Virology 377 (2008) 184–196

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of *Musca domestica*, MdSGHV

Alejandra Garcia-Maruniak [a,*], James E. Maruniak [a], William Farmerie [b], Drion G. Boucias [a]

[a] Department of Entomology and Nematology, PO Box 110620, University of Florida, Gainesville, Florida 32611-0620, USA
[b] Interdisciplinary Center for Biotechnology Research, Genomics Core, University of Florida, Gainesville, Florida 32610-3622, USA

ARTICLE INFO

Article history:
Received 27 February 2008
Returned to author for revision
25 March 2008
Accepted 14 April 2008
Available online 21 May 2008

Keywords:
Salivary gland hypertrophy virus
Insect viruses
Insect salivary gland
Musca domestica
SGHV
Genome sequence

ABSTRACT

The genome of the virus that causes salivary gland hypertrophy in *Musca domestica* (MdSGHV) was sequenced. This non-classified, enveloped, double stranded, circular DNA virus had a 124,279bp genome. The G + C content was 43.5% with 108 putative methionine-initiated open reading frames (ORFs). Thirty ORFs had homology to database proteins: eleven to proteins coded by both baculoviruses and nudiviruses (p74, pif-1, pif-2, pif-3, odv-e66, rr1, rr2, iap, dUTPase, MMP, and Ac81-like), seven to nudiviruses (mcp, dhfr, ts, tk and three unknown proteins), one to baculovirus (Ac150-like), one to herpesvirus (dna pol), and ten to cellular proteins. Mass spectrum analysis of the viral particles' protein components identified 29 structural ORFs, with only p74 and odv-e66 previously characterized as baculovirus structural proteins. Although most of the homology observed was to nudiviruses, phylogenetic analysis showed that MdSGHV was not closely related to them or to the baculoviruses.

Introduction

Presently, three different dipterans are known to harbor salivary gland hypertrophy viruses (SGHV). These include the house fly *Musca domestica* (Coler et al., 1993), the narcissus bulb fly *Merodon equestris* (Amargier et al., 1979), and various tsetse fly species *Glossina* spp. (Gouteux, 1987; Jaenson, 1978; Minter-Goedbloed and Minter 1989; Otieno et al., 1980; Shaw and Moloo, 1993). Comparisons among different virus-host fly interactions have demonstrated that the different viruses share several morphological and pathological properties (Amargier et al., 1979; Coler et al., 1993; Ellis and Maudlin, 1987; Jaenson, 1978; Kokwaro et al., 1990, 1991; Odindo et al., 1986). Electron microscopy of virus particles in hypertrophic salivary glands showed heterogeneous, elongate (650 by 75 nm), enveloped rod-shaped virions (Geden et al., 2008). In all three fly hosts, the virus replicated in the nuclei of salivary gland cells, resulting in nuclear and salivary gland hypertrophy (SGH) that is easily visualized when the insects are dissected. Infected adults do not exhibit any external disease symptoms, but viral infection is known to inhibit reproduction (Lietze et al., 2007).

The best-studied SGHVs are those associated with the tsetse fly complex that were detected over sixty years ago in feral flies (Burtt, 1945; Whitnall, 1934). SGHV infection in tsetse flies, in addition to causing SGH, results in testicular degeneration and ovarian abnormalities (Feldmann et al., 1992; Jura et al., 1988; Sang et al., 1998, 1999). The tsetse SGHV can infect the female milk gland, resulting in maternal transmission to larvae in this viviparous, hematophagous insect. Venereal transmission between infected and healthy tsetse flies has been reported (Jura et al., 1989). The natural incidence of the tsetse SGHV is low; only 0.4 to 5% of the field-collected flies display SGH (Ellis and Maudlin, 1987; Jaenson, 1978; Jura et al., 1988; Odindo, 1982; Odindo et al., 1981). In colonized *G. morsitans*, incidence of symptomatic SGH ranges from 1.1% (Jura et al., 1993) to 4.0% (Kokwaro et al., 1990). However, in recent years, the tsetse SGHV has been reported to infect and sometimes decimate the tsetse fly colonies being used for sterile insect release programs. Recently, the development of a diagnostic polymerase chain reaction (PCR) has demonstrated that the tsetse SGHV can exist in an asymptomatic state in host flies (Abd-Alla et al., 2007). Insect colonies displaying low levels (< 5%) of SGH symptoms produce a 100% tsetse SGHV PCR-positives.

The SGHV from *M. domestica* (MdSGHV) is orally transmitted by adult feeding on contaminated substrates and is capable of inducing the symptomatic SGH within days. Like the tsetse SGHV, infection by MdSGHV acts to rapidly and completely inhibit egg production in viremic females (Lietze et al., 2007). Examination of the translational and transcriptional products has demonstrated that the infection blocks hexamerin and yolk protein synthesis. In addition, infection with MdSGHV disrupts house fly mating behavior in that symptomatic

* Corresponding author. Fax: +1 352 392 0190.
E-mail addresses: alemar@ufl.edu (A. Garcia-Maruniak), jem@ifas.ufl.edu (J.E. Maruniak), wgf@biotech.ufl.edu (W. Farmerie), pathos@ufl.edu (D.G. Boucias).

