

# Nudivirus Genomics and Phylogeny

Yongjie Wang<sup>1,\*</sup>, Olaf R.P. Bininda-Emonds<sup>2</sup>, and Johannes A. Jehle<sup>3</sup>

<sup>1</sup>Laboratory of Marine and Food Microbiology,  
College of Food Science and Technology,  
Shanghai Ocean University, Shanghai,

<sup>2</sup>Institute for Biology and Environmental Sciences (IBU),  
Carl von Ossietzky University Oldenburg, Oldenburg,

<sup>3</sup>Institute for Biological Control,  
Federal Research Centre for Cultivated Plants,  
Julius Kühn-Institut, Darmstadt,

<sup>1</sup>China

<sup>2,3</sup>Germany

## 1. Introduction

The nudiviruses (NVs) are a diverse group of arthropod-specific large DNA viruses. They form rod-shaped, enveloped virions, and replicate in the nucleus of infected cells. Nudivirus genomes are covalently closed circles of double stranded DNA molecules. Some nudiviruses have been used as potential bio-control agent for management of economically important arthropod pests (Burand 1998, Huger 1966). A variety of non-occluded rod-shaped dsDNA viruses replicating in the host nucleus have been observed in various host species, belonging to Lepidoptera, Trichoptera, Diptera, Siphonaptera, Hymenoptera, Neuroptera, Coleoptera, Homoptera, Thysanura, Orthoptera, Acarina, Araneina, and Crustacea. They had been considered as “non-occluded baculoviruses” (Huger and Krieg 1991) or more recently as nudiviruses (Burand, 1998). Most of these viruses were identified solely based on morphological features and very limited biological data. Accordingly, it remains unclear whether they are evolutionarily monophyletic or polyphyletic lineages, and whether they are genetically related to each other, to the well-investigated baculoviruses, or to other large dsDNA viruses.

Thus far, only a few nudiviruses have somehow been studied in detail. The *Oryctes rhinoceros* nudivirus (OrNV), formerly known as the rhinoceros beetle virus or *Oryctes* baculovirus, was discovered in the 1960s and has been widely used to control rhinoceros beetle (*O. rhinoceros*) in coconut and oil palm in Southeast Asia and the Pacific (Huger 1966, Jackson et al. 2005). It has an enveloped rod-shaped virion and replicates in the nucleus of infected midgut and fat body cells (Huger 1966, Payne 1974, Payne et al. 1977). *Heliothis zea* nudivirus 1 (HzNV-1), formerly known as Hz-1 virus or the non-occluded baculovirus Hz-1, was originally described as a persistent viral infection in the IMC-Hz-1 cell line isolated

---

\*Corresponding Author

from the adult ovarian tissues of the corn earworm *Heliothis zea* (Granados et al. 1978). It can also persistently infect several other lepidopterous cell lines, e.g. IPLB-1075 (*H. zea*), IPLB-SF-21 (*Spodoptera frugiperda*), IPLB-65Z (*Lymantria dispar*) and TN-368 (*Trichoplusia ni*) (Granados et al. 1978, Kelly et al. 1981, Lu and Burand 2001). In contrast, clear infections have not been observed when the virus was inoculated into larvae of *H. zea*, *H. armigera*, *Estigmene acrea*, *S. frugiperda*, and *S. littoralis* (Granados et al. 1978, Kelly et al. 1981). The potential molecular mechanisms associated with this defective host infection of HzNV-1 need to be explored, which will shed light on the viral evolution. *Gryllus bimaculatus* nudivirus (GbNV) infects nymphs and adults of several field crickets *G. bimaculatus*, *G. campestris*, *Teleogryllus oceanicus* and *T. commodus*, and replicates in the nuclei of the infected fat body cells (Huger 1985). *Heliothis zea* nudivirus 2 (HzNV-2), previously known as gonad-specific virus, *H. zea* reproductive virus or Hz-2V, was first observed in the gonads of adult corn earworm *H. zea*. Its infection brings about deformities of the reproductive organs of insect hosts, which in turn lead to sterility in both female and male moths (Burand and Rallis 2004, Raina et al. 2000). HzNV-2 is also able to infect other Noctuid species and to replicate in two lepidopteran insect cell lines of TN-368 and Ld652Y, derived from ovarian tissues (Burand and Lu 1997, Lu and Burand 2001, Raina and Lupiani 2006).

### 1.1 Infection cycles and gene expression

Only very limited data on the infection cycle of nudiviruses are available. Their life cycle in either cell culture or natural hosts is still poorly understood. HzNV-1 has a bi-phasic infection process of latency and productivity in its life cycle. In the latent phase of infection, viruses either exist as episomes or insert their DNA into the host genome (Lin et al. 1999), and keep latency for many passages in the infected insect cells (Chao et al. 1992, Lin et al. 1999, Wood and Burand 1986); virus particles are undetectable in most of these latently infected cells. Sometimes virions are released from as few as 0.2% of latently infected cells, resulting in the presence of low viral titers (around  $10^3$  PFU/ml) in the culture medium (Chao et al. 1998, Lin et al. 1999). During the productive infection cycle, in contrast, high titers of virus progeny are produced, resulting in the death of most cells. Often, however, a small proportion of the cells, usually less than 5%, are latently infected, and viruses stay in these cells for a prolonged period of time (Chao et al. 1992, Wood and Burand 1986). Upon in vitro infection, OrNV appears to attach to and subsequently internalizes into cultured cells by pinocytosis (Crawford and Sheehan 1985), a mechanism involving the formation of invaginations by the cell membrane, which close and break off to generate virus-containing vacuoles in the cytoplasm. While it remains unknown how the viral DNA is released into the cytoplasm and eventually enters the nucleus. During the later stage of replication, along with the cytopathic changes to the nucleus, the virogenic stroma is developed, where the viral envelopes and nucleocapsid shells are produced and subsequently packaged with viral DNA. At last, the matured virions enter the cytoplasm followed by budding through the cell membrane (Crawford and Sheehan 1985).

In vitro sequential expression of viral genes encoding structural and intracellular proteins has been divided into early, intermediate and late stage in the replication cycle of OrNV (Crawford and Sheehan 1985). The temporal gene expression profiles of HzNV-1 during productive infection are divided into three stages: (i) the early stage, 0 to 2 h p.i.; (ii) the intermediate stage, 2 to 6 h p.i.; and (iii) the late stage, which includes all virus-specific

events appearing after 6 h p.i. (Chao et al. 1992). Persistency-associated transcript 1 (PAT1), expressed by persistency-associated gene 1 (*pag1*), is the only detectable transcript during latent infection of HzNV-1 (Chao et al. 1998).

## 1.2 Taxonomy and nomenclature

Given that they share similar structural and replication aspects with baculoviruses of insects, nudiviruses were previously classified as the so-called “non-occluded baculoviruses” (NOBs) (Huger and Krieg 1991). NOBs were later removed from the family *Baculoviridae* because no genetic data were available which would have supported their relationship (Mayo 1995). Nudiviruses have been also referred to as intranuclear bacilliform viruses (IBVs). Notably, unlike baculoviruses, nudiviruses generally lack occlusion bodies (OBs). The genus name *Nudivirus* has been proposed to accommodate this group. Based on the currently available morphological and molecular data, the following demarcation criteria were proposed for classification of a candidate virus into the genus *Nudivirus*: (i) Viral genome is consist of large circular dsDNA molecule; (ii) A set of conserved core genes are shared among members and viruses propagate in the nuclei of infected host cells; (iii) Morphology of virion is rod-shaped and enveloped; (iv) Viruses are transmitted per oral and/or per parenteral route, and infect larvae and/or adults with diverse tissue and cell tropisms (Wang et al. 2007a). Obviously, these demarcation criteria need to be complemented with more biological properties, such as virion properties, infection and replication strategies, as well as host range and virus ecology, becoming available. To name a nudivirus species, it was suggested to follow the nomenclature for other large eukaryotic dsDNA viruses, host name with the suffix name of nudivirus (Wang et al. 2007c).

