Journal of General Virology (2009), **90**, 334–346 **DOI** 10.1099/vir.0.006783-0

Two viruses that cause salivary gland hypertrophy in *Glossina pallidipes* and *Musca domestica* are related and form a distinct phylogenetic clade

Alejandra Garcia-Maruniak,[1]† Adly M. M. Abd-Alla,[2]† Tamer Z. Salem,[1,3] Andrew G. Parker,[2] Verena-Ulrike Lietze,[1] Monique M. van Oers,[4] James E. Maruniak,[1] Woojin Kim,[5] John P. Burand,[5] François Cousserans,[6] Alan S. Robinson,[2] Just M. Vlak,[4] Max Bergoin[6] and Drion G. Boucias[1]

Correspondence
Alejandra Garcia-Maruniak
alemar@ufl.edu

[1]Department of Entomology and Nematology, PO Box 110620, University of Florida, Gainesville, FL 32611-0620, USA

[2]Entomology Unit, FAO/IAEA Agriculture & Biotechnology Laboratory, IAEA Laboratories Seibersdorf, A-2444 Seibersdorf, Austria

[3]Department of Microbial Molecular Biology, AGERI, Agricultural Research Center, Giza 12619, Egypt

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

[5]Department of Plant, Soil and Insect Science, University of Massachusetts, Amherst, MA 01003, USA

[6]Laboratoire de Pathologie Comparée, Université Montpellier II, Montpellier, France

Glossina pallidipes and *Musca domestica* salivary gland hypertrophy viruses (GpSGHV and MdSGHV) replicate in the nucleus of salivary gland cells causing distinct tissue hypertrophy and reduction of host fertility. They share general characteristics with the non-occluded insect nudiviruses, such as being insect-pathogenic, having enveloped, rod-shaped virions, and large circular double-stranded DNA genomes. MdSGHV measures 65×550 nm and contains a 124 279 bp genome (~44 mol% G+C content) that codes for 108 putative open reading frames (ORFs). GpSGHV, measuring 50×1000 nm, contains a 190 032 bp genome (28 mol% G+C content) with 160 putative ORFs. Comparative genomic analysis demonstrates that 37 MdSGHV ORFs have homology to 42 GpSGHV ORFs, as some MdSGHV ORFs have homology to two different GpSGHV ORFs. Nine genes with known functions (*dnapol*, *ts*, *pif-1*, *pif-2*, *pif-3*, *mmp*, *p74*, *odv-e66* and *helicase-2*), a homologue of the conserved baculovirus gene *Ac81* and at least 13 virion proteins are present in both SGHVs. The amino acid identity ranged from 19 to 39 % among ORFs. An (A/T/G)TAAG motif, similar to the baculovirus late promoter motif, was enriched 100 bp upstream of the ORF transcription initiation sites of both viruses. Six and seven putative microRNA sequences were found in MdSGHV and GpSGHV genomes, respectively. There was genome. Collinearity between the two SGHVs, but not between the SGHVs and the nudiviruses. Phylogenetic analysis of conserved genes clustered both SGHVs in a single clade separated from the nudiviruses and baculoviruses. Although MdSGHV and GpSGHV are different viruses, their pathology, host range and genome composition indicate that they are related.

Received 25 August 2008
Accepted 24 October 2008

INTRODUCTION

A group of double-stranded (ds) DNA viruses, referred to as salivary gland hypertrophy viruses (SGHV), has been detected in the house fly *Musca domestica* (Coler *et al.*, 1993), the narcissus bulb fly *Merodon equestris* (Amargier *et al.*, 1979) and various tsetse fly *Glossina* species (Jaenson, 1978; Otieno *et al.*, 1980; Gouteux, 1987; Minter-Goedbloed & Minter, 1989; Shaw & Moloo, 1993). SGHV infection in the nuclei of salivary gland cells, leads to salivary gland hypertrophy (SGH). Infected adult house flies do not exhibit any overt disease symptoms, but in

†These authors contributed equally to this work.

Supplementary material is available with the online version of this paper.

tsetse flies infected males can be distinguished by a rounded, enlarged abdomen (A. M. M. Abd-Alla and others, unpublished data). The viral infection of *M. domestica* salivary gland hypertrophy virus (MdSGHV) and *Glossina pallipides* salivary gland hypertrophy virus (GpSGHV) is known to reduce reproduction in the house fly and tsetse fly, respectively (Sang *et al.*, 1998, 1999; Lietze *et al.*, 2007).

Most of the research has focused on GpSGHV and MdSGHV. Although both viruses induce SGH, they possess distinct structural and molecular characteristics. Both viruses are enveloped, rod-shaped with a circular dsDNA genome, and both contain a complex of major and minor structural peptides (Coler *et al.*, 1993; Garcia-Maruniak *et al.*, 2008). The MdSGHV, measuring approximately 65 × 550 nm, contains a 124 279 bp genome (~44 mol% G+C content) that codes for 108 methionine-initiated putative open reading frames (ORFs) (Garcia-Maruniak *et al.*, 2008). GpSGHV, measuring 50 × 1000 nm (Abd-Alla *et al.*, 2007), contains a 190 032 bp genome (28 mol% G+C content) with 160 putative ORFs (Abd-Alla *et al.*, 2007, 2008). In addition to the biochemical differences, there are some pathological differences between GpSGHV and MdSGHV that may reflect adaptations to their respective host systems. The tissue tropism differs between these viruses. MdSGHV appears to only replicate in the salivary gland tissue and only sterilizes female house flies (Lietze *et al.*, 2007; Geden *et al.*, 2008). No morphological aberrations have been observed in female accessory glands, in spermathecae, or in the male ejaculatory tract, and testes of infected males contain viable sperm (Lietze *et al.*, 2007). GpSGHV, in addition to infecting salivary glands, has been reported to replicate in the female milk gland as well as in gonadal tissues, resulting in testicular degeneration and ovarian abnormalities, thus reducing the reproductive potential in both sexes of the host (Sang *et al.*, 1998, 1999).

The major mode of MdSGHV transmission within adult house flies is believed to be horizontal by feeding on contaminated food sources, and the field incidence of MdSGHV in feral house fly populations may reach up to 34 % (Geden *et al.*, 2008). In contrast, the natural incidence of the tsetse SGHV is lower, with only 0.4–15.6 % of the field-collected flies displaying SGH (Jaenson, 1978; Otieno *et al.*, 1980; Odindo *et al.*, 1981; Odindo, 1982; Ellis & Maudlin, 1987; Jura *et al.*, 1988). In nature, the major mode of SGHV transmission in tsetse flies is believed to be vertical from mother to offspring either by transovum transfer of the virus to the embryo (Jura *et al.*, 1989) or via virus-infected milk glands to the developing larvae (Sang *et al.*, 1996).

Symptomatic expression of SGHV infection has been found in collapsed tsetse fly colonies such as the Ethiopian colony of *G. pallidipes* maintained in the Entomology Unit of the FAO/IAEA Agriculture and Biotechnology Laboratory, Seibersdorf, Austria (Abd-Alla *et al.*, 2007). Unlike the situation in nature, horizontal transmission of GpSGHV has been implicated in colonized tsetse fly populations that feed in cohorts on blood-containing membrane systems. However, in most cases the incidence of SGH symptoms is low in colonized tsetse, for example in *Glossina morsitans* colonies the incidence of symptomatic SGH ranged from 1.1 (Jura *et al.*, 1993) to 4.0 % (Kokwaro *et al.*, 1990). Utilizing a non-destructive PCR detection method, Abd-Alla *et al.* (2007) detected GpSGHV PCR positives in 100 % of the colonized tsetse flies that did not exhibit SGH. Similar screening by PCR detected MdSGHV positives in less than 0.1 % of the tested colonized *M. domestica* flies (V.-U. Lietze, unpublished data).

The genome sequences from GpSGHV (Abd-Alla *et al.*, 2008) and MdSGHV (Garcia-Maruniak *et al.*, 2008) show that they were not closely related to any other large entomopathogenic DNA viruses available in the current databases. In this study, we analyse these two viruses and compare the extent of homology between the ORFs, the putative promoter regions, the genome collinearity, and phylogenetic relatedness of the two SGHVs to each other and to other large invertebrate DNA viruses. This information provides a basis to determine the relationship between these two viruses, explain their key biological features and outline to what extent they are different from other large dsDNA invertebrate viruses.