0042-6822/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2008.04.010

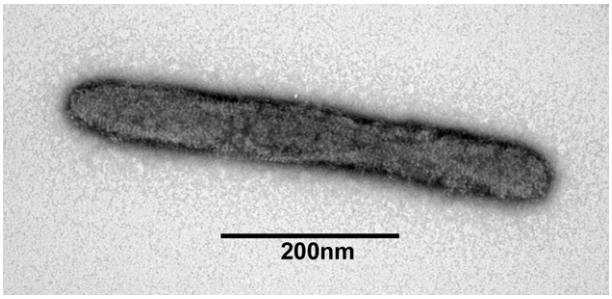


Fig. 1. Transmission electron micrograph of MdSGHV particle purified on Nycodenz gradients and stained with uranyl acetate.

females refuse to copulate with healthy males and symptomatic males show reduce avidity to initiate courtship with healthy females. Thus, the MdSGHV-induced mechanisms underlying the inhibition of house fly reproduction appear to function on both physiological and behavioral levels (Lietze et al., 2007).

Comparisons reveal that MdSGHV possesses biological properties that are distinct from the tsetse SGHV. The house fly is highly susceptible to the virus. Injection of an estimated 100 virions results in 100% SGH of test flies (Lietze et al., 2007), whereas injection with millions of virions results in ~ 20% SGH in colony-reared adult tsetse flies (Abd-Alla, personal communication). The SGHVs in tsetse flies have been reported to be transmitted vertically from mother to offspring, whereas the MdSGHV has been shown to be only orally transmitted (Lietze et al., 2007). The incidence of symptomatic SGH, reported to be as high as 34% in feral house fly populations (Geden et al., 2008), is much greater than the 0.4 to 5.0% in tsetse fly populations (Jaenson, 1978; Otieno et al., 1980; Odindo et al., 1981; Odindo, 1982; Ellis and Maudlin, 1987; Jura et al., 1988).

The MdSGHV, as well as the tsetse SGHV, share general characteristics with the non-occluded insect nudiviruses, such as being insect-pathogenic, having an enveloped, rod-shaped morphology, and possessing a circular dsDNA genome (Wang et al., 2007a). Members of this group were once placed within subgroup C of the Baculoviridae but have since been placed in the unassigned nudivirus group and classified as *sedis incertae*. Analyses of complete genome sequences of over 40 baculoviruses have shown the presence of a core of 29 conserved genes (Garcia-Maruniak et al., 2004; Lauzon et al., 2004). The *Heliothis zea* nudiviruses 1 and 2 (HzNV-1, HzNV-2), the *Gryllus bimaculatus* nudivirus (GbNV), and the *Oryctes rhinoceros* nudivirus (OrNV) have been shown to contain 15 core baculovirus genes (Cheng et al., 2002; Wang et al., 2007a,b,c). Analysis of selected nudivirus sequences has indicated that they form a monophyletic sister group that apparently diverged from a common ancestor of the baculoviruses (Wang et al., 2007b). Whether the SGHV fits within this monophyletic nudivirus group is unknown. Preliminary studies on both the tsetse fly and housefly SGHVs have demonstrated that they possess structural and biological properties that are distinct from the insect nudiviruses (Coler et al., 1993; Abd-Alla et al., 2007). During submission of this manuscript, the complete genome sequence of the related GpSGHV became available (Abd-Alla et al., 2008). In order to determine the genetic basis for the SGHV group classification and to further understand the MdSGHV, its genome has been sequenced. The results of this effort are presented in this manuscript.

Results and discussion

Virus morphology and SDS-PAGE analysis

Injection of the filtered homogenate from MdSGHV infected salivary glands into healthy houseflies resulted in 100% infection. Gradient purification on osmotically stabililized Nycodenz gradients produced a uniform population of enveloped rod-shaped virus particles measuring 65 by 550 nm (Fig. 1). Previous attempts involving sucrose gradient purification resulted in preparations consisting of partly enveloped nucleocapsids (Coler et al., 1993). Negative staining of the Nycodenz-purified virus revealed enveloped particles possessing a unique braided, bead-like surface topography not observed with other rod-shaped insect viruses (Fig. 1). Furthermore, the ends of the virus particle were rounded and not blunt-ended as observed with the nudiviruses. The length of the enveloped MdSGHV, although shorter than the *Glossina pallipedes* GpSGHV (1.0 μm in length) (Odindo et al., 1986), was significantly longer than the enveloped forms of GbNV and OrNV (100 by 200 nm, Wang et al., 2007a).