Presently, nudiviruses comprise five tentative species, OrNV, GbNV, HzNV-1, HzNV-2, and *Penaeus monodon* nudivirus (PmNV) (Wang and Jehle 2009). Considering their similarities to baculoviruses and, on the other hand, taking their distinct biological, ecological features and virion properties into account, the establishment of an independent family “Nudiviridae” within a new order “Baculovirales” along with the *Baculoviridae* seems most appropriate. The establishment of an order “Baculovirales” would allow subsequent flexible integration of other “baculovirus-related” but highly diverged viruses, such as the proposed “Hytrosaviridae” (Abd-Alla et al. 2009) or the *Nimaviridae*, without taxonomic re-definition of the family *Baculoviridae*.

## 2. Genome structure

### 2.1 Genome size

HzNV-1 was the first completely sequenced nudivirus (Cheng et al. 2002). Its genome is 228,089 bp in size, has a G+C content of 42%, and encodes 154 ORFs (Table 1). HzNV-1 ORFs are randomly distributed on both DNA strands with 45% clockwise orientation and 55% counterclockwise orientation. HzNV-2, the close relative to HzNV-1, has a genome of 231,621 bp, only slightly longer than that of HzNV-1, with a G+C content of 42% identical to HzNV-1 (Wang et al. 2007a). Later on, the genome of OrNV, the first discovered nudivirus, was partially sequenced (Wang et al. 2007c). Recently, the complete genome of OrNV was successfully achieved using DNA generated with multiple displacement amplification (MDA) (Wang et al. 2008). The OrNV genome is 127,615 bp in size with a G+C content of

42% and contains 139 ORFs (Table 1, Fig. 1) (Wang et al. 2008, Wang et al. 2011). Thus far, the smallest nudivirus genome sequenced is GbNV, which is 96,944 bp in length with a G+C content of 28% and contains 98 ORFs. Among them, 58% are in clockwise distribution and 42% are in reverse direction (Table 1, Fig. 2) (Wang et al. 2007b). Genome sequencing of other nudiviruses such as the *Tipula oleracea* nudivirus (ToNV) (E. Herniou, personal communication) is ongoing. Partial nucleotide sequences of the shrimp PmNP genome are already accessible in GenBank.

Virus	Size in bp	GC content (mol %)	No. of ORFs	Clockwise orientation* (No. of ORF/ ORF%)	Gene density (kbp per ORF)	No. of Rsr
HzNV-1	228,089	41.8	154	69/45	1.47	3
GbNV	96,944	28.0	98	57/58	0.93	17
OrNV	127,615	42.0	139	64/46	0.82	20

Table 1. Characteristics of nudivirus genomes. \*Clockwise orientation means in the same orientation as the DNA polymerase B ORF. Rsr = Repetitive sequence regions.

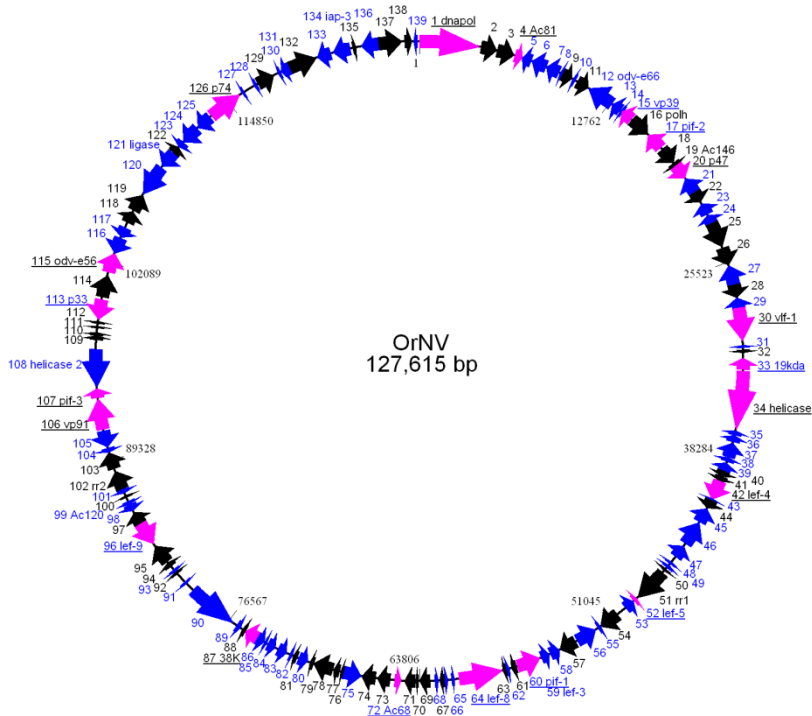


Fig. 1. The genome map of OrNV. ORFs and their transcriptional directions are indicated in arrows. Black color, clockwise coding; blue color, counterclockwise coding; pink color, the 20 baculovirus core gene homologues.

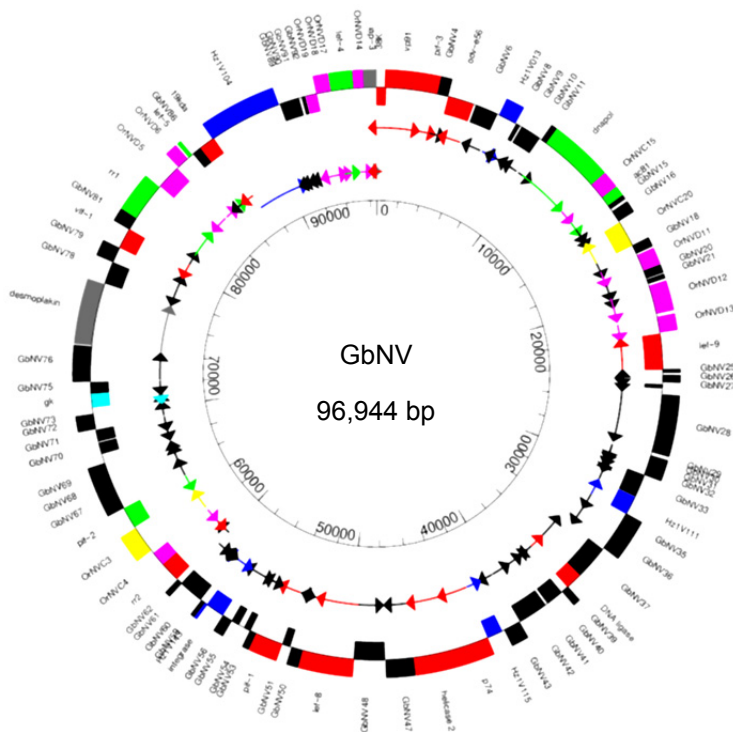


Fig. 2. Circular map of the GbNV genome. ORFs and their transcriptional directions are indicated in boxes and arrows, respectively. Red color, baculovirus and HzNV-1 homologues; black color, GbNV specific ORFs; blue color, HzNV-1 homologues; green color, baculovirus, HzNV-1 and OrNV homologues; pink color, OrNV homologues; yellow color, HzNV-1 and OrNV homologues; light blue, cellular homologues; grey color, baculovirus homologues. Taken from Wang et al. (2007b) with permission from the American Society for Microbiology.