METHODS

Comparison of the MdSGHV and GpSGHV ORFs. All putative ORFs identified in GpSGHV (GenBank accession no. EF568108) (Abd-Alla *et al.*, 2008) and MdSGHV (GenBank accession no. EU522111) (Garcia-Maruniak *et al.*, 2008) were compared to the GenBank protein database using the National Center for Biotechnology Information BLAST searches (Altschul *et al.*, 1990, 1997), including the non-redundant standard protein–protein BLAST (BLASTP) searches for viruses (taxid 10239). ORFs were considered homologous when the E-values were 0.01 or lower. The percentage of amino acid identity between the GpSGHV and MdSGHV ORFs and the size of the compared amino acid regions were also annotated. Since the SGHVs share general characteristics with the non-occluded insect nudiviruses and shrimp nimaviruses, the genomes from *Heliothis zea* 1 nudivirus (HzNV-1) (Cheng *et al.*, 2002), *Gryllus bimaculatus* nudivirus (GbNV) (Wang *et al.*, 2007b), *Oryctes rhinoceros* nudivirus (partial sequence) (Wang *et al.*, 2007c) and the shrimp white spot syndrome virus (WSSV) (van Hulten *et al.*, 2001) were also compared to MdSGHV and GpSGHV separately. Also in this case, homology was only accepted when the BLASTP E-values were lower than 0.01.

Genomic organization. The genomic organization of MdSGHV and GpSGHV was compared according to the gene order by gene parity plot analysis (Hu *et al.*, 1998) and by the collinearity of conserved regions (not specifically ORFs) with a syntenic map (Lauzon *et al.*, 2006; Wolff *et al.*, 2008). In both analyses, the two genomes were linearized taking the *dnapol* gene as the first ORF. For the syntenic map, the genomes were aligned using the BLAST2 sequence tool with the TBLASTX program. They were translated in all six reading frames and the encoded peptides were compared. The DOUBLE ACT v2 program was used to produce the input comparison file for both genomes to be used with the Artemis Comparative Tool (ACT)

provided by the Sanger Centre (http://www.sanger.ac.uk/Software/ACT/). The cut-off score used with the DOUBLE ACT v2 program was 30. The high-score pairs (HSPs) accepted for these analyses were from 30 to 161, with identities ranging from 21 to 73 %. In order to compare MdSGHV and GpSGHV to HzNV-1 or GbNV, each combination was used with DOUBLE ACT v2 to produce a comparison file for each pair (MdSGHV/HzNV-1, GpSGHV/HzNV-1, MdSGHV/GbNV, GpSGHV/GbNV and MdSGHV/GpSGHV). ACT was used to compare the complete genome content of MdSGHV and GpSGHV to HzNV-1 or GbNV in two separate syntenic maps.

Identification of pre-microRNA (miRNA) sequences. Structural and thermodynamic analyses were employed to predict putative pre-miRNA candidates from MdSGHV and GpSGHV genomes. The sRNAloop program (Grad *et al.*, 2003) was used to scan the entire genome to search for stem–loop structures (parameters: hairpin length no more than 75 nt, loop longer than 3 nt and cut-off threshold score of 22). After removing redundant and low mol% G+C entries, the screened stem–loop sequences were analysed by a structure-based analysis tool with a cut-off threshold score of 90 (Ritchie *et al.*, 2007). Finally, RNAfold (Hofacker, 2003) was used to screen thermodynamically stable pre-miRNA sequences that have less than -25 kcal mol^{-1} of free energy.

Identification of promoter motifs. In order to identify putative promoter motifs, the 100 and 200 bp regions upstream of the predicted ORFs in GpSGHV and MdSGHV were analysed for the relative enrichment of 4 and 5 mer sequence motifs. The occurrence of the motifs in promoter regions was compared to their presence in the whole genome. For this analysis, the G+C mol% content of the virus was taken into account and assigned repeat regions were removed to avoid bias towards repeated sequences. Analysis was performed using scripts in Perl (http://www.perl.com/) as described previously (Marks *et al.*, 2006). A similar analysis was carried out for *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV), *Chilo iridescent* virus (CIV) and HzNV-1 for comparison.

Phylogenetic analysis. In order to establish the relationship of GpSGHV and MdSGHV to other large dsDNA virus families, phylogenetic trees of homologous genes in the respective families were constructed using CLUSTAL W protein alignment and MEGA 4 software (Molecular Evolutionary Genetics Analysis) (Tamura *et al.*, 2007) using the neighbour-joining method, and were confirmed with bootstrap analysis with heuristic search and 500 replicates.

RESULTS AND DISCUSSION

General comparison of MdSGHV and GpSGHV

Although both viruses induce SGH in their hosts, they have distinct biological, structural and molecular characteristics. The vast majority of information available for the SGHVs deals with the biological changes they inflict on their specific hosts upon infection (Table 1). The difference in their genome size (MdSGHV=124 279 bp vs GpSGHV=190 032 bp) is reflected by a difference in the viral particle size. The nucleocapsid topology of MdSGHV differs from that of GpSGHV in that a braided structure is apparent by the electron microscopy of MdSGHV virion particles (Garcia-Maruniak *et al.*, 2008) and not in the GpSGHV flexible rods (A. M. M. Abd-Alla, personal communication). To see whether these two viruses are genetically related, a comparative genomic analysis has been carried out.

Genomic comparison of MdSGHV and GpSGHV

Both viruses have a circular dsDNA genome (Abd-Alla *et al.*, 2008; Garcia-Maruniak *et al.*, 2008). The GpSGHV genome is 65 753 bp longer than the MdSGHV genome with 52 more putative ORFs and a much lower G+C content (Table 1). The number of direct repeats (*drs*) is higher in MdSGHV (18 vs 14); however, the total percentage of repeated sequences was 1.74 % of the genome in MdSGHV and 2.84 % in GpSGHV due to longer repeated regions in GpSGHV. Interestingly, seven and nine of the *drs* were found inside putative ORFs in MdSGHV and GpSGHV, respectively. Any effects caused by the *drs* on the translation of these putative proteins are still to be investigated. The function of the *drs* in these genomes are to be determined as they might serve as origins of DNA replication or enhancers of transcription.

BLAST analysis identified 37 MdSGHV ORFs homologous with 42 GpSGHV ORFs (Table 2). The differences in ORF homologue numbers were due to five instances where the individual MdSGHV ORFs 025, 033, 036, 074 and 082 had homology to the following pair of GpSGHV ORFs 096/097, 107/108, 110/111, 032/033 and 030/031, respectively. Further BLAST and motif analysis of these pairs showed two instances of ORF duplications followed by mutations. GpSGHV107/108 included ATPase motifs homologous with MdSGHV033, while GpSGHV110/111 had zinc-dependent matrix metalloproteinase motifs similar to the one with MdSGHV036. However, the amino acid identity within the GpSGHV pairs was only 33 %, suggesting that independent changes occurred to differentiate each GpSGHV ORF. The amino acid identity between GpSGHV096 and GpSGHV097 was only 23 % and no conserved motif was found. In two other instances, two GpSGHV ORFs were included within one MdSGHV ORF. GpSGHV032/033 were homologous with two different regions of MdSGHV074. Similarly, GpSGHV030/031 shared homology to different locations in MdSGHV082.

Nine genes with known functions and a gene with unknown function, but with homology to a baculovirus conserved ORF, were present in the two SGHVs. These are DNA polymerase (*dnapol*), thymidylate synthase (*ts*), *per os* infectivity factors 1, 2 and 3 (*pif-1*, *pif-2* and *pif-3*), peroral infectivity factor (*p74*), matrix metalloproteinase (*mmp*), occlusion-derived virus e66 (*odv-e66*), *helicase-2* and *Ac81-like* protein (Table 2). The remainder of the shared ORFs did not have homology to known genes of other large invertebrate dsDNA viruses, which indicated that the SGHVs may have a set of specific genes required for virion morphogenesis for infecting cells of the adult flies' salivary glands and for inducing SGH-specific symptoms that are not caused either by baculovirus or nudivirus infections.