SDS-PAGE of purified MdSGHV produced a complex of Coomassie R-250 blue-stained bands ranging from 10 to >150 kDa (Fig. 2). The presence and distribution of the two major-stained (bands b and h) and moderately-stained bands agreed with the protein profile previously reported (Coler et al., 1993) and those results obtained

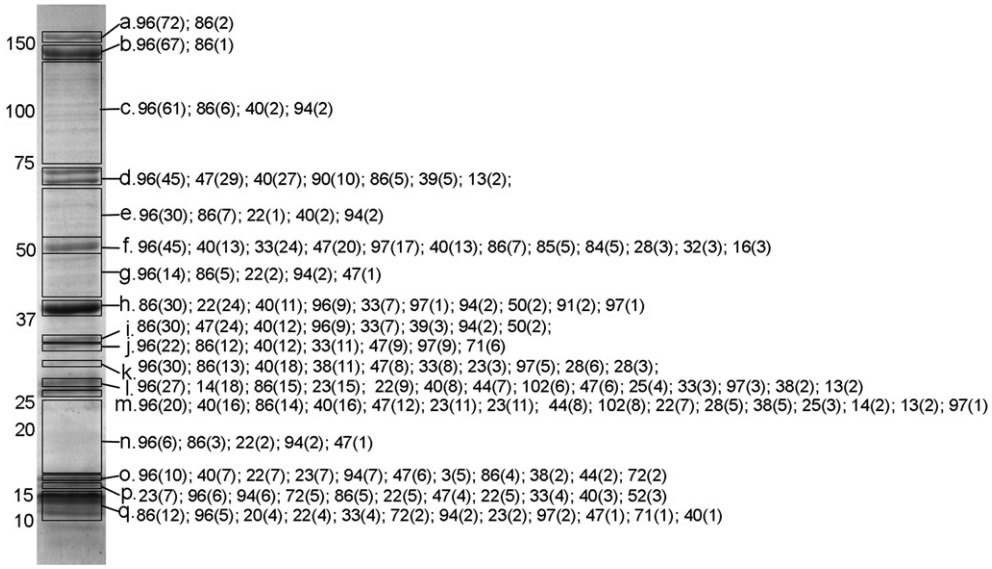


Fig. 2. SDS-PAGE of MdSGHV stained with Coomassie blue. Molecular masses (kDa) of the Bio-Rad markers are denoted on the left side of the gel. The letters represent the excised gel sections that were subjected to GeLC-MS/MS analysis. Adjacent numbers represent the different ORFs that were detected in the denoted gel sections. The numbers in the parentheses represent the number of unique peptides that identified the designated ORFS.

with various other (putative) nudiviruses (Boucias et al., 1989; Payne, 1974). However, the low molecular bands (< 20kDa) were not resolved in a previous analysis (Coler et al., 1993) that utilized a non-gradient 12% separating gel. In addition to the 16 bands, multiple light-stained bands were detected throughout the gel, suggesting proteolysis had occurred during sample storage and/or preparation.

Genome sequence analysis

A total of 60.5μg of DNA was extracted from a sample containing an estimated 2.5mg of purified virus. End-labeling of undigested MdSGHV DNA was unsuccessful, suggesting a circular genome structure. This result was further confirmed when the genome was completely sequenced, showing that the double-stranded circular DNA of MdSGHV was 124,279bp in size, somewhat smaller than the originally described 137 kbp, estimated by restriction endonuclease analysis (Coler et al., 1993). The end-labeled fragments of the digested DNA showed this viral preparation, to be more homogeneous (derived from a single gland) than the virus extracted from the pooled glands preparation previously analyzed (Coler et al., 1993). The sequence analysis, however, confirmed that our viral preparation was still polymorphic. At least 66 bases were found to be different when sequences from several DNA fragment clones were aligned. The differences at a given nucleotide position were always found to be nucleotide transitions. The most abundant base (present in the majority of the clones sequenced) was chosen as the consensus. The location of each polymorphic base was mapped in the genome (data not shown) as potential targets for assessing polymorphism of field-collected viral samples. The MdSGHV genome is smaller than the 185 kbp estimated for GpSGHV (Abd-Alla et al., 2007). It should be noted that the estimated genome sizes reported among the nudiviruses range from 96 to 228kb (Wang et al., 2007a). The G + C content of MdSGHV was 43.5%, which is very similar to several baculoviruses, HzNV-1 and OrNV, but it is higher than GbNV, which has only a 28% G + C content (Cheng et al., 2002; Wang et al., 2007b,c). The computer generated EcoRI digestion of the MdSGHV genome resulted in 52 fragments ranging from 10,219bp (EcoRI-A) to 30bp (EcoRI-z). From those, 47 fragments have been cloned, resulting in the construction of 90% of the MdSGHV EcoRI genome library. The physical map is shown in Fig. 3.