## 2.2 Gene order

Similar to what is observed in other viral families (e.g., the *Baculoviridae*), gene order is poorly conserved in nudivirus genomes as well. OrNV and GbNV share a number of gene clusters, comprising 2–7 collinearly arranged genes, distributed throughout their genomes (Wang and Jehle 2009). In contrast, only two gene clusters were detected between OrNV and HzNV-1 (Wang and Jehle 2009, Wang et al. 2011). However, a gene cluster of *helicase*, *pif-4/19 kda*, and/or *lef-5* is present in all three nudivirus genomes (Fig. 3), which is similar to the conserved core gene cluster of four genes of *helicase*, *pif-4/19 kda*, *38K* and *lef-5* in all sequenced baculoviruses (Herniou et al. 2003, Jehle and Backhaus 1994). Hence, core gene clustering strongly supports the hypothesis of a common ancestor of nudiviruses and baculoviruses.

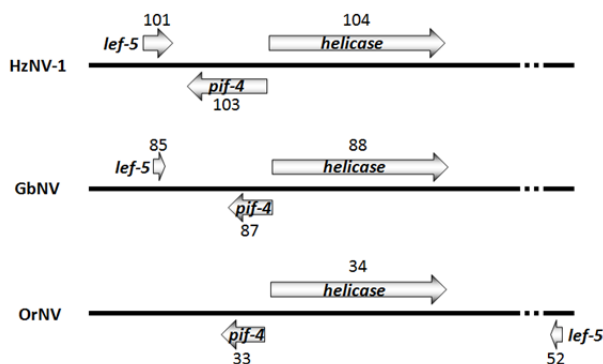


Fig. 3. Conserved gene cluster of *helicase*, *pif-4*, and/or *lef-5* on the genomes of HzNV-1, GbNV and OrNV. ORF is indicated by boxed arrows; number above and below the boxed arrows represents the sequential ordering of ORFs on the viral genomes; line in bold represents the viral genomes; dashed bold line indicates the omitted genomic ranges.

### 2.3 Repetitive regions

Repetitive sequence regions (Rsr) were detected in all three sequenced nudivirus genomes. They are variable in length and numbers and are distributed throughout the genome. They are homologous neither to each other within and between genomes, nor to those of other large dsDNA viruses, such as baculoviruses, hytrosaviruses and white spot syndrome virus (WSSV). Rsr appear to be a universal feature of all large dsDNA viruses.

## 3. Nudivirus gene structure

### 3.1 Promoter motifs

A promoter motif of TTATAGTAT was identified at the upstream regulatory regions of HzNV-1 late gene *p34* (ORF79) and *p51* (ORF64) (Guttieri and Burand 1996, Guttieri and Burand 2001). It was also found within 200 bp of the initiation codon of HzNV-1 ORF81 based on in silico sequence analysis (Cheng et al. 2002). Although consensus early and late promoter motif sequences similar to those of baculoviruses were predicted in nudivirus ORFs, convincing experimental data remain unavailable (Cheng et al. 2002, Wang et al. 2007c).

### 3.2 Untranslated regions

In the HzNV-1 transcripts, the early gene *hhi1* (HzNV-1 HindIII fragment 1 gene) contains 270 nucleotides (nts) of 5' untranslated region (UTR) which, together with its upstream 62 bps, compose *hhi1* early promoter (Wu et al. 2008, Wu et al. 2010); the HzNV-1 late gene *p34* (ORF79) possesses 16 and 17 nts of 5' UTR, respectively, differing by 1 nt, and both 5' UTRs overlap with the identified 9 bp late promoter motif of the *p34* (Guttieri and Burand 1996); as for the HzNV-1 late gene *p51* (ORF64), the major late transcriptional initiation site is at -205 bp relative to the translational start codon and seven minor late start sites locate at various positions upstream of this primary site (Guttieri and Burand 2001). The putative

polyadenylation signals (AATAAA) downstream of the stop codon of the *p34* and *p51* were found (Guttieri and Burand 1996, Guttieri and Burand 2001). Thus far, nothing is known on how UTR mediate the translational efficiency of nudivirus genes.

### 3.3 Open reading frames (ORFs)

Computer-assisted ORF prediction included all sequences starting with ATG followed by 50 or more amino acid (aa) codons and minimum overlap with other ORFs. ORFs with less than 50 aa are only considered as putative genes in cases of clear homology to ORFs in other dsDNA viruses.

## 4. Gene content and conserved gene functions

There are 66, 34, and 33 homologous genes shared by OrNV and GbNV, OrNV and HzNV-1, and GbNV and HzNV-1, respectively (Table 2), suggesting that OrNV and GbNV are more closely related to each other than to HzNV-1. OrNV, GbNV and HzNV-1 have 33 genes in common (Table 2). Strikingly, 20 out of them are homologues of baculovirus core genes, which are present in all 54 baculovirus genomes that have been deposited in GenBank as of July 2011. Baculovirus 31 core genes play crucial role in virus replication cycle and are the evolutionarily conserved marker genes in identification, classification and phylogeny of baculoviruses (Herniou et al. 2003, Herniou and Jehle 2007, Jehle et al. 2006a, Jehle et al. 2006b, van Oers and Vlaskovska 2007). Nine other ORFs are likely involved in DNA replication, repair and recombination, and nucleotide metabolism; one is homologous to baculovirus *iap-3* gene; two others are nudivirus-specific ORFs of unknown function (Table 2). The presence of 20 baculovirus core genes in nudiviruses strongly indicates that nudiviruses and baculoviruses are the closest lineages among the viruses known so far.

Besides in nudiviruses, homologues to baculovirus core genes were also detected in two salivary gland hypertrophy viruses (SGHVs) MdSGHV infecting the house fly *Musca domestica* and GpSGHV infecting the tsetse fly *Glossina pallidipes* (Abd-Alla et al. 2008, Garcia-Maruniak et al. 2008). GpSGHV and MdSGHV share 37 homologous ORFs and are phylogenetically closely related (Garcia-Maruniak et al. 2009). Surprisingly, several core gene homologues of baculoviruses was identified in the marine WSSV as well (Wang et al. 2011), suggesting that WSSV, as suspected since it was observed, is evolutionarily related, albeit distantly, to baculoviruses. Most strikingly, nudiviruses, SGHVs and WSSV have the homologues to the genes encoding peroral infectivity factors (*p74*, *pif-1*, *pif-2* and *pif-3*) (Wang et al. 2011). These four *pif* genes are conserved among all sequenced baculoviruses and are absolutely crucial for successful peroral infection of insect hosts. As midgut infection is the essential first step in the invasion of baculoviruses, PIFs may be the key determinants of host range and virulence. Accordingly, it seems to be reasonable to hypothesize that a highly conserved interaction mode of viruses and hosts upon primary infection is present in nudiviruses, baculoviruses, SGHVs and WSSV. However, only limited data on the function of the PIF proteins has been delineated in baculoviruses (Slack and Arif 2007), let alone in nudiviruses, SGHVs and WSSV. Obviously, deeply exploring of the molecular mechanisms of the PIF proteins as well as their homologues is crucial for better understanding of host range, zoonotic behaviour, and epizootic or enzootic disease of these viruses. In addition, nudiviruses appear to share homologues of the transcription apparatus of baculoviruses, suggesting that a similar mode of late gene transcription is used in

nudiviruses as well (Wang et al. 2011). Taken together, this finding provides crucial clues to the origin and evolution of arthropod specific large dsDNA viruses. The biochemical and biological function of the genes predicted in nudiviruses remains unknown. Only the occlusion body protein-encoding gene of PmNV has been molecularly characterised, revealing no homology to any other genes deposited in Genbank (Chaivisuthangkura et al. 2008).