The proteomic analysis of the MdSGHV purified virions revealed the presence of 29 encoded proteins (Garcia-Maruniak *et al.*, 2008), 15 of which share homology with GpSGHV (Table 2), including p74. This indicates that (i) these homologous proteins will most probably be present in the GpSGHV virions and may belong to the SGHV core

Table 1. Biological, structural and molecular characteristics of MdSGHV and GpSGHV

Feature	MdSGHV	GpSGHV	References
Biological characteristic			
Replication site	Salivary glands	Salivary glands, milk glands and ovarioles	Coler *et al.* (1993); Geden *et al.* (2008); Jaenson (1978); Jura *et al.* (1988); Sang *et al.* (1996)
Infection phenotype	Symptomatic	Symptomatic and asymptomatic	V.-U. Lietze (unpublished); Abd-Alla *et al.* (2007)
Symptoms besides SGH	Undeveloped ovaries	Gonadal abnormalities in both sexes	Coler *et al.* (1993); Jura *et al.* (1988); Lietze *et al.* (2007)
Vertical transmission	None	Milk gland, transovarial and transovum	Jura *et al.* (1988, 1989); Lietze *et al.* (2007); Sang *et al.* (1996)
Horizontal transmission	Salivary secretion	Salivary secretion (laboratory colony)	Abd-Alla *et al.* (2007); V.-U. Lietze & D. G. Boucias (unpublished)
Sterilizing agent	Female infertility	Male and female infertility	Lietze *et al.* (2007); Sang *et al.* (1999)
Impact on host behaviour	Mating disruption	Impaired feeding	Lietze *et al.* (2007); Sang *et al.* (1998)
Structural characteristic			
Virion size	65 × 550 nm	50 × 1000 nm	Garcia-Maruniak *et al.* (2008); Abd-Alla *et al.* (2007)
Virion topography	Braided, bead-like surface	Flexible rods	Garcia-Maruniak *et al.* (2008); Abd-Alla *et al.* (2008)
Molecular characteristic			
Size (bp)	124 279	190 032	Abd-Alla *et al.* (2008); Garcia-Maruniak *et al.* (2008)
G+C content (mol%)	44	28	Same as above
Total no. ORFs	108	160	Same as above
No. of *drs* (% of genome)	18 (1.74)	14 (2.84)	Same as above
MdSGHV/GpSGHV shared ORFs	37	42	Present study
No. ORFs homologous to nudiviruses	17	11	Present study
No. ORFs homologous to WSSV	6	4	Present study
No. ORFs homologous to baculoviruses	12	11	Abd-Alla *et al.* (2008); Garcia-Maruniak *et al.* (2008)
No. ORFs homologous to herpesvirus	1	1	Same as above
No. ORFs homologous to entomopoxvirus	1	16	Same as above

Table 2. Comparison of the MdSGHV and GpSGHV ORFs

V, Virion-associated ORFs identified by mass spectroscopy from MdSGHV viral particles (Garcia-Maruniak *et al.*, 2008). Genes in bold are known baculovirus genes conserved in MdSGHV and GpSGHV.

Homologous to MdSGHV					MdSGHV			GpSGHV				Homologous to GpSGHV				
WSSV ORF	**OrNV ORF**	**GbNV ORF**	**HzNV-1 ORF**	**Best match gene/org**	**ORF**	**Length (aa)**	**Gene**	**ORF**	**Length (aa)**	**Identity* vs MdSGHV**	**E-value**	**Best match gene/org**	**HzNV-1 ORF**	**GbNV ORF**	**OrNV ORF**	**WSSV ORF**
514	C17		131	GpSGHV	**1**	977	***dnapol***	**79**	953	31 % (983)	2.00E-119	MdSGHV				
				GpSGHV	**4**	131		**82**	159	26 % (147)	2.00E-06	MdSGHV				
	C7			OrNV	**10**	256	*mcp*									
			7	HzNV-1	**11**	170	*dhfr*									
54	C12		109	OrNV	**12**	294	***ts***	**36**	115	30 % (81)	4.00E-06	WSSV	110		C12	67
				GpSGHV	**V13**	644		**83**	694	27 % (249)	4.00E-20	MdSGHV				
				GpSGHV	**V16**	499		**86**	592	19 % (468)	5.00E-08	MdSGHV				
				GpSGHV	**17**	553		**88**	652	23 % (258)	1.00E-04	MdSGHV				
				GpSGHV	**V22**	343		**93**	329	24 % (319)	9.00E-30	MdSGHV				
				GpSGHV	**V25**	376		**97**	394	33 % (331)	3.00E-49	MdSGHV				
								96	381	22 % (334)	1.00E-06	MdSGHV				
				GpSGHV	**27**	169		**99**	160	31 % (118)	8.00E-08	MdSGHV				
		52	55	GpSGHV	**29**	644	***pif-1***	**102**	652	35 % (653)	2.00E-84	MdSGHV	55	52		
				GpSGHV	**30**	692		**104**	660	24 % (654)	8.00E-47	MdSGHV				
				GpSGHV	**V33**	497		**107**	521	33 % (378)	2.00E-41	MdSGHV				
								108	545	28 % (335)	1.00E-25	MdSGHV				
			70	GpSGHV	**36**	196	***mmp***	**110**	201	27 % (183)	1.00E-14	MdSGHV	71			
								111	219	27 % (162)	1.00E-08	MdSGHV				
								147	176			WSSV				179
								148	275			WSSV				179
			11	GpSGHV	**V39**	707	***p74***	**1**	696	39 % (696)	1.00E-143	MdSGHV	11			
				GpSGHV	**46**	395		**6**	361	28 % (366)	6.00E-32	MdSGHV				
	C6			GpSGHV	**V47**	700	***odv-e66***	**5**	353	35 % (344)	6.00E-46	MdSGHV				
112			69	Grouper iridovirus	**V52**	150	*dUTPase*									
				GpSGHV	**55**	416		**41**	413	26 % (406)	2.00E-15	MdSGHV				
188		63	73	LdMNPV	**62**	343	*rr2*									
172	D8	82	95	SpltMNPV	**65**	785	*rr1*									
			71	Sf ascovirus	**69**	203	*tk*									
				GpSGHV	**70**	967		**40**	901	35 % (929)	9.00E-153	MdSGHV				
				GpSGHV	**V71**	333		**154**	338	32 % (70)	4.00E-04	MdSGHV				
				GpSGHV	**V72**	136		**43**	144	26 % (133)	2.00E-05	MdSGHV				
				GpSGHV	**73**	390		**44**	359	33 % (92)	2.00E-06	MdSGHV				
				GpSGHV	**74**	698		**32**	259	34 % (259)	4.00E-34	MdSGHV				

Table 2. cont.

Homologous to MdSGHV					MdSGHV			GpSGHV				Homologous to GpSGHV				
WSSV ORF	**OrNV ORF**	**GbNV ORF**	**HzNV-1 ORF**	**Best match gene/org**	**ORF**	**Length (aa)**	**Gene**	**ORF**	**Length (aa)**	**Identity* vs MdSGHV**	**E-value**	**Best match gene/org**	**HzNV-1 ORF**	**GbNV ORF**	**OrNV ORF**	**WSSV ORF**
								33	348	35 % (352)	7.00E-50	MdSGHV				
								35	87			ts herpes-virus	110			67
	C3	65	68	OrNV	**76**	407	OrNV ORFC3									
		98	135	MSV entomo-poxvirus	**78**	142	*iap*									
				GpSGHV	**82**	359		**31**	285	29 % (244)	3.00E-20	MdSGHV				
								30	133	33 % (102)	3.00E-07	MdSGHV				
				GpSGHV	**83**	1780		**45**	1728	25 % (1141)	2.00E-89	MdSGHV				
				GpSGHV	**V84**	509		**46**	533	23 % (513)	3.00E-30	MdSGHV				
				GpSGHV	**V86**	381		**50**	291	23 % (203)	7.00E-04	MdSGHV				
				GpSGHV	**87**	544		**51**	660	22 % (525)	2.00E-08	MdSGHV				
	C2	66	123	GpSGHV	**89**	379	***pif-2***	**53**	360	38 % (352)	6.00E-65	MdSGHV	123	66	C1	
				GpSGHV	**V90**	672		**71**	608	24 % (249)	1.00E-09	MdSGHV				
				GpSGHV	**V96**	1473		**65**	1421	19 % (1150)	2.00E-16	MdSGHV				
				GpSGHV	**V97**	482		**64**	595	19 % (579)	3.00E-02	MdSGHV				
				GpSGHV	**100**	434		**61**	494	24 % (377)	2.00E-15	MdSGHV	3			
				GpSGHV	**V102**	257		**72**	269	36 % (90)	3.00E-14	MdSGHV				
				GpSGHV	**104**	770	***helicase-2***	**74**	712	31 % (618)	6.00E-59	MdSGHV		46		
				GpSGHV	**106**	242	***pif-3***	**76**	211	31 % (151)	1.00E-07	LdMNPV				
				GpSGHV	**107**	1125		**77**	1059	21 % (957)	3.00E-31	MdSGHV		12		
	C14			GpSGHV	**108**	205	***Ac81***	**78**	236	36 % (238)	7.00E-30	MdSGHV				

*Per cent of amino acid identity (no. amino acid used in the comparison).

genes, and (ii) both viruses may have similar routes for the early infection steps in their respective hosts. Proteomic analysis of the GpSGHV purified virions will be required to support these assumptions.