Nucleotide 1 of the MdSGHV genome was considered to be the A of the ATG initiation codon for the DNA polymerase gene (*dnapol*). This gene was chosen as the start site in an effort to find a common gene present in all DNA viruses, so that other unclassified DNA viruses could be easily compared. A total of 108 putative ORFs were identified in the MdSGHV genome (Table 1, Fig. 3). There were 53 ORFs (49%) transcribed in the clockwise direction and 55 (51%) in the opposite direction. Interestingly, there were several clusters of ORFs transcribing in one direction. A balance in transcriptional direction is often found in baculoviruses and nudiviruses. A total of 22 ORFs overlapped adjacent ORFs with a maximum of 158 nucleotides between ORFs 105 and 106. These two ORFs had opposite transcriptional orientations and were over 240aa in size. No biological activity could be assigned to ORF 105, since there was no protein homology, nor known motifs, while ORF 106 had homology to the baculovirus *pif-3* gene. Elimination of ORF 105 just to avoid overlapping was considered unsupported at this stage in a poorly known virus, so it was accepted as a putative ORF.

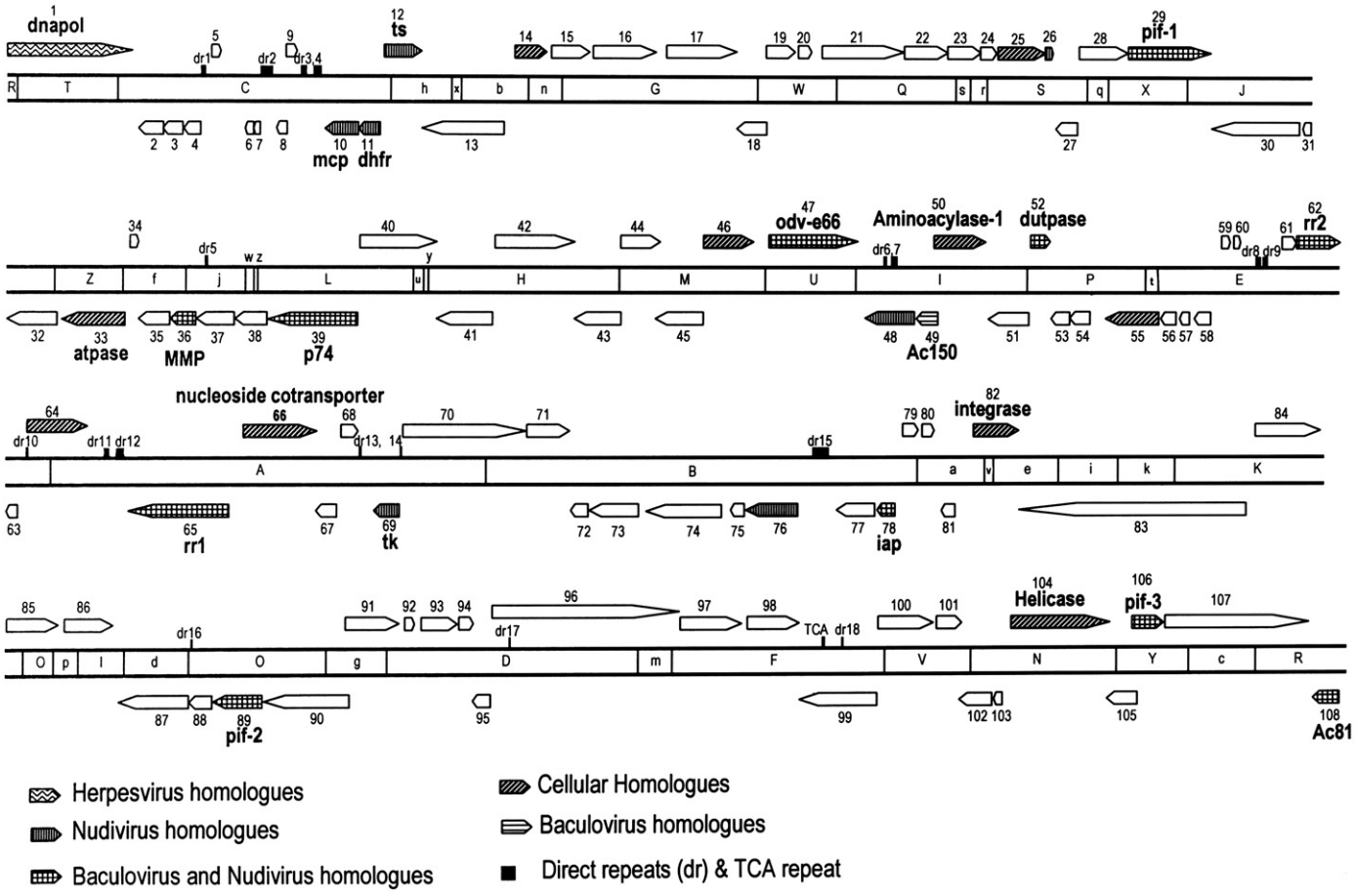


Fig. 3. Linear representation of the MdSGHV genome and EcoRI physical map. The genome location, relative size and transcriptional direction of each putative ORF is indicated by the arrows. Arrow patterns and color indicate homology to organisms found in the current protein database. Dark boxes indicate location of direct repeats.