Function	Name	OrNV	GbNV	HzNV-1	PmNV
DNA replication, repair, and recombination	<i>dnapol</i>	1	12	131	N.d.
	<i>helicase</i>	34	88	104	N.d.
	<i>helicase 2</i>	108	46	60	N.d.
	<i>integrase</i>	75	57	144	+
	<i>ligase</i>	121	38	36	N.d.
Nucleotide metabolism	<i>lef-3</i>	59	86	-	N.d.
	<i>rr1</i>	51	82	95	N.d.
	<i>rr2</i>	102	63	73	N.d.
	<i>tk</i>	58	74	115	+
	<i>tk</i>	117	34	111	+
Transcription	<i>tk</i>	125	44	71	+
	<i>tk</i>	137	17	51	+
	<i>p47</i>	20	69	75	N.d.
	<i>lef-4</i>	42	96	98	N.d.
	<i>lef-8</i>	64	49	90	N.d.
Oral infectivity	<i>lef-9</i>	96	24	75	+
	<i>lef-5</i>	52	85	101	+
	<i>vlf-1</i>	30	80	121	+
	<i>p74</i>	126	45	11	N.d.
	<i>pif-1</i>	60	52	55	N.d.
Packaging, assembly, and morphogenesis	<i>pif-2</i>	17	66	123	N.d.
	<i>pif-3</i>	107	3	88	N.d.
	<i>19 kda/pif-4</i>	33	87	103	N.d.
	<i>odv-e56/pif-5</i>	115	5	76	N.d.
	<i>polh/gran</i>	16	65	69	N.d.
Inhibition of apoptosis	<i>ac68</i>	72	55	74	N.d.
	<i>38 K</i>	87	1	10	+
Unknown function	<i>iap-3</i>	134	98	138	N.d.
	<i>vp39</i>	15	64	89	N.d.
	<i>vp91</i>	106	2	46	N.d.
	<i>ac81</i>	4	14	33	N.d.
	<i>ac92</i>	113	7	13	N.d.
		47	19	30	N.d.
		76	58	143	N.d.
		3	13	-	N.d.
		18	67	-	N.d.
		22	72	-	N.d.



Function	Name	OrNV	GbNV	HzNV-1	PmNV
		23	74	-	N.d.
		24	75	-	N.d.
		25	76	-	N.d.
		27	78	-	N.d.
		29	81	-	N.d.
		39	93	-	N.d.
		40	94	-	N.d.
		41	95	-	N.d.
		44	97	-	N.d.
		45	23	-	N.d.
		46	22	-	N.d.
		53	84	-	N.d.
		54	83	-	N.d.
		61	51	-	N.d.
		79	59	-	N.d.
		80	60	-	N.d.
		86	61	-	N.d.
		90	28	-	N.d.
		95	9	-	N.d.
		104	62	-	N.d.
		105	43	-	N.d.
		114	6	-	N.d.
		116	33	-	N.d.
		118	35	-	N.d.
		119	36	-	N.d.
		120	37	-	N.d.
		122	39	-	N.d.
		123	41	-	N.d.
		132	48	-	N.d.
		6	-	109	N.d.
		-	-	52	+
		-	-	64	+
		-	-	93	+
		-	-	118	+
		-	-	141	+

Table 2. Homologous genes conserved in nudiviruses. -: Absent; +: Present; N.d.: Not determined. The predicted ORFs in nudiviruses are presented in number. Homologues to baculovirus core genes are marked in bold face.

## 5. Nudivirus phylogeny

### 5.1 Phylogenetic analysis

Due to the poorness of information of other distinguishing features, single gene phylogeny and/or phylogenomics became the most important approach to delineate the relationship of concerned viruses on strain and species level. However, single gene phylogenies have fallen

increasingly into disfavor given the recognition that gene trees can often differ substantially from the underlying species tree due to a variety of evolutionary events in addition to simply stochastic or analytical error. This is likely to be especially true in DNA viruses with the substantial evolutionary dynamics intrinsic to their genomes. Thus, (i) horizontal gene transfer (HGT), i.e., exchange with other viruses, symbiotic bacteria and hosts; (ii) homologous or non-homologous recombination with other viruses; (iii) gene/domain duplication and rearrangement; and (iv) lineage specific gene loss/expansion all impose significant complications on both the bioinformatic detection of orthologous genes and on the accuracy of the resulting gene trees with respect to the overall species tree (Shackelton and Holmes 2004).

To overcome these problems, a set of conserved genes were analysed using both the supertree and supermatrix approaches. Multiple sequence alignments of individual genes were performed using any of T-Coffee (Notredame, Higgins and Heringa 2000), MUSCLE (Edgar 2004), ClustalW/X (Chenna et al. 2003), MAFFT (Katoh et al. 2002) and Kalign (Lassmann and Sonnhammer 2005), and were manually refined as needed. Sequence alignment quality was assessed by using MUMSA (Lassmann and Sonnhammer 2005). In particular, 20 of the 30 baculovirus core genes (Table 2) were analysed, considering that they are evolutionarily more conserved than other nonessential genes and that homologues to all or most of them are present in NVs, SGHVs, and WSSV as well as, albeit more distantly, in other large eukaryotic dsDNA viruses such as NCLDV (nucleocytoplasmic large DNA viruses) and herpesviruses (Table 3).

NALDV			NCLDV							Herpes-virus		Homologue	
Baculo-virus	Nudi-virus	SGHV	WSSV	ASFV	Pox-virus	Irido-virus	Asco-virus	Phy-codnavirus	Mimi-virus		Definition	Putative function	
■	■	■	■	■	■	■	■	■	■	■	DNA polymerase B	DNA replication, recombination and repair	
■	■	■	-	■	■	■	■	■	-	Helicase / Helicase III (D5R like helicase)			
■	■	■	-	■	■	■	■	□	■	-	LEF-8 / Rpb2	Transcription	
■	■	■	-	■	■	■	■	□	■	-	LEF-9 / Rpb1		
■	■	■	-	■	■	■	□	■	■	-	LEF-5 / TFIIIS-like transcription factor		
■	■	■	■	■	■	■	■	■	■	-	P33 / Thiol oxidoreductase	Virion packaging and morphogenesis	

Table 3. Ancient core genes identified in NALDVs, WSSV, NCLDVs, and herpesviruses. Black squares, homologue detected in all available genomes; grey squares, in many but not in all available genomes; white squares, in few available genomes; -, not detected. Homologue definition, gene name in NALDVs / in NCLDVs. Ascovirus is considered to be a member of the NCLDVs because of its high sequence similarity to iridovirus.

The supertree and supermatrix framework represent alternative strategies to the issue of data combination. In the supermatrix approach, all the primary character data are combined into a single supermatrix that is analysed using standard phylogenetic methods (de Queiroz and Gatesy 2007). By contrast, the supertree approach combines phylogenetic trees derived from individual partitions of the full data set (here the individual gene trees) to likewise derive a single, joint phylogenetic estimate (Bininda-Emonds 2004a). Thus, the supertree approach addresses conflict and congruence at the level of the source trees rather than at the level of the primary data (Bininda-Emonds 2004b). Although this approach has been

criticised because of the inherent loss of information (among others, see de Queiroz and Gatesy 2007), numerous simulation studies have demonstrated that this loss of information is not detrimental in practice (see Bininda-Emonds 2004a). Moreover, the contrasting approaches of the supertree and supermatrix frameworks form the basis of the global congruence framework (Bininda-Emonds 2004b), whereby increased confidence is placed in those clades common to both approaches and increased attention is demanded on conflicting solutions, particularly when each is strongly supported.