Since the nudivirus and nimavirus virions share some structural (large circular dsDNA, dispersed homologous repeat regions), morphological (rod-shaped) and biological (nuclear replication) characteristics with the SGHVs, comparative analyses were carried out between the SGHVs and the nudiviruses HzNV-1, GbNV and OrNV, and the nimavirus WSSV. MdSGHV had more homologous ORFs to HzNV-1, GbNV, OrNV and WSSV (13, 6, 8 and 5, respectively) than GpSGHV (7, 4, 2 and 3, respectively) (Table 2). However, the number of homologous ORFs was very limited and restricted to DNA replication and oral infectivity genes. The protein identity values of the encoded proteins were always higher between the two SGHVs than with the nudiviruses.

Genomic organization

Gene parity plots have been successfully used to align and compare the ORF organization in baculovirus genomes (Hu *et al.*, 1998; van Oers & Vlak, 2007). Using this approach the gene order of both SGHV genomes and each of the SGHVs to HzNV-1 and GbNV (Fig. 1) were compared. The SGHVs revealed two major regions of collinear ORFs that suggest common ancestry of the two SGHVs interspersed by a large gap and a distinct inversion (Fig. 1a). The analysis also showed that the GpSGHV ORFs from 112 to 160 (from a total of 160 ORFs) did not present organizational homology to the MdSGHV ORFs with one exception (GpSGHV154). The ORFs sharing homology to MdSGHV were mostly clustered between ORFs 30 and 111 (Fig. 1a). In contrast, the parity plots of either SGHV versus either nudivirus, HzNV-1 (Fig. 1b and c) or GbNV (Fig. 1d and e) indicated that they have a few ORFs in common and no linear correspondence, thus distancing the SGHVs from the nudiviruses.

Syntenic maps make it possible to compare general genomic organization and visualize regions of both coding and non-coding sequence conservation. The ACT has been a useful tool since 2001 to compare genomes (Cole *et al.*, 2001) and has been applied successfully to test collinearity between baculovirus genomes (Lauzon *et al.*, 2006; Oliveira

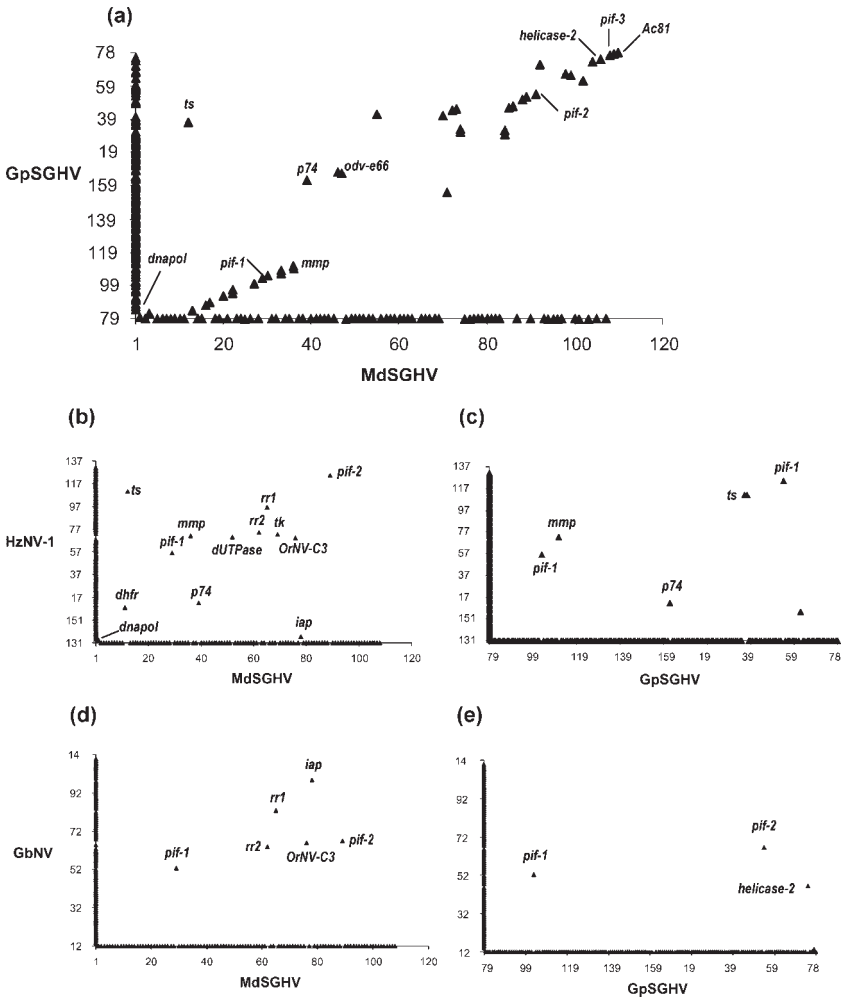

Fig. 1. Gene parity plots comparing gene order of (a) MdSGHV and GpSGHV, (b) MdSGHV and HzNV-1, (c) GpSGHV and HzNV-1, (d) MdSGHV and GbNV, and (e) GpSGHV and GbNV. All the genomes were linearized to start the comparison at the ORF corresponding to DNA polymerase. The non-homologous ORFs of the two SGHVs were located on the *x* and *y* axis, respectively. Known viral genes are indicated in the panels.

et al., 2006; Wolff *et al.*, 2008). The syntenic map between MdSGHV and GpSGHV showed significant collinearity between the two genomes (Fig. 2). A comparable level of collinearity has also been observed when several representatives of the baculovirus NPVs group II were compared (Wolff *et al.*, 2008). A total of 161 conserved regions (not ORFs) sharing amino acid identities of 25–73 % were found between MdSGHV and GpSGHV by the syntenic analysis. When this analysis was extended to include HzNV-1 (Fig. 2a) and GbNV (Fig. 2b), the number of conserved regions was reduced, and there was a lack of collinearity with the SGHVs. The syntenic relationship between these viruses was based on the translated sequence regions that provided an advantage over comparing ORFs, as most of the ORFs of MdSGHV and GpSGHV remain to be validated. These genomic analyses strongly suggest that GpSGHV and MdSGHV are related and distinct from other large invertebrate DNA viruses.

Virus-derived miRNAs have been found to be important regulatory factors involved in the control of both viral (Barth *et al.*, 2008; Gupta *et al.*, 2006) and host gene