Table 1
MdSGHV putative ORFs

ORF[a]	Position	Intergene distance (bp)[b]	Size			Best BlastP match				Motifs
			nt.	aa	kDa	ORF or protein encoded	Species	aa identity[c]	E-value	
1	1>2934	153	2934	977	114	DNA Polymerase	Retroperitoneal fibromatosis-associated herpesvirus	27% (766) 1013	5.00E-60	DNA_pol_B, DNA_pol_B_exo
2	3088 < 3669	-10	582	193	23					
S 3	3660 < 4130	26	471	156	19					TM
4	4157 < 4552	231	396	131	15					
5	4784>4993	547	210	69	8					
6	5541 < 5759	-1	195	72	8					
7	5759 < 5920	390	162	53	6					
8	6311 < 6562	-30	252	83	10					
9	6533>6787	703	255	84	10					SP
10	7491 < 8261	-16	771	256	29	Mitochondrial carrier protein	*Oryctes rhinoceros* nudivirus	38% (261) 262	7.00E-46	5TM, SP, Mito_carr
11	8246 < 8758	64	513	170	20	Dihydrofolate reductase	*Heliothis zea* nudivirus 1	26% (163) 264	2.00E-09	SP, Dihydrofolate Reductase
12	8823>9707	106	885	294	34	Thymidylate synthase	*Oryctes rhinoceros* nudivirus	53% (293) 321	9.00E-88	Thymidylat_synt
S 13	9814 < 11748	185	1935	644	75					
S 14	11934>12683	73	750	249	29	Molybdopterin oxidoreductase	Mycobacterium sp	28% (194) 743	7.00E-02	
15	12757>13656	95	900	299	35					
S 16	13752>15251	271	1500	499	58					
17	15523 > 17184	-4	1662	553	64					TM, SP
18	17181 < 17882	-22	702	233	26					
19	17861>18538	65	678	225	26					SP
S 20	18604>18918	264	315	104	11					2TM
21	19183>21117	10	1935	644	72					
S 22	21128>22159	-40	1032	343	40					
S 23	22120>22890	-11	771	256	29					
24	22880>23266	27	387	128	16					
S 25	23294>24424	-8	1131	376	41	Hypothetical repeat containing protein	*Tetrahymena thermophila*	26% (120) 2640	6.00E-04	2TM
26	24417>24599	112	183	60	7					
27	24712 < 25221	39	510	169	20					
S 28	25261>26406	5	1146	381	42					
29	26412>28346	28	1935	644	74	Per os infectivity factor (pif-1)	*Neodiprion abietis* NPV	24% (357) 537	1.00E-08	TM, SP
30	28375 < 30453	43	2079	692	79					
31	30497 < 30697	12	201	66	7					TM/SP
S 32	30710 < 31891	101	1182	393	45					
S 33	31993 < 33486	95	1494	497	56	Putative vacuolar sorting ATPase	*Candida albicans*	26% (268) 439	1.00E-12	AAA ATPASE
34	33582>33776	44	195	64	7					TM, SP
35	33821 < 34549	15	729	242	27					TM, SP
36	34565 < 35155	9	591	196	23	Matrix metalloproteinase (MMP)	Invertebrate iridescent virus 6	34% (101) 264	1.00E-07	TM, SP, metalloproteases (ZnMc_MMP)
37	35165 < 36049	9	885	294	33					TM, SP
S 38	36059 < 36811	8	753	250	29					
S 39	36820 < 38943	53	2124	707	82	p74 protein	*Heliothis zea* nudivirus 1	22% (316) 682	1.00E-07	TM, Baculo_p74_N
S 40	38997>40823	31	1827	608	69					
41	40855 < 42180	-17	1326	441	52					
42	42164>44047	12	1884	627	72					
43	44060 < 45169	-52	1110	369	44					
S 44	45118>46053	-24	936	311	36					SP
45	46030 < 47172	-61	1143	380	45					
46	47112>48299	128	1188	395	46	Hypothetical protein	Bradyrhizobium sp.	28% (140) 457	2.00E-05	
S 47	48428>50530	465	2103	700	81	Occlusion-derived virus envelope protein (ODV-E66)	*Glossina pallidipes* SGHV	35% (346) 346	7.00E-42	TM, SP, Baculo_E66

(*continued on next page*)

Table 1 (*continued*)