For the supertree analyses, phylogenetic analyses of the individual gene trees were performed under a maximum likelihood (ML) framework using RAxML 7.0.4 (Stamatakis, Hoover and Rougemont 2008). Optimal substitution matrices for each amino acid data were selected initially using the Perl script ProteinModelSelector (<http://icwww.epfl.ch/~stamatak/index-Dateien/Page443.htm>) as implemented in batchRAxML (<http://www.molekulare.systematik.uni-oldenburg.de/33997.html>) and then applied for the full ML analysis of each gene tree. In all cases, rate heterogeneity between sites was accounted for using the CAT approximation of the gamma distribution (Stamatakis 2006). The former is an approximation of the latter that is both computationally more efficient in terms of its memory demands and overall speed, and provides equivalent results (Stamatakis 2006). However, all final likelihood values were obtained under a true gamma distribution. ML analysis used the new fast bootstrapping approach (Stamatakis et al. 2008) that simultaneously obtained the ML tree as well as estimates of nodal support based on a non-parametric bootstrap (Felsenstein 1985). Bootstrap values were based on 1000 replicates. Gene trees were rooted on the herpesviruses HHV-3, HHV-4, and HHV-5 (as a monophyletic group) because they share the minimum number of conserved ancestral genes with the other viruses (Table 3); trees lacking herpesviruses were treated as unrooted.

The supertree analysis used the method of matrix representation with parsimony (MRP) (Baum 1992, Ragan 1992), whereby the topology of each gene tree was then encoded using additive binary coding: for each node in turn, all taxa descended from that node are scored as "1", all taxa otherwise present on the tree are scored as "0", and all remaining taxa as "?". Semi-rooted coding was employed in that rooted gene trees included an all-zero fictitious outgroup taxon to root the supertree; for unrooted gene trees, this taxon was coded using "?" (Bininda-Emonds, Beck and Purvis 2005). The matrix representations of all source trees were then combined into a single matrix that was analyzed using maximum parsimony (MP). Individual pseudocharacters in the matrix were weighted according to the bootstrap support of their corresponding nodes, a procedure that improves the accuracy of the supertree analysis by helping account for differential support within the primary character matrices (Bininda-Emonds and Sanderson 2001). MP searches in PAUP\* v4.0b10 (Swofford 2002) used a heuristic search strategy based on a random addition sequence (10000 replicates), TBR branch swapping, and with up to 50000 equally most parsimonious trees (MPTs) being saved. The supertree was taken to be the 50% majority-rule consensus of all MPTs. Support for the nodes in the supertree was estimated using the rQS index (Bininda-Emonds et al. 2003, Price, Bininda-Emonds and Gittleman 2005) restricted to informative gene trees only; analyses used the Perl script QualiTree (<http://www.molekularesystematik.uni-oldenburg.de/33997.html>).

The rQS index measures the number of gene trees that explicitly support or conflict with a given node on the supertree. Values of 1 and -1 indicate universal support or conflict, respectively, among the set of gene trees (Fig. 4).

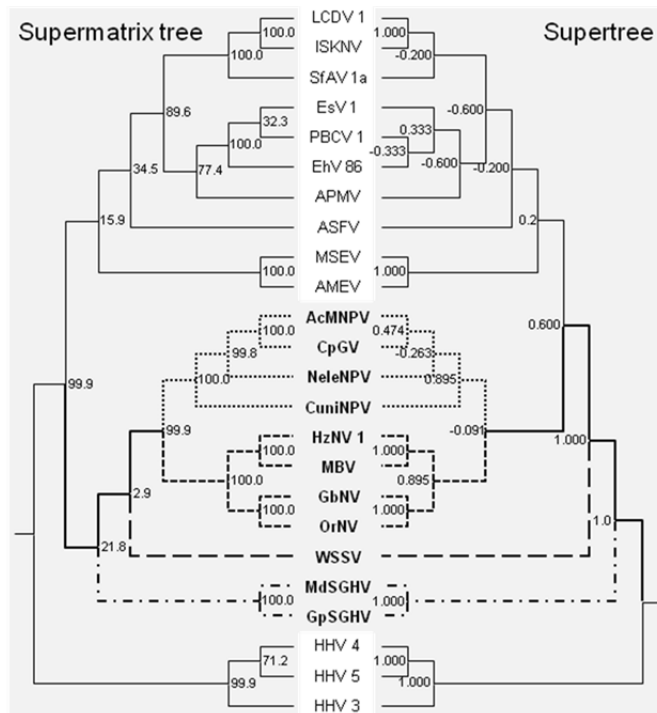


Fig. 4. Combined data trees based on the 20 conserved baculovirus core gene sequences: (A) ML supermatrix tree derived from a simultaneous analysis of the concatenated sequences and (B) weighted MRP supertree of the 20 gene trees in Wang et al, 2011. The latter represents the 50% majority-rule consensus of 71 equally most parsimonious solutions. Nodal support is given as non-parametric bootstrap frequencies ( $n = 1000$ ) determined from the supermatrix data set / degree of support among the informative source trees for a given node as measured by the rQS index. Branch lengths in (A) are proportional to the average number of substitutions per site per unit time. GenBank accession numbers for these viral genomes and virus full names are listed as follows: NC\_001623 (*Autographa californica* NPV, AcMNPV), NC\_002816 (*Cydia pomonella* GV, CpGV), NC\_005906 (*Neodiprion lecontei* NPV, NeleNPV), NC\_003084 (*Culex nigripalpus* NPV, CuniNPV), NC\_004156 (*Heliothis zea* NV 1, HzNV-1), NC\_009240 (*Gryllus bimaculatus* NV, GbNV), EU747721 (*Oryctes rhinoceros* NV, OrNV), NC\_010356 (*Glossina pallidipes* SGHV, GpSGHV), NC\_010671 (*Musca domestica* SGHV, MdSGHV), NC\_003225 (Shrimp white spot syndrome virus, WSSV), NC\_001659 (African swine fever virus, ASFV), NC\_002520 (*Amsacta moorei* EV, AMEV), NC\_001993 (*Melanoplus sanguinipes* EV, MSEV), NC\_008361 (*Spodoptera frugiperda* AV 1a, SfAV-1a), NC\_001824 (Lymphocystis disease virus 1, LCDV-1), NC\_003494 (Infectious spleen and kidney necrosis virus, ISKNV), NC\_006450 (*Acanthamoeba polyphaga* mimivirus, APMV), NC\_000852 (*Paramecium bursaria* Chlorella virus 1, PBCV-1), NC\_007346 (*Emiliania huxleyi* virus 86, EhV-86), NC\_002687 (*Ectocarpus siliculosus* virus 1, EsV-1), NC\_001348 (Human herpesvirus 3, HHV-3), NC\_001347 (Human herpesvirus 5, HHV-5), and NC\_007605 (Human herpesvirus 4, HHV-4).

For the supermatrix analysis, all individual gene data sets were concatenated into a single, larger matrix that was analyzed using RAxML. Analysis used the same method as for the individual gene trees, except that a partitioned model was used whereby each gene partition was modeled individually according to the optimal model of evolution determined previously. Support values for each tree were also estimated using the support measure for the other technique. In other words, the rQS index was also applied to the supermatrix tree to estimate the support for its nodes across the gene trees and the bootstrap values for the nodes on the supertree were estimated using the 1000 bootstrap replicate trees derived from the supermatrix analysis (Fig. 4).

## 5.2 Phylogeny and evolution

### 5.2.1 Common ancestry of NVs, baculoviruses and SGHVs

In the light of gene content analysis, an evolutionary link among NVs, baculoviruses, SGHVs and WSSV is most plausible. Consequently, it should be possible to analyze their phylogenetic relationship on the basis of their shared conserved ancestral genes. When these 20 single gene trees were inferred, most of the nodes showed medium to high bootstrap values, with average values across an entire gene tree ranging from  $57.6 \pm 10.9$  (*helicase*;  $n = 17$  nodes) and  $99.0 \pm 0.8$  (*p47*;  $n = 4$  nodes), suggesting the trees are topologically reliable on the whole (Wang et al. 2011).