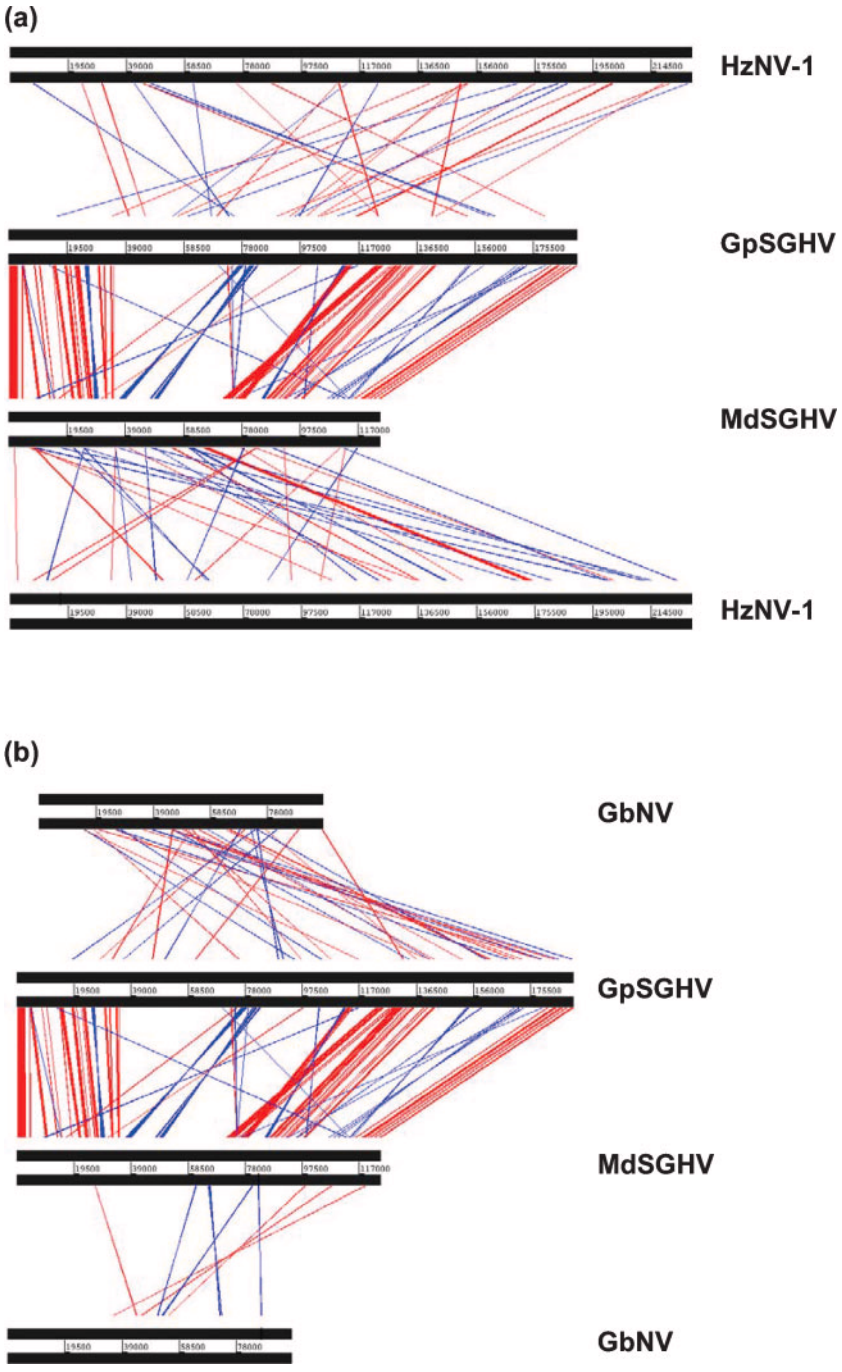

Fig. 2. Syntenic maps comparing the overall level of collinearity of MdSGHV and GpSGHV genomes. HzNV-1 (a) and GbNV (b) were compared to MdSGHV and GpSGHV as well. Blue lines indicate inversions; red lines indicate levels of identity among HSPs. Bands are between conserved regions, not necessarily ORFs. Genome position in base pairs is shown between the lines of each dsDNA viral genome. The number of conserved regions does not reflect the level of homology in each region.

expression (Sullivan *et al.*, 2005; Stern-Ginossar *et al.*, 2007). miRNAs have been found in intergenic regions, introns, non-coding region of exons (Kim & Nam, 2006) and also in the middle of coding regions (Hussain *et al.*, 2008). Six and seven pre-miRNAs were found in MdSGHV and GpSGHV, respectively. These pre-miRNAs were dispersed along the genomes, in both DNA strands, and either inside or in between ORFs (Supplementary Fig. S1 available in JGV Online). This is different from the baculovirus type species AcMNPV in which only four predicted miRNAs were found located in previously identified homologous regions (hr) of the viral genome (data not shown). Currently, we do not have data about the biological function of these putative SGHV miRNAs. However, since both SGHVs cause hypertrophy of infected host salivary gland tissues and both have sequences predicted to code for pre-miRNAs it would be interesting to validate their presence and determine if they regulate gene expression in their respective hosts, the house fly and tsetse fly.

Promoter analysis

Promoter analysis of the SGHV genomes was performed to see if these promoters share common transcriptional properties with other groups of large dsDNA viruses (Marks *et al.*, 2006). Baculoviruses, iridoviruses and SGHVs are rich in A and T nucleotides and more so are their promoters. In MdSGHV, the 4 mer TAAG motif was strongly overrepresented in the 100 and 200 bp sequences upstream of the identified ORFs and could represent a viral gene promoter motif (Table 3, Supplementary Table S1 available in JGV Online). The TAAG motif was also strongly represented in the GpSGHV upstream regions of identified ORFs (Abd-Alla *et al.*, 2008). This TAAG motif is ubiquitously found in promoters of baculovirus late genes (Table 3) and is known to be the canonical sequence to initiate baculovirus late transcripts (Blissard & Rohrmann, 1990). The predominance of this motif in the region upstream of SGHV ORFs might well predict a common role in transcription regulation as with baculoviruses. This TAAG motif was not enriched in promoter regions of CIV ORFs, and this is in line with the phylogenetic distance of iridoviruses from baculoviruses and the SGHV group. The nudivirus HzNV-1 upstream regions were not strongly enriched for TAAG motifs (Table 3), despite its phylogenetically smaller distance to the baculoviruses and SGHV group. From the analysis of 5 mer sequences, a consensus motif [(A/G/T) TAAG (A/G/C)] could be extracted for both SGHV viruses (see also Supplementary Table S1), and this is also reminiscent of the situation in baculoviruses (Marks *et al.*, 2006). Further analyses such as the role of the SGHV ORF motifs in transcriptional regulation are needed to confirm that the TAAG motif is indeed a crucial element of SGHV (late) viral promoters.

Phylogenetic analysis

Both MdSGHV (MdSGHV001) and GpSGHV (SGHV079) encode a putative protein homologous to the delta catalytic subunit of DNA polymerase from several eukaryotic organisms (Abd-Alla *et al.*, 2008; Garcia-Maruniak *et al.*, 2008). The amino acid identity between MdSGHV and GpSGHV DNA polymerase was 31 % (Table 2). The best match for both ORFs to viral genes was to herpesvirus DNA polymerase with 22.6 and 24.9 % amino acid identity for MdSGHV001 and GpSGHV079, respectively. The phylogenetic analysis clearly showed that the DNA polymerases from MdSGHV and GpSGHV are in the same clade, and they branched distantly from the baculoviruses and nudiviruses (data not shown). These findings agreed with the topologies generated for the GpSGHV and MdSGHV DNA polymerases by Abd-Alla *et al.* (2008) and Garcia-Maruniak *et al.* (2008), respectively.

The genes encoding *per os* infectivity factors (pifs) and p74 structural proteins are common to all sequenced nudiviruses and baculoviruses, and they are essential for oral infectivity (Kuzio *et al.*, 1989; Kikhno *et al.*, 2002; Pijlman *et al.*, 2003; Ohkawa *et al.*, 2005; Song *et al.*, 2008). These genes are known or presumed to encode envelope proteins typical of the occlusion-derived virus phenotype, which are important for the entry process into the midgut epithelial cells (Slack & Arif, 2007). Both MdSGHV and GpSGHV encode putative proteins homologous with the baculovirus pif proteins and the *odv-e66* gene conserved in Lepidopteran baculoviruses (Abd-Alla *et al.*, 2008; Garcia-Maruniak *et al.*, 2008). In both viruses a homologue of the conserved baculovirus gene *odv-e66* was found (GpSGHV005 and MdSGHV047), which encodes a structural component of the occlusion-derived virion of baculoviruses. MdSGHV contains a homologue of *Ac150*, which encodes a baculovirus virion protein that may modulate oral infection but is not essential for it to occur (Zhang *et al.*, 2005). No such gene homologue was found in GpSGHV ORFs. The phylogenetic analysis of a concatenated p74, pif-1, pif-2 and pif-3 amino acid sequence (Fig. 3a) and of odv-e66 (Fig. 3b) showed that the relationship between the structural proteins in MdSGHV and GpSGHV was very close, and in all cases, both viruses shared the same clade having amino acid identities of 39, 35, 38, 31 and 35 % for p74, pif-1, pif-2, pif-3 and odv-e66, respectively (Table 2). This tree was strongly supported by bootstrap analysis.

A homologue of AcMNPV ORF 81 (Ac81), present in all the baculoviruses and nudiviruses, was also found in MdSGHV108 and GpSGHV078. Both proteins had amino acid identities of 36 % and branched very close to each other in the same clade (Fig. 3c). Although this gene belongs to the conserved core genes of baculoviruses, its function has not been defined (Slack & Arif, 2007).

