new look at a useful biocontrol agent. J Invertebr Pathol 89: 91–94
22. Jancovich JK, Mao J, Chinchar VG, Wyatt C, Case ST, Kumar S, Valente G, Subramanian S, Davidson EW, Collins JP, Jacobs BL (2003) Genomic sequence of a ranavirus (family *Iridoviridae*) associated with salamander mortalities in North America. Virology 316: 90–103

23. Jehle JA, Blissard GW, Bonning BC, Cory JS, Herniou EA, Rohrmann GF, Theilmann DA, Thiem SM, Vlak JM (2006) On the classification and nomenclature of baculoviruses: a proposal for revision. Arch Virol 151: 1257–1266
24. Knebel-Morsdorf D, Quadt I, Li Y, Montier L, Guarino LA (2006) Expression of baculovirus late and very late genes depends on LEF-4, a component of the viral RNA polymerase whose guanyltransferase function is essential. J Virol 80: 4168–4173
25. Ko R, Okano K, Maeda S (2000) Structural and functional analysis of the *Xestia c-nigrum* granulovirus matrix metalloproteinase. J Virol 74: 11240–11246
26. Kool M, van den Berg PM, Tramper J, Goldbach RW, Vlak JM (1993) Location of two putative origins of DNA replication of *Autographa californica* nuclear polyhedrosis virus. Virology 192: 94–101
27. Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5: 150–163
28. Lu A, Miller LK (1997) Regulation of baculovirus late and very late gene expression. In: Miller LK (ed) The baculoviruses. Plenum Press, New York, pp 193–216
29. Mayo MA (1995) Unassigned viruses. In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) Virus taxonomy: The Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien New York, pp 504–507
30. Mayo MA, Christian PD, Hillman BI, Brunt AA, Desselberger V (2005) The unassinged viruses. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy: The Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, New York, pp 1129–1144
31. Melki R, Rommelaere H, Leguy R, Vandekerckhove J, Ampe C (1996) Cofactor A is a molecular chaperone required for beta-tubulin folding: functional and structural characterization. Biochemistry 35: 10422–10435
32. Mohan KS, Gopinathan KP (1991) Physical mapping of the genomic DNA of the *Oryctes rhinoceros* baculovirus, KI. Gene 107: 343–344
33. Monsarrat P, Duthoit JL, Vago C (1974) Demonstration of virions of the Baculo-virus type in the genital system of the coleopteran, *Oryctes rhinoceros* L. C R Acad Sci Hebd Seances Acad Sci D 278: 3259–3261
34. Nolasco S, Bellido J, Gonçalves J, Zabala JC, Soares H (2005) Tubulin cofactor A gene silencing in mammalian cells induces changes in microtubule cytoskeleton, cell cycle arrest and cell death. FEBS Lett 579: 3515–3524
35. Payne CC (1974) The isolation and characterization of a virus from *Oryctes rhinoceros*. J Gen Virol 25: 105–116
36. Payne CC, Compson D, de Looze SM (1977) Properties of the nucleocapsids of a virus isolated from *Oryctes rhinoceros*. Virology 77: 269–280
37. Pearson M, Bjornson R, Pearson G, Rohrmann G (1992) The *Autographa californica* baculovirus genome: evidence for multiple replication origins. Science 257: 1382–1384
38. Perryman SM, Rossana C, Deng TL, Vanin EF, Johnson LF (1986) Sequence of a cDNA for mouse thymidylate synthase reveals striking similarity with the prokaryotic enzyme. Mol Biol Evol 3: 313–321
39. Pijlman GP, Pruijssers AJ, Vlak JM (2003) Identification of *pif-2*, a third conserved baculovirus gene required for *per os* infection of insects. J Gen Virol 84: 2041–2049
40. Ramle M, Wahid MB, Norman K, Glare TR, Jackson TA (2005) The incidence and use of *Oryctes* virus for control of rhinoceros beetle in oil palm plantations in Malaysia. J Invertebr Pathol 89: 85–90
41. Richards NK, Glare TR, Aloalii I, Jackson TA (1999) Primers for the detection of *Oryctes* virus from *Scarabaeidae* (Coleoptera). Mol Ecol 8: 1552–1553
42. Schäffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res 29: 2994–3005
43. Schroeder DC, Oke J, Malin G, Wilson WH (2002) Coccolithovirus (*Phycodnaviridae*): characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. Arch Virol 147: 1685–1698
44. van Hulten MC, Witteveldt J, Peters S, Kloosterboer N, Tarchini R, Fiers M, Sandbrink H, Lankhorst RK, Vlak JM (2001) The white spot syndrome virus DNA genome sequence. Virology 286: 7–22
45. Vlak JM, Bonami JR, Flegel TW, Kou GH, Lightner DV, Lo CF, Loh PC, Walker PW (2005) Nimaviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy: The Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, New York
46. Yoneda A, Ogawa H, Kojima K, Matsumoto I (1998) Characterization of the ligand binding activities of vitronectin: interaction of vitronectin with lipids and identification of the binding domains for various ligands using recombinant domains. Biochemistry 37: 6351–6360
47. Zelazny B (1972) Studies on *Rhabdionvirus oryctes* I. Effect on Larvae of *Oryctes rhinoceros* and inactivation of the virus. J Invertebr Pathol 20: 235–241