The supermatrix (on the basis of the 20 core genes indicated in Table 2) and the supertree using these 20 single core gene trees in (Wang et al. 2011) analyses were performed. Both the supermatrix tree and supertree were highly congruent (Fig. 4). In both cases, the monophyly of each of the NVs, baculoviruses, and SGHVs was strongly supported, the branching patterns within each of the baculovirus and NV clades were also in good agreement with the current picture of their phylogeny, and a common ancestor of baculoviruses and NVs was suggested (Fig. 4). Hence, we recognized from both the supertree and supermatrix tree that baculoviruses and NVs are monophyletic; they can be considered as the minimally forming group that we term the nuclear arthropod-specific large DNA viruses (NALDVs). Both methods conflict in positioning the SGHVs within (or at least as sister lineage to the NALDV; supermatrix tree) or outside (supertree) the NALDV group. For each tree, the preferred position enjoys better support than that from the other analysis based on the most appropriate support measure. For instance, the supertree placement of the SGHVs has an rQS index value of 0.455 compared to a value of 0.143 supporting the grouping of SGHVs with the baculoviruses and NVs. However, whereas the supermatrix placement of the SGHVs enjoys some rQS support (0.143 as mentioned), the supertree placement has no bootstrap support whatsoever (0.6 compared to 99.4). Thus, the supermatrix placement of the SGHVs as sister lineage to baculoviruses and NVs, and the SGHVs being members of the NALDVs seem to be justifiable.

### 5.2.2 The “Monodon baculovirus” represents a nudivirus

Blastp searches revealed that a number of nudivirus homologues are present in the partially sequenced genome of the so-called “Monodon baculovirus” of the shrimp *Penaeus monodon* (21,150 bp in total; GenBank accession no. EU246943, EU246944, EF458632, AY819785).

When using the annotated shrimp MBV ORFs as query in BLAST similarity search, best hits were frequently found with HzNV-1 (Wang et al. 2011). Different phylogenetic analyses, including single gene tree inference as well as both supermatrix and supertree analyses, of the homologues of baculovirus core genes *lef-9*, *vlf-1*, *lef-5*, *38K*, revealed unequivocally an obvious relationship between MBV and the non-occluded HzNV-1 (Fig. 4) (Wang et al. 2011). Given that seven other MBV and HzNV-1 ORFs are also highly similar, it is strongly suggested to consider MBV as an occluded member of the NVs and to rename it to *Penaeus monodon* nudivirus (PmNV) (Wang et al. 2011).

### 5.2.3 WSSV might be related to the NALDVs

The position of WSSV differs between the two trees, however, being nested deep with the NALDVs in the supermatrix tree and as sister to the clade of NCLDVs plus NALDVs in the supertree (Fig. 4). Support for either position based on either the rQS index or the bootstrap is worse than that for other clades in the tree. The different positions for WSSV reflect how the two different methods used deal with the restricted, conflicting information that is available for this virus. Although WSSV shares six genes with the other viruses, only two of these (*DNA polymerase* and *p33*) are phylogenetically informative as to its potential placement with respect to the NALDV and NCLDV groups because homologue counterparts are available in both groups. The remaining four genes (*p74*, *pif-1*, *pif-2*, and *pif-3*) are restricted to baculoviruses, NVs, SGHVs, and WSSV only. The resulting trees are therefore essentially unrooted and it is not possible to determine if WSSV nests within NALDVs (contradicting the supermatrix placement) or is sister to them (consistent with both placements). Of the two informative genes, only *p33* has associated WSSV with the NALDVs; the *DNA polymerase* has grouped it within the NCLDVs (Wang et al. 2011). The supermatrix analysis is influenced largely by the relative number of amino acids (aa) supporting a given position. In the current context, *DNA polymerase* with ~3000 aa residues is clearly outweighing the ~1000 aa residues of *p33*, thereby favouring the placement of WSSV with the NCLDVs. By contrast, the supertree analysis is more sensitive to the number of trees supporting a given position and, importantly, the relative node support within those trees (in a weighted supertree analysis). Thus, although the *DNA polymerase* tree places WSSV within the NCLDVs, this position is very poorly supported and outweighed by its more robust placement within NALDVs in the *p33* tree (Wang et al. 2011). As a result, WSSV was excluded from the NCLDV group in the supermatrix tree.

Thus, the phylogenetic analyses are equivocal with respect to the evolutionary relationships of WSSV based on the current data set and more genes need to be sampled to resolve its placement. Nevertheless, other sources of evidence suggest that WSSV is more closely related to the NALDVs than to other DNA viruses. Notably, WSSV shares six conserved homologous genes with the NALDVs, but rarely possesses homologous genes with numerous other marine viruses colonising the same aquatic ecological niches. It therefore seems that WSSV is a very ancient virus that has undergone extremely divergent evolution, as witnessed by the branch lengths generally subtending this virus (Fig. 4). This fact, in turn, hampers identification of its gene homologues and reconstruction of its phylogenetic affinities using present-day alignment based methods. In contrast, when an alignment-free whole-proteome phylogenetic analysis was applied, WSSV clustered with SGHVs (Wu et

al. 2009), which coincidentally is in agreement with the presented hypothesis of WSSV's evolutionary link to the NALDVs. However, in the study by Wu et al. (2009) the SGHV and WSSV were placed within the herpesviruses, although there is no evidence of relationship among these viruses, when considering structural, biological and other genome features.

#### 5.2.4 A common ancestry of nudiviruses, baculoviruses, hytrosaviruses, and WSSV

Taking together, 20 baculovirus core gene homologues were identified in nudiviruses, 12 in SGHVs, and six in WSSV, respectively. Consequently, this shared gene content of baculoviruses, nudiviruses, SGHVs, and WSSV is an important evidence for a proposed common ancestry of these viruses. Any other explanation, e. g., horizontal gene transfer of these genes, seems to be less probable. Therefore it is proposed that baculoviruses, nudiviruses, hytrosaviruses, and WSSV most likely shared a common ancestor and form a highly diverse group of nuclear arthropod-specific large DNA viruses (Wang and Jehle 2009; Wang et al. 2011).

### 6. Acknowledgement

This work was funded by grants from Shanghai Municipal Education Commission (the Eastern Scholar Project and the Leading Academic Discipline Project) and Shanghai Municipal Science and Technology Commission (Project no. 10540503000).

### 7. References

- Abd-Alla, A. M., F. Cousserans, A. G. Parker, J. A. Jehle, N. J. Parker, J. M. Vlak, A. S. Robinson & M. Bergoin (2008) Genome analysis of a *Glossina pallidipes* salivary gland hypertrophy virus reveals a novel, large, double-stranded circular DNA virus. *J Virol*, Vol. 82, No. 9, (May 2008), pp. 4595-4611, ISSN 1098-5514
- Abd-Alla, A. M., J. M. Vlak, M. Bergoin, J. E. Maruniak, A. Parker, J. P. Burand, J. A. Jehle & D. G. Boucias (2009) Hytrosaviridae: a proposal for classification and nomenclature of a new insect virus family. *Arch Virol*, Vol. 154, No. 6, (2009), pp. 909-918, ISSN 1432-8798
- Bézier, A., 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 (2009) Polydnnaviruses of braconid wasps derive from an ancestral nudivirus. *Science*, Vol. 323, No. 5916, (February 2009), pp. 926-930, ISSN 1095-9203
- Baum, B. R. (1992) Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon*, Vol. 41, No. 1, (1992), pp. 3-10.
- Bininda-Emonds, O. R. (2004a) The evolution of supertrees. *Trends Ecol Evol*, Vol. 19, No. 6, (June 2004), pp. 315-322, ISSN 0169-5347
- Bininda-Emonds, O. R. (2004b) Trees versus characters and the supertree/supermatrix "paradox". *Syst Biol*, Vol. 53, No. 2, (April 2004), pp. 356-359, ISSN 1063-5157