Helicase is one of the 29 conserved baculovirus core genes. ORFs homologous to helicase-2 have been described in HzNV-1, HzNV-2, OrNV (Wang *et al.*, 2007a), GpSGHV (ORF 74) and MdSGHV (ORF 104) (Abd-Alla *et al.*, 2008; Garcia-Maruniak *et al.*, 2008). Although *helicase-2* is one of the conserved nudivirus genes (Wang *et al.*, 2007a), we did

Table 3. Relative enrichment of 4 and 5 nt motifs in 100 bp upstream regions of MdSGHV ORFs, in comparison to GpSGHV, HzNV-1 and CIV

AcMNPV is used as a reference. Only the 10 motifs with highest relative enrichment are shown for each virus. Sequences in bold are significantly enriched. Motifs containing TAAG are underlined. See also Supplementary Table S1.

Virus	4 mer motif				5 mer motif			
	Motif	Occurrence in genome (% of expected occurrence)	Occurrence in upstream regions (% of expected occurrence)	Relative enrichment in upstream regions	Motif	Occurrence in genome (% of expected occurrence)	Occurrence in upstream regions (% of expected occurrence)	Relative enrichment in upstream regions
AcMNPV	**TAAG**	393 (29)	90 (114)	4.0	**ATAAG**	127 (32)	45 (190)	6.0
	TATA	1314 (66)	172 (149)	2.3	**TAAGG**	49 (17)	16 (96)	5.5
	ATAA	1973 (101)	222 (198)	1.9	**TAAGA**	110 (28)	29 (123)	4.4
	ATAT	1616 (81)	170 (147)	1.8	**TAAGT**	104 (25)	25 (103)	4.1
	AGTA	671 (49)	70 (89)	1.8	**GTAAG**	85 (30)	18 (108)	3.6
	AAGG	473 (51)	41 (76)	1.5	**TTAAG**	126 (31)	24 (99)	3.2
	GATA	867 (63)	74 (94)	1.5	**ATATA**	420 (72)	74 (214)	3.0
	CACT	612 (64)	52 (95)	1.5	**CAGTA**	144 (52)	23 (139)	2.7
	AATA	2230 (115)	186 (166)	1.4	**TATAT**	420 (70)	67 (189)	2.7
	ATTA	1957 (98)	163 (141)	1.4	**TATAA**	488 (84)	76 (220)	2.6
MdSGHV	**TAAG**	317 (26)	67 (86)	4.7	**TAAGG**	37 (14)	13 (110)	7.8
	CTTA	317 (26)	32 (223)	2.3	**ATAAG**	133 (39)	37 (241)	6.2
	ATAA	317 (26)	121 (140)	2.3	**TAAGA**	113 (33)	31 (202)	6.1
	CCCC	317 (26)	33 (33)	2.2	**GTAAG**	65 (25)	14 (119)	4.8
	CTAG	317 (26)	16 (33)	2.1	**TAAGC**	52 (20)	10 (85)	4.3
	TATA	317 (26)	112 (43)	2.0	**TTAAG**	73 (22)	13 (86)	4.0
	TAAA	317 (26)	90 (127)	2.0	**TATAA**	264 (60)	44 (223)	3.7
	AAGC	317 (26)	33 (169)	2.0	**CTTAT**	133 (40)	21 (140)	3.5
	CTAA	317 (26)	35 (92)	1.9	**CTAGG**	22 (11)	3 (33)	3.0
	AGAG	317 (26)	43 (251)	1.9	**CCTAG**	22 (11)	3 (33)	3.0
GpSGHV	**TAAG**	848 (32)	106 (96)	3.0	**ATAAG**	338 (35)	70 (172)	4.9
	AGTC	506 (53)	36 (88)	1.7	**TAAGA**	297 (31)	46 (113)	3.7
	AGGT	536 (52)	38 (87)	1.7	**GTCAG**	80 (57)	11 (187)	3.2
	TAGG	458 (44)	32 (73)	1.6	**GTAAG**	128 (34)	16 (100)	3.0
	AGTA	1515 (58)	103 (93)	1.6	**AGGGC**	57 (100)	7 (289)	2.9
	CAGT	814 (85)	55 (135)	1.6	**TCCGC**	52 (109)	6 (297)	2.7
	GCGC	178 (123)	12 (195)	1.6	**CCTTA**	83 (26)	9 (67)	2.6
	AAGT	1703 (65)	113 (102)	1.6	**CGCGC**	28 (143)	3 (362)	2.5
	TAGT	1451 (58)	95 (89)	1.5	**GCGCA**	79 (148)	8 (354)	2.4
	GTAG	676 (66)	44 (101)	1.5	**TAAGC**	99 (28)	10 (67)	2.4
Hz1-NV	**TATT**	2.764 (85)	187 (169)	2.0	**ACTTA**	173 (26)	16 (69)	2.7
	TAAC	1171 (50)	75 (95)	1.9	**TTTTT**	986 (103)	89 (273)	2.6
	AATA	2764 (86)	176 (161)	1.9	**TAATT**	521 (55)	47 (147)	2.6
	TTTT	3541 (108)	225 (202)	1.9	**CTTAG**	114 (23)	10 (60)	2.6
	ATAC	2340 (101)	147 (186)	1.8	**TTATT**	794 (84)	67 (207)	2.5
	TTAT	2279 (70)	141 (127)	1.8	**ATACC**	428 (88)	36 (216)	2.5
	GTAT	2340 (101)	141 (179)	1.8	**ATAGG**	265 (56)	22 (136)	2.4
	ATAA	2279 (71)	132 (121)	1.7	**TGATA**	400 (60)	33 (145)	2.4
	TATA	2728 (85)	158 (144)	1.7	**GATAA**	305 (46)	25 (111)	2.4
	TTTA	4102 (126)	237 (214)	1.7	**CGTTA**	212 (44)	17 (103)	2.4
CIV	**TAAA**	7280 (115)	735 (219)	1.9	**ATAAA**	2192 (98)	275 (230)	2.4
	ATAA	5120 (81)	494 (147)	1.8	**AATAA**	2083 (93)	252 (211)	2.3
	TTAA	6960 (107)	614 (177)	1.7	**TAGGG**	42 (34)	5 (75)	2.2
	AATA	5438 (86)	450 (134)	1.6	**GACGC**	36 (69)	4 (144)	2.1
	TAAG	1247 (52)	101 (79)	1.5	**GCGAG**	18 (37)	2 (76)	2.1
	AAAT	7966 (126)	636 (189)	1.5	**GGCCC**	18 (84)	2 (177)	2.1
	TTTA	7280 (108)	566 (158)	1.5	**GGGCC**	18 (90)	2 (188)	2.1
	AAAA	10516 (171)	788 (241)	1.4	**TAAAT**	2272 (98)	236 (192)	2.0
	TAGT	1153 (46)	85 (64)	1.4	**ATAAG**	475 (55)	49 (108)	1.9
	TATA	3830 (59)	282 (81)	1.4	**TAAAG**	1019 (119)	105 (231)	1.9