ORF[a]	Position	Intergene distance (bp)[b]	Size			Best BlastP match				Motifs
			nt.	aa	kDa	ORF or protein encoded	Species	aa identity[c]	E-value	
48	50996 < 52168	16	1173	390	43	Hypothetical protein LNTAR 25495	*Lentisphaera araneosa*	26% (246) 794	1.00E-03	SP
49	52185 < 52688	-147	504	167	19	Ac ORF 150-like	*Autographa californica* NPV	30% (98) 99	3.00E-06	Chitin binding perotrophin A domain (CBM-14)
S 50	52542>53756	31	1215	404	46	Aminoacylase 1	*Sus scrofa*	28% (413) 407	2.00E-37	AMINOACYLASE-1, Zn-dependent exopeptidases
51	53788 < 54741	86	954	317	36					
S 52	54828>55280	111	453	150	16	dUTPase	Grouper iridovirus	51% (141) 159	7.00E-33	dUTPase
53	55392 < 55832	-46	441	146	18					
54	55787 < 56272	118	486	161	19					
55	56391 < 57641	36	1251	416	49	Hypothetical protein LNTAR_10481	*Lentisphaera araneosa*	36% (84) 234	6.00E-04	
56	57678 < 58025	178	348	115	13					
57	58204 < 58425	71	222	73	8					TM/SP
58	58497 < 58892	309	396	131	16					
59	59202>59399	173	198	65	7					SP
60	59573>59728	905	156	51	6					
61	60634>60957	21	324	107	12					
62	60979>62010	102	1032	343	40	Ribonucleoside diphosphate reductase small subunit (rr2)	*Lymantria dispar* NPV	72% (337) 348	4.00E-142	TM, Ribonuc_red_sm (RNRR2)
63	62113 < 62385	205	273	90	10					
64	62591>64018	974	1428	475	52	Zinc finger protein	*Danio rerio*	44% (49) 450	4.00E-03	zinc finger (C3HC4), RING
65	64993 < 67350	303	2358	785	89	Ribonucleoside diphosphate reductase large subunit (rr1)	*Spodoptera litura* NPV	61% (786) 770	0	Ribonuc_red_lgC (RNR_1)
66	67654>69393	-12	1740	579	64	Sodium/nucleoside cotransporter	*Aedes aegypti*	24% (451) 602	4.00E-32	13TM, Sodium/Nucleoside cotransporter (Nucleos_tra 2_C)
67	69382 < 69870	79	489	162	19					
68	69950>70348	388	399	132	15					2TM
69	70737 < 71348	71	612	203	24	Thymidine kinase (tk)	*Spodoptera frugiperda* ascovirus	29% (195) 210	1.00E-12	Deoxynucleoside kinase (dNK)
70	71420>74323	3	2904	967	111					
S 71	74327>75328	57	1002	333	37					
S 72	75386 < 75796	42	411	136	16					TM
73	75839 < 77011	23	1173	390	46					
74	77035 < 79131	46	2097	698	79					Galactose-binding domain-like
75	79178 < 79504	20	327	108	12					
76	79525 < 80673	881	1149	382		OrNV ORF C3	*Oryctes rhinoceros* nudivirus	38% (386) 448	6.00E-63	PIN domain-like
77	81555 < 82451	24	897	298	33					
78	82476 < 82904	52	429	142	17	Inhibitor of apoptosis (iap)	ORF MSV248 *Melanoplus sanguinipes* entomopoxvirus	33% (140) 150	2.00E-19	IAP, BIR
79	82957>83325	265	369	122	14					TM, SP
80	83591>83875	259	285	94	11					
81	84135 < 84419	192	285	94	10					
82	84612>85691	63	1080	359	43	Integrase/recombinase protein	*Sulfolobus acidocaldarius*	24% (153) 287	4.00E-02	Phage_integrase, DNA breaking & rejoining (INT_REC_C)
83	85755 < 91097	156	5343	1780	206					
S 84	91254>92783	60	1530	509	60					
S 85	92844>94055	154	1212	403	46					
S 86	94210>95355	70	1146	381	43					
87	95426 < 97060	42	1635	544	64					
88	97103 < 97642	25	540	179	21					2TM, SP
89	97668 < 98807	56	1140	379	42	Per os infectivity factor 2 (pif-2)	*Ecotropis obliqua* NPV	26% (186) 392	3.00E-10	TM, SP, Baculo_44
S 90	98864 < 100882	-59	2019	672	78					
S 91	100824>102095	125	1272	423	48					
92	102221>102442	126	222	73	8					TM, SP
93	102569>103420	64	852	283	32					TM, SP
S 94	103485>103820	-4	336	111	13					TM
95	103817<104239	20	423	140	16					

Table 1 (*continued*)

ORF[a]	Position	Intergene distance (bp)[b]	Size			Best BlastP match				Motifs
			nt.	aa	kDa	ORF or protein encoded	Species	aa identity[c]	E-value	
S 96	104260>108681	37	4422	1473	165					
S 97	108719>110167	162	1449	482	57					
98	110330>111544	-18	1215	404	47					9TM
99	111527<113365	18	1839	612	68					
100	113384>114751	6	1368	455	50					
101	114758>115354	-42	597	198	23					
S 102	115313<116086	30	774	257	30					
103	116117<116320	161	204	67	8					
104	116482>118794	-14	2313	770	89	Helicase-like	Parabacteroides	34% (72) 708	8.00E-02	p-loop containing triphosphate hydrolases
105	118781<119506	-158	726	241	28					TM, SP
106	119349>120077	1	729	242	28	Per os infectivity factor 3 (pif-3)	*Spodoptera litura* GV	25% (153) 215	5.00E-03	TM, SP, DUF666
107	120079>123456	188	3378	1125	131					
108	123645<124262	16	618	205	24	Ac ORF 81-like	*Neodiprion lecontei* NPV	32% (108) 175	1.00E-03	2TM