- Bininda-Emonds, O. R., R. M. Beck & A. Purvis (2005) Getting to the roots of matrix representation. *Syst Biol*, Vol. 54, No. 4, (August 2005), pp. 668-672, ISSN 1063-5157
- Bininda-Emonds, O. R., K. E. Jones, S. A. Price, R. Grenyer, M. Cardillo, M. Habib, A. Purvis & J. L. Gittleman (2003) Supertrees are a necessary not-so-evil: a comment on Gatesy et al. *Syst Biol*, Vol. 52, No. 5, (October 2003), pp. 724-729, ISSN 1063-5157
- Bininda-Emonds, O. R. & M. J. Sanderson (2001) Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. *Syst Biol*, Vol. 50, No. 4, (August 2001), pp. 565-579, ISSN 1063-5157
- Burand, J. P. 1998. Nudiviruses. In *The Insect Viruses*, eds. L. K. Miller & L. A. Ball, 69-90. New York: Plenum Press.
- Burand, J. P. & H. Lu (1997) Replication of a Gonad-Specific Insect Virus in TN-368 Cells in Culture. *J Invertebr Pathol*, Vol. 70, No. 2, (September 1997), pp. 88-95, ISSN 1096-0805
- Burand, J. P. & C. P. Rallis (2004) In vivo dose-response of insects to Hz-2V infection. *Virology*, Vol. 1, No. (2004), pp. 15, ISSN 1743-422X
- Chaivisuthangkura, P., C. Tawilert, T. Tejangkura, S. Rukpratanporn, S. Longyant, W. Sithigorngul & P. Sithigorngul (2008) Molecular isolation and characterization of a novel occlusion body protein gene from *Panaeus monodon* nucleopolyhedrovirus. *Virology*, Vol. 381, No. 2, (November 2008), pp. 261-267, ISSN 1096-0341
- Chao, Y. C., S. T. Lee, M. C. Chang, H. H. Chen, S. S. Chen, T. Y. Wu, F. H. Liu, E. L. Hsu & R. F. Hou (1998) A 2.9-kilobase noncoding nuclear RNA functions in the establishment of persistent Hz-1 viral infection. *J Virol*, Vol. 72, No. 3, (March 1998), pp. 2233-2245, ISSN 0022-538X
- Chao, Y. C., H. A. Wood, C. Y. Chang, H. J. Lee, W. C. Shen & H. T. Lee (1992) Differential expression of Hz-1 baculovirus genes during productive and persistent viral infections. *J Virol*, Vol. 66, No. 3, (March 1992), pp. 1442-1448, ISSN 0022-538X
- Cheng, C. H., S. M. Liu, T. Y. Chow, Y. Y. Hsiao, D. P. Wang, J. J. Huang & H. H. Chen (2002) Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the *Baculoviridae*. *J Virol*, Vol. 76, No. 18, (September 2002), pp. 9024-9034, ISSN 0022-538X
- Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins & J. D. Thompson (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*, Vol. 31, No. 13, (July 2003), pp. 3497-3500, ISSN 1362-4962
- Crawford, A. M. & C. Sheehan (1985) Replication of *Oryctes* baculovirus in cell culture: viral morphogenesis, infectivity and protein synthesis. *J Gen Virol.*, Vol. 66, No. (1985), pp. 529-539.
- de Queiroz, A. & J. Gatesy (2007) The supermatrix approach to systematics. *Trends Ecol Evol*, Vol. 22, No. 1, (January 2007), pp. 34-41, ISSN 0169-5347
- Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, Vol. 5, (August 2004), pp. 113, ISSN 1471-2105
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, Vol. 39, (1985), pp. 783-791.



- Garcia-Maruniak, A., A. M. Abd-Alla, T. Z. Salem, A. G. Parker, V. U. Lietze, M. M. van Oers, J. E. Maruniak, W. Kim, J. P. Burand, F. Cousserans, A. S. Robinson, J. M. Vlák, M. Bergoin & D. G. Boucias (2009) Two viruses that cause salivary gland hypertrophy in *Glossina pallidipes* and *Musca domestica* are related and form a distinct phylogenetic clade. *J Gen Virol*, Vol. 90, Pt 2, (February 2009), pp. 334-346, ISSN 0022-1317
- Garcia-Maruniak, A., J. E. Maruniak, W. Farmerie & D. G. Boucias (2008) Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of *Musca domestica*, MdSGHV. *Virology*, Vol. 377, No. 1, (July 2008), pp. 184-196, ISSN 0042-6822
- Granados, R. R., T. Nguyen & B. Cato (1978) An insect cell line persistently infected with a baculovirus-like particle. *Intervirology*, Vol. 10, No. 5, (1978), pp. 309-317, ISSN 0300-5526
- Guttieri, M. C. & J. P. Burand (2001) Location, nucleotide sequence, and regulation of the p51 late gene of the hz-1 insect virus: identification of a putative late regulatory element. *Virus Genes*, Vol. 23, No. 1, (2001), pp. 17-25, ISSN 0920-8569
- Guttieri, M. C. & J. P. Burand (1996) Nucleotide sequence, temporal expression, and transcriptional mapping of the p34 late gene of the Hz-1 insect virus. *Virology*, Vol. 223, No. 2, (September 1996), pp. 370-375, ISSN 0042-6822
- Herniou, E. A. & J. A. Jehle (2007) Baculovirus phylogeny and evolution. *Curr Drug Targets*, Vol. 8, No. 10, (October 2007), pp. 1043-1050, ISSN 1873-5592
- Herniou, E. A., J. A. Olszewski, J. S. Cory & D. R. O'Reilly (2003) The genome sequence and evolution of baculoviruses. *Annu Rev Entomol*, Vol. 48, (2003), pp. 211-234, ISSN 0066-4170
- Huger, A. M. (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*, Vol. 8, No. 1, (March 1966), pp. 38-51, ISSN 0022-2011
- Huger, A. M. (1985) A new virus disease of crickets (Orthoptera: Gryllidae) causing macronucleosis of fatbody. *J. Invertebr. Pathol.*, Vol. 45, No. 1, (1985), pp. 108-111.
- Huger, A. M. & A. Krieg. 1991. Baculoviridae. Nonoccluded Baculoviruses. In *Atlas of Invertebrate Viruses*, eds. J. R. Adams & J. R. Bonami, 287-319. Boca Raton: CRC Press, Inc.
- Jackson, T. A., A. M. Crawford & T. R. Glare (2005) *Oryctes* virus - time for a new look at a useful biocontrol agent. *J Invertebr Pathol.*, Vol. 89, No. 1, (2005), pp. 91-94.
- Jehle, J. A. & H. Backhaus (1994) Genome organization of the DNA-binding protein gene region of *Cryptophlebia leucotreta* granulosis virus is closely related to that of nuclear polyhedrosis viruses. *J Gen Virol*, Vol. 75 ( Pt 7), (July 1994), pp. 1815-1820. ISSN 0022-1317
- Jehle, J. A., G. W. Blissard, B. C. Bonning, J. S. Cory, E. A. Herniou, G. F. Rohrmann, D. A. Theilmann, S. M. Thiem & J. M. Vlák (2006a) On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol*, Vol. 151, No. 7, (July 2006), pp. 1257-1266, ISSN 0304-8608