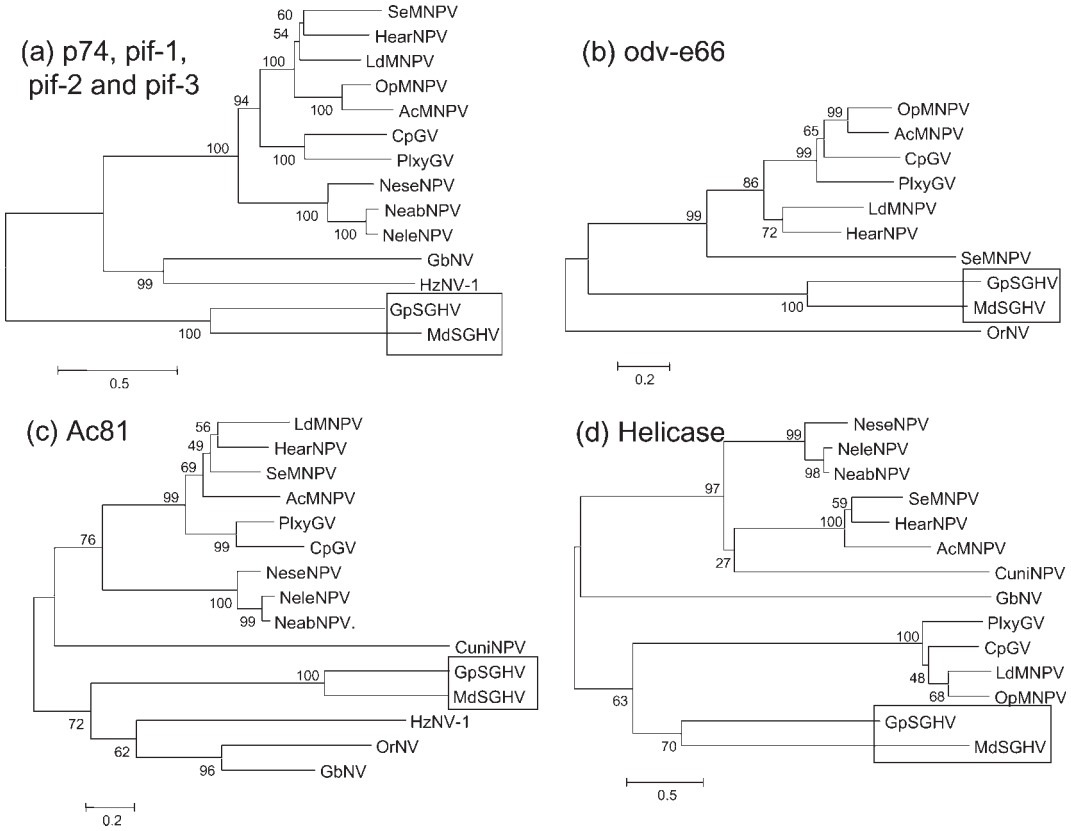

Fig. 3. Neighbour-joining phylogenetic trees of (a) combined p74, pif-1, pif-2 and pif-3, (b) odv-e66, (c) Ac81 and (d) helicase proteins and their homologues in baculoviruses and nudiviruses. The following viruses (GenBank accession number in square brackets) were included: Lepidopteran-specific NPVs: *Autographa californica* (Ac) MNPV [NC_001623], *Orgyia pseudotsugata* (Op) MNPV [NC_001875], *Lymantria dispar* (Ld) MNPV [NC_001973], *Spodoptera exigua* (Se) MNPV [NC_002169], *Helicoverpa armigera* (Hear) NPV [NC_002654]; Granuloviruses: *Cydia pomonella* (Cp) GV [NC_002816], *Plutella xylostella* (Plxy) GV [NC_002593]; Hymenopteran-specific NPVs: *Neodiprion sertifer* (Nese) NPV [NC_005905], *N. lecontei* (Nele) NPV [NC_005906], *N. abietis* (Neab) NPV [NC_008252]; Dipteran-specific NPV: *Culex nigripalpus* (Cuni) NPV [NC_003084]; and Nudiviruses: *Heliothis zea* (Hz) NV-1 [NC_004156], *Gryllus bimaculatus* (Gb) NV [NC_009240], *Oryctes rhinoceros* (Or) NV [DQ665871, DQ665870). Distances were calculated using Poisson correction. Homogeneous substitution pattern among lineages with gamma distributed rate among sites (gamma parameter 2.25) was employed for reconstruction of the trees. The robustness of the tree was tested using bootstrap analysis (500 replicates). Numbers on the nodes indicate bootstrap values.

not find homology to the SGHVs helicases. The SGHVs putative helicase-2 shared 31 % amino acid identities and clustered in a single clade (Fig. 3d), separated from the conserved *helicase* core gene of baculoviruses. The biological activity of the SGHV helicase-2 has to still be investigated to confirm if it is involved in DNA repair and recombination as was found for baculoviruses (Pearson & Rohrmann, 1998).

Although the SGHVs have some structural (non-occluded virion, circular dsDNA) and biological (oral infectivity, nuclear replication) similarities with baculoviruses and nudiviruses they have major differences especially in their pathology. None of the nudiviruses or baculoviruses induce SGH that cause a decrease (or even inhibition of) of fertility. The lack of gene order conservation between the SGHVs and nudiviruses was demonstrated when either their ORFs or their syntenic maps were compared to one another. Furthermore, the distance obtained in the phylogenetic trees indicated that the SGHVs, although related to each other, are not close to other large dsDNA viruses infectious to insects. Future studies on MdSGHV and GpSGHV could reveal what is the genetic basis for adaptation to hosts (house and tsetse flies) with different biological backgrounds. Information obtained from additional SGHVs should provide further insight into the genetics and unique pathology of this new group of viruses and a framework for addressing the taxonomic position of the SGHVs.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support provided by the National Institute of Health (NIAID R21 A1073501-01), USDA/NRI

(2007-35302-18127) and the research agreement no. 14263 with Joint FAO/IAEA Division IAEA.

REFERENCES

Abd-Alla, A., Bossin, H., Cousserans, F., Parker, A., Bergoin, M. & Robinson, A. (2007). Development of a non-destructive PCR method for detection of the salivary gland hypertrophy virus (SGHV) in tsetse flies. *J Virol Methods* **139**, 143–149.

Abd-Alla, A. M. M., Cousserans, F., Parker, A. G., Jehle, J. A., Parker, N. J., Vlak, J. M., Robinson, A. S. & Bergoin, M. (2008). Genome analysis of a *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) reveals a novel large double-stranded circular DNA virus. *J Virol* **82**, 4595–4611.

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

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Amargier, A., Lyon, J. P., Vago, C., Meynadier, G. & Veyrunes, J. C. (1979). Mise en evidence et purification d'un virus dans la proliferation monstreuse glandulaire d'insectes. Etude sur *Merodon equestris* (Diptera, Syrphidae). Note. *C R Seances Acad Sci D* **289**, 481–484 (in French).

Barth, S., Pfuhl, T., Mamiani, A., Ehses, C., Roemer, K., Kremmer, E., Jäker, C., Höck, J., Meister, G. & Grässer, F. A. (2008). Epstein–Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. *Nucleic Acids Res* **36**, 666–675.

Blissard, G. W. & Rohrmann, G. F. (1990). Baculovirus diversity and molecular biology. *Annu Rev Entomol* **35**, 127–155.

Cheng, C.-H., Liu, S.-M., Chow, T.-Y., Hsiao, Y.-Y., Wang, D.-P., Huang, J.-J. & Chen, H.-H. (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.

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honoré, N., Garnier, T., Churcher, C. & other authors (2001). Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007–1011.

Coler, R. R., Boucias, D. G., Frank, J. H., Maruniak, J. E., Garcia-Canedo, A. & Pendland, J. C. (1993). Characterization and description of a virus causing salivary gland hyperplasia in the housefly, *Musca domestica*. *Med Vet Entomol* **7**, 275–282.

Ellis, D. S. & Maudlin, I. (1987). Salivary gland hyperplasia in wild caught tsetse from Zimbabwe. *Entomol Exp Appl* **45**, 167–173.

Garcia-Maruniak, A., Maruniak, J. E., Farmerie, W. & Boucias, D. G. (2008). Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of *Musca domestica*, MdSGHV. *Virology* **377**, 184–196.

Geden, C. J., Lietze, V.-U. & Boucias, D. (2008). Seasonal prevalence and transmission of salivary gland hypertrophy virus of house flies (Diptera: Muscidae). *J Med Entomol* **45**, 42–51.

Gouteux, J. P. (1987). Prevalence of enlarged salivary glands in *Glossina palpalis*, *G. pallicera*, and *G. nigrofusca* (Diptera: Glossinidae) from the Vavoua area, Ivory Coast. *J Med Entomol* **24**, 268.

Grad, Y., Aach, J., Hayes, G. D., Reinhart, B. J., Church, G. M., Ruvkun, G. & Kim, J. (2003). Computational and experimental identification of *C. elegans* microRNAs. *Mol Cell* **11**, 1253–1263.

Gupta, A., Gartner, J. J., Sethupathy, P., Hatzigeorgiou, A. G. & Fraser, N. W. (2006). Anti-apoptotic function of a microRNA encoded by the HSV-1 latency-associated transcript. *Nature* **442**, 82–85.

Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res* **31**, 3429–3431.

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

Hussain, M., Taft, R. J. & Asgari, S. (2008). An insect virus-encoded microRNA regulates viral replication. *J Virol* **82**, 9164–9170.

Jaenson, T. G. T. (1978). Virus-like rods associated with salivary gland hyperplasia in tsetse, *Glossina pallidipes*. *Trans R Soc Trop Med Hyg* **72**, 234–238.

Jura, W. G. Z. O., Odhiambo, T. R., Otieno, L. H. & Tabu, N. O. (1988). Gonadal lesions in virus-infected male and female tsetse, *Glossina pallidipes* (Diptera: Glossinidae). *J Invertebr Pathol* **52**, 1–8.