[a] S corresponds to ORFs identified as structural proteins. See Table 2 for details.
[b] "-" indicates number of nucleotides overlapping between adjacent ORFs.
[c] % of amino acid identity (number of aa used in the identity comparison) total number of aa from the compared protein.

MdSGHV gene content analysis

From the putative 108 ORFs found in MdSGHV, only 30 had significant homology to proteins from the current database (Table 1, Fig. 3). From those, 11 had homology to proteins coded by both baculovirus and nudivirus genes, seven to nudivirus, one to baculovirus, one to herpesvirus, and ten to cellular genes. From the 12 ORFs homologous with proteins encoded by baculovirus genes, only five were from the conserved core of 29 genes (Garcia-Maruniak et al., 2004). The highest number of homologous ORFs was found with the nudiviruses (18 in total), although just seven of those are present in all of the nudiviruses completely sequenced to date: HzNV-1 virus (Cheng et al., 2002), GbNV (Wang et al., 2007b), and HzNV-2 virus (information obtained from Wang et al., 2007a).

Analysis of structural proteins

A total of 29 MdSGHV ORFs encoding structural proteins were identified from the GeLC-MS/MS analysis of the SDS-PAGE gel of the purified virus preparation (Fig. 2; Table 2). The number of proteins detected was less than that detected in enveloped baculovirus (Braunegel et al., 2003; Perera et al., 2007). For example, Perera and colleagues, using the GeLC-MS/MS approach, identified 44 structural proteins in the nucleopolyhedrovirus of *Culex nigripalpus.* In many cases in our analysis, the calculated molecular mass of the gel band contained peptide data that corresponded to an ORF(s) that encoded for a similar-sized peptide. For example, the major band b (~ 145kDa) contained peptides that accounted for 67% coverage of ORF 96 (165kDa), whereas band h (~ 38kDa) contained peptides that accounted for 61% and 70% coverage of ORF 86 (43 kDA) and ORF 22 (40kDa) respectively. Additionally, the LC-MS/MS analysis identified peptide fragments that originated from an array of ORFs that had molecular masses different from the calculated values of the gel fragments. For example, similar motifs identified from ORFs 96 and 86 were found in all gel fragments, suggesting a breakdown of certain protein complexes. Alternatively, motifs from ORF 94 (13kDa) were detected in bands h (43kDa) and i (35kDa), suggesting an incomplete solubilization of the structural peptides. Of the 29 structural ORFs, only five ORFs (13, 16, 39, 50, and 91) had less than 10% coverage. Fourteen ORFs had 10–50% coverage, and ten ORFs had coverage greater than 50% (Table 2). It should be noted that all 29 ORFs identified as *bona fide* structural peptides contained unique peptide fragments, ranging from 6 to 27 amino acids that had a 100% match to the translated MdSGHV putative ORFs. The structural ORFs appeared to be clustered loosely on the genome; a total of 17 structural ORFs were located between MdSGHV013 and MdSGHV052, and eight structural ORFs were located between MdSGHV084 and MdSGHV097.

MdSGHV proteins homologous to proteins involved in DNA replication

Two ORFs, MdSGHV001 and MdSGHV104, had homology to the DNA polymerase and helicase enzymes, respectively (Table 1), and both are known to be involved in DNA replication. The best match of MdSGHV001 to viral genes was to a herpesvirus DNA polymerase with 22.8% amino acid identity. No significant BLASTP matches were obtained with either the baculovirus or nudivirus DNA polymerases, even when the comparison was done with the local database containing only nudiviruses ORFs. The DNA polymerase of MdSGHV had homology to a delta 1 catalytic subunit similar to the DNA-directed DNA polymerase type B family that has a DNA binding polymerase and a 3′-5′ exonuclease activity (Shamoo and Steitz, 1999). The conserved nucleotide binding site K(3×)NSxYG(2×)G motif at the C terminus was found almost intact in the MdSGHV DNA polymerase putative peptide (KLCANAIYGLLG starting at aa 631). In order to asses the relationship of MdSGHV to other DNA virus families, a phylogenetic analysis was done on the DNA polymerase gene. The results show that MdSGHV DNA polymerase was closer related to the baculovirus DNA polymerase than to the nudivirus DNA polymerase (Fig. 4A). MdSHGV001 clustered close to Herpesviridae, iridoviruses and ascoviruses. The baculovirus, poxvirus and herpesvirus families appeared to have monophyletic DNA polymerase lineages. Without additional family members, viruses such as iridescent virus, ascoviruses, densoviruses, shrimp white spot syndrome virus (SWSSV), nudiviruses and MdSGHV do not show direct descent and origins of their DNA polymerases.