- Jehle, J. A., M. Lange, H. Wang, Z. Hu, Y. Wang & R. Hauschild (2006b) Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology*, Vol. 346, No. 1, (March 2006), pp. 180-193, ISSN 0042-6822
- Katoh, K., K. Misawa, K. Kuma & T. Miyata (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*, Vol. 30, No. 14, (July 2002), pp. 3059-3066, ISSN 1362-4962
- Kelly, D. C., T. Lescott, M. D. Ayres, D. Carey, A. Coutts & K. A. Harrap (1981) Induction of a nonoccluded baculovirus persistently infecting *Heliothis zea* cells by *Heliothis armigera* and *Trichoplusia ni* nuclear polyhedrosis viruses. *Virology*, Vol. 112, No. 1, (July 1981), pp. 174-189, ISSN 0042-6822
- Lassmann, T. & E. L. Sonnhammer (2005) Automatic assessment of alignment quality. *Nucleic Acids Res*, Vol. 33, No. 22, (2005), pp. 7120-7128, 1362-4962
- Lin, C. L., J. C. Lee, S. S. Chen, H. A. Wood, M. L. Li, C. F. Li & Y. C. Chao (1999) Persistent Hz-1 virus infection in insect cells: evidence for insertion of viral DNA into host chromosomes and viral infection in a latent status. *J Virol*, Vol. 73, No. 1, (January 1999), pp. 128-139, ISSN 0022-538X
- Lu, H. & J. P. Burand (2001) Replication of the gonad-specific virus Hz-2V in Ld652Y cells mimics replication in vivo. *J Invertebr Pathol*, Vol. 77, No. 1, (January 2001), pp. 44-50, ISSN 0022-2011
- Mayo, M. A. 1995. Unassigned Viruses. In *Virus Taxonomy: The Sixth Report of the International Committee on Taxonomy of Viruses*, eds. 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, 504-507. Wien: Springer-Verlag.
- Notredame, C., D. G. Higgins & J. Heringa (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, Vol. 302, No. 1, (September 2000), pp. 205-217, ISSN 0022-2836
- Payne, C. C. (1974) The isolation and characterization of a virus from *Oryctes rhinoceros*. *J Gen Virol*, Vol. 25, No. 1, (October 1974), pp. 105-116, ISSN 0022-1317
- Payne, C. C., D. Compson & S. M. de looze (1977) Properties of the nucleocapsids of a virus isolated from *Oryctes rhinoceros*. *Virology*, Vol. 77, No. 1, (1977), pp. 269-280.
- Price, S. A., O. R. Bininda-Emonds & J. L. Gittleman (2005) A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). *Biol Rev Camb Philos Soc*, Vol. 80, No. 3, (August 2005), pp. 445-473, ISSN 1464-7931
- Ragan, M. A. (1992) Phylogenetic inference based on matrix representation of trees. *Mol Phylogenet Evol*, Vol. 1, No. 1, (March 1992), pp. 53-58, ISSN 1055-7903
- Raina, A. K., J. R. Adams, B. Lupiani, D. E. Lynn, W. Kim, J. P. Burand & E. M. Dougherty (2000) Further characterization of the gonad-specific virus of corn earworm, *Helicoverpa zea*. *J Invertebr Pathol*, Vol. 76, No. 1, (July 2000), pp. 6-12, ISSN 0022-2011
- Raina, A. K. & B. Lupiani (2006) Acquisition, persistence, and species susceptibility of the Hz-2V virus☆. *Journal of Invertebrate Pathology*, Vol. 93, No. 2, (2006), pp. 71-74, ISSN 0022-2011
- Shackelton, L. A. & E. C. Holmes (2004) The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol.*, Vol. 12, No. 10, (2004), pp. 458-465

- Slack, J. & B. M. Arif (2007) The baculoviruses occlusion-derived virus: virion structure and function. *Adv Virus Res*, Vol. 69, (2007), pp. 99-165, ISSN 0065-3527
- Stamatakis, A. 2006. Phylogenetic Models of Rate Heterogeneity: A High Performance Computing Perspective. In *Proceedings of 20th IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS2006), High Performance Computational Biology Workshop*. Rhodos, Greece.: Proceedings on CD.
- Stamatakis, A., P. Hoover & J. Rougemont (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol*, Vol. 57, No. 5, (October 2008), pp. 758-771, ISSN 1076-836X
- Swofford, D. L. 2002. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). 4, Sinauer Associates, Sunderland, Massachusetts
- van Oers, M. M. & J. M. Vlak (2007) Baculovirus genomics. *Curr Drug Targets*, Vol. 8, No. 10, (October 2007), pp. 1051-1068, ISSN 1873-5592
- Wang, Y., O. R. Bininda-Emonds, M. M. van Oers, J. M. Vlak & J. A. Jehle (2011) The genome of *Oryctes rhinoceros* nudivirus provides novel insight into the evolution of nuclear arthropod-specific large circular double-stranded DNA viruses. *Virus Genes*, Vol. 42, No. 3, (June 2011), pp. 444-456, ISSN 1572-994X
- Wang, Y., J. P. Burand & J. A. Jehle (2007a) Nudivirus genomics: diversity and classification. *Virologica Sinica*, Vol. 22, No.1, (2007), pp. 128-136
- Wang, Y. & J. A. Jehle (2009) Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: new insights on an old topic. *J Invertebr Pathol*, Vol. 101, No. 3, (July 2009), pp. 187-193, ISSN 1096-0805
- Wang, Y., R. G. Kleespies, A. M. Huger & J. A. Jehle (2007b) The genome of *Gryllus bimaculatus* nudivirus indicates an ancient diversification of baculovirus-related nonoccluded nudiviruses of insects. *J Virol*, Vol. 81, No. 10, (May 2007), pp. 5395-5406, ISSN 0022-538X
- Wang, Y., R. G. Kleespies, M. B. Ramle & J. A. Jehle (2008) Sequencing of the large dsDNA genome of *Oryctes rhinoceros* nudivirus using multiple displacement amplification of nanogram amounts of virus DNA. *J Virol Methods*, Vol. 152, No. 1-2, (September 2008), pp. 106-108, ISSN 0166-0934
- Wang, Y., M. M. van Oers, A. M. Crawford, J. M. Vlak & J. A. Jehle (2007c) Genomic analysis of *Oryctes rhinoceros* virus reveals genetic relatedness to *Heliothis zea* virus 1. *Arch Virol*, Vol. 152, No. 3, (2007), pp. 519-531, ISSN 0304-8608
- Wood, H. A. & J. P. Burand (1986) Persistent and productive infections with the Hz-1 baculovirus. *Curr Top Microbiol Immunol*, Vol. 131, (1986), pp. 119-133, ISSN 0070-217X
- Wu, G. A., S. R. Jun, G. E. Sims & S. H. Kim (2009) Whole-proteome phylogeny of large dsDNA virus families by an alignment-free method. *Proc Natl Acad Sci U S A*, Vol. 106, No. 31, (August 2009), pp. 12826-12831, ISSN 1091-6490
- Wu, Y. L., C. Y. Liu, C. P. Wu, C. H. Wang, S. T. Lee & Y. C. Chao (2008) Cooperation of ie1 and p35 genes in the activation of baculovirus AcMNPV and HzNV-1 promoters. *Virus Res*, Vol. 135, No. 2, (August 2008), pp. 247-254, ISSN 0168-1702

Wu, Y. L., C. P. Wu, S. T. Lee, H. Tang, C. H. Chang, H. H. Chen & Y. C. Chao (2010) The early gene hhi1 reactivates *Heliothis zea* nudivirus 1 in latently infected cells. *J Virol*, Vol. 84, No. 2, (January 2010), pp. 1057-1065, ISSN 1098-5514