Jura, W. G. Z. O., Otieno, L. H. & Chimtawi, M. M. B. (1989). Ultrastructural evidence for trans-ovum transmission of the DNA virus of tsetse *Glossina pallidipes* (Diptera: Glossinidae). *Curr Microbiol* **18**, 1–4.

Jura, W. G. Z. O., Zdarek, J. & Otieno, L. H. (1993). A simple method for artificial infection of tsetse, *Glossina morsitans morsitans* larvae with the DNA virus of *G. pallidipes*. *Insect Sci Appl* **14**, 383–387.

Kikhno, I., Gutierrez, Z., Croizier, L., Croizier, G. & Lopez Ferber, M. (2002). Characterization of *pif*, a gene required for the *per os* infectivity of *Spodoptera littoralis* nucleopolyhedrovirus. *J Gen Virol* **83**, 3013–3022.

Kim, V. N. & Nam, J. W. (2006). Genomics of microRNA. *Trends Genet* **22**, 165–173.

Kokwaro, E. D., Nyindo, M. & Chimtawi, M. (1990). Ultrastructural changes in salivary glands of tsetse, *Glossina morsitans morsitans*, infected with virus and rickettsia-like organisms. *J Invertebr Pathol* **56**, 337–346.

Kuzio, J., Jaques, R. & Faulkner, P. (1989). Identification of p74, a gene essential for virulence of baculovirus occlusion bodies. *Virology* **173**, 759–763.

Lauzon, H. A., Garcia-Maruniak, A., Zanotto, P. M., Clemente, J. C., Herniou, E. A., Lucarotti, C. J., Arif, B. M. & Maruniak, J. E. (2006). Genomic comparison of *Neodiprion sertifer* and *Neodiprion lecontei* nucleopolyhedroviruses and identification of potential hymenopteran baculovirus-specific open reading frames. *J Gen Virol* **87**, 1477–1489.

Lietze, V.-U., Geden, C. J., Blackburn, P. & Boucias, D. G. (2007). Effects of salivary gland hypertrophy virus on the reproductive behavior of the house fly, *Musca domestica*. *Appl Environ Microbiol* **73**, 6811–6818.

Marks, H., Ren, X. Y., Sandbrink, H., van Hulten, M. C. & Vlak, J. M. (2006). In silico identification of putative promoter motifs of White Spot Syndrome Virus. *BMC Bioinformatics* **7**, 309.

Minter-Goedbloed, E. & Minter, D. M. (1989). Salivary gland hyperplasia and trypanosome infection of *Glossina* in two areas of Kenya. *Trans R Soc Trop Med Hyg* **83**, 640–641.

Odindo, M. O. (1982). Incidence of salivary gland hypertrophy in field populations of the tsetse *Glossina pallidipes* on the south Kenyan coast. *Insect Sci Appl* **3**, 59–64.

Odindo, M. O., Sabwa, D. M., Amutalla, P. A. & Otieno, W. A. (1981). Preliminary tests on the transmission of virus-like particles to the tsetse *Glossina pallidipes*. *Insect Sci Appl* **2**, 219–221.

Ohkawa, T., Washburn, J. O., Sitapara, R., Sid, E. & Volkman, L. E. (2005). Specific binding of *Autographa californica* M nucleopolyhedrovirus occlusion-derived virus to midgut cells of *Heliothis virescens*

larvae is mediated by products of *pif* genes *Ac119* and *Ac022* but not by *Ac115*. *J Virol* **79**, 15258–15264.

Oliveira, J. V. de C., Wolff, J. L. C., Garcia-Maruniak, A., Ribeiro, B. M., de Castro, M. E. B., de Souza, M. L., Moscardi, F., Maruniak, J. E. & Zanotto, P. M. de A. (2006). Genome of the most widely used viral biopesticide: *Anticarsia gemmatalis* multiple nucleopolyhedrovirus. *J Gen Virol* **87**, 3233–3250.

Otieno, L. H., Kokwaro, E. D., Chimtawi, M. & Onyango, P. (1980). Prevalence of enlarged salivary glands in wild populations of *Glossina pallidipes* in Kenya, with a note on the ultrastructure of the affected organ. *J Invertebr Pathol* **36**, 113–118.

Pearson, M. N. & Rohrmann, G. F. (1998). Characterization of a baculovirus-encoded ATP-dependent DNA ligase. *J Virol* **72**, 9142–9149.

Pijlman, G. P., Pruijssers, A. J. & Vlak, J. M. (2003). Identification of *pif-2*, a third conserved baculovirus gene required for *per os* infection of insects. *J Gen Virol* **84**, 2041–2049.

Ritchie, W., Legendre, M. & Gautheret, D. (2007). RNA stem-loops: to be or not to be cleaved by RNAse III. *RNA* **13**, 457–462.

Sang, R. C., Jura, W. G. Z. O., Otieno, L. H. & Ogaja, P. (1996). Ultrastructural changes in the milk gland of tsetse *Glossina morsitans centralis* (Diptera; Glissinidae) female infected by a DNA virus. *J Invertebr Pathol* **68**, 253–259.

Sang, R. C., Jura, W. G. Z. O., Otieno, L. H. & Mwangi, R. W. (1998). The effects of a DNA virus infection on the reproductive potential of female tsetse flies, *Glossina morsitans centralis* and *Glossina morsitans morsitans* (Diptera: Glossinidae). *Mem Inst Oswaldo Cruz* **93**, 861–864.

Sang, R. C., Jura, W. G. Z. O., Otieno, L. H., Mwangi, R. W. & Ogaja, P. (1999). The effects of a tsetse DNA virus infection on the functions of the male accessory reproductive gland in the host fly *Glossina morsitans centralis* (Diptera; Glossinidae). *Curr Microbiol* **38**, 349–354.

Shaw, M. K. & Moloo, S. K. (1993). Virus-like particles in Rickettsia within the midgut epithelial cells of *Glossina morsitans centralis* and *Glossina brevipalpis*. *J Invertebr Pathol* **61**, 162–166.

Slack, J. & Arif, B. M. (2007). The baculoviruses occlusion-derived virus: virion structure and function. *Adv Virus Res* **69**, 99–165.

Song, J., Wang, R., Deng, F., Wang, H. & Hu, Z. H. (2008). Functional studies of *per os* infectivity factors of *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus. *J Gen Virol* **89**, 2331–2338.

Stern-Ginossar, N., Elefant, N., Zimmermann, A., Wolf, D. G., Saleh, N., Biton, M., Horwitz, E., Prokocimer, Z., Prichard, M. & other authors (2007). Host immune system gene targeting by a viral miRNA. *Science* **317**, 376–381.

Sullivan, C. S., Grundhoff, A. T., Tevethia, S., Pipas, J. M. & Ganem, D. (2005). SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* **435**, 682–686.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.

van Hulten, M. C. W., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R. K. & Vlak, J. M. (2001). The white spot syndrome virus DNA genome sequence. *Virology* **286**, 7–22.

van Oers, M. M. & Vlak, J. M. (2007). Baculovirus genomics. *Curr Drug Targets* **8**, 1051–1068.

Wang, Y., Burand, J. P. & Jehle, J. A. (2007a). Nudivirus genomics: diversity and classification. *Virol Sin* **22**, 128–136.

Wang, Y., Kleespies, R. G., Huger, A. M. & Jehle, J. A. (2007b). The genome of *Gryllus bimaculatus* nudivirus indicates an ancient diversification of baculovirus-related nonoccluded nudiviruses of insects. *J Virol* **81**, 5395–5406.

Wang, Y., van Oers, M. M., Crawford, A. M., Vlak, J. M. & Jehle, J. A. (2007c). Genomic analysis of *Oryctes rhinoceros* virus reveals genetic relatedness to *Heliothis zea* virus 1. *Arch Virol* **152**, 519–531.

Wolff, J. L., Valicente, F. H., Martins, R., Oliveira, J. V. & Zanotto, P. M. (2008). Analysis of the genome of *Spodoptera frugiperda* nucleopolyhedrovirus (SfMNPV-19) and of the high genomic heterogeneity in group II nucleopolyhedroviruses. *J Gen Virol* **89**, 1202–1211.

Zhang, J. H., Ohkawa, T., Washburn, J. O. & Volkman, L. E. (2005). Effect of Ac150 on virulence and pathogenesis of *Autographa californica multiple nucleopolyhedrovirus* in noctuid hosts. *J Gen Virol* **86**, 1619–1627.