The best match of MdSGHV104 was to a *Parabacteroides* helicase-like protein, although the homology was low. Helicase is one of the 29 conserved baculovirus core genes, and ORFs with homology to helicase-2 have been described in HzNV-1, HzNV-2 and OrNV (Wang et al., 2007a). MdSGHV104, however, presented no significant homology to either the baculovirus or the nudivirus peptides. The viral replication genes from MdSGHV seem to be distinct from other

Table 2
Structural proteins identified from MdSGHV by mass spectrum analysis

MdSGHV ORF	MdSGHV predicted peptide length (aa)	Number of unique peptide fragments with 100% identity to MdSGHV predicted proteins	Coverage of total MdSGHV predicted protein (%)
3	156	5	33
13[a]	644	13	5
14	249	18	76
16[a]	499	3	7
20	104	3	20
22	343	25	70
23	256	16	75
25	376	4	14
28	381	7	20
32	393	4	11
33	497	25	64
38	250	11	60
39[a]	707	5	7
40	608	29	65
44	311	10	34
47	700	30	55
50[a]	404	2	4
52	150	3	27
71	333	6	19
72	136	5	43
84	509	5	10
85	403	5	16
86	381	29	70
90	672	10	19
91[a]	423	3	9
94	111	8	47
96	1473	82	71
97	482	19	59
102	257	8	30

The percent coverage and number of unique fragments represents the pooled amino acid sequence attained from all gel slices analyzed using the GeLC-MS/MS analysis.
[a] Indicates MdSGHV ORFs with protein coverage of less than 10%.

insect DNA viruses. Interestingly, no baculovirus transcriptional regulator gene homologues were found in the MdSGHV genome. Also, there was no homology to any of the early transcriptional genes, to the late transcriptional activators, nor to the genes encoding for known viral RNA polymerases. Analysis of the DNA replication and transcriptional genes of other SGHVs could help decipher the strategy used by these viruses.

MdSGHV proteins homologous to baculovirus structural proteins

Six MdSGHV ORFs had homology to known baculovirus structural proteins (Table 1): MdSGHV029 (pif-1), MdSGHV039 (p74), MdSGHV047 (odv-e66), MdSGHV049 (Ac150), MdSGHV089 (pif-2) and MdSGHV106 (pif-3). While *pif-1*, *pif-2*, *pif-3* and the *p74* genes belong to the core of conserved baculovirus genes and are present in all completely sequenced nudiviruses, *odv-e66* is present only in the nucleopolyhedrovirus (alphabaculovirus) and granulovirus (betabaculovirus) that infect Lepidoptera hosts (Jehle et al., 2006) and in OrNV (Wang et al., 2007c). On the other hand, *Ac150*-like is present in all GVs but not all NPVs and is not present in nudiviruses. The per os infectivity factor (pif) proteins seem to be associated with the baculovirus occlusion-derived viruses (ODVs), are released from occlusion bodies by the alkaline environment of the insect midgut, and have been reported to be responsible for both oral infectivity and host-to-host dispersal of these viruses. It has been shown that p74 (Kuzio et al., 1989), pif-1 (Kikhno et al., 2002), pif-2 (Pijlman et al., 2003), and pif-3 (Ohkawa et al., 2005) are essential proteins for oral infection. The phylogenetic analysis of the combined p74, pif-1, pif-2 and pif-3 proteins showed a distant relationship between MdSGHV and the baculoviruses (Fig. 4B), and that there is no common ancestor with the nudiviruses with the current number of viruses with pif proteins analyzed. There appears to be a monophyletic relationship amongst the baculovirus combined pifs including lepidopteran, hymenopteran and dipteran viruses. The HzNV-1 and GbNV pifs appear to branch together, but the MdSGHV pifs are basal, as are the nudiviruses to the baculovirus pif proteins.

MdSGHV049 had homology to the *Autographa californica* NPV (AcMNPV) ORF 150 (Ac150). Transmission electron microscopy coupled with immunogold labeling of Ac150 show this protein is associated mainly with enveloped virions before and after they had been occluded within the nuclei of infected Sf-21 cells, and it is also found in budded virus (Lapointe et al., 2004). Deletion of Ac150 decreases oral infectivity, but has no effect when the budded viruses (BV) are injected into the insect hemocoel indicating that Ac150, can be considered a *per os* infection factor that mediates, but is not essential for oral infection (Zhang et al., 2005). The C terminus of MdSGHV049 (from aa 98 to 152) had a chitin-binding peritrophin-A (CBM-14) domain similar to Ac150. This domain is found in many chitin-binding proteins including peritrophic matrix proteins of insects and animal chitinases (Tellam et al., 1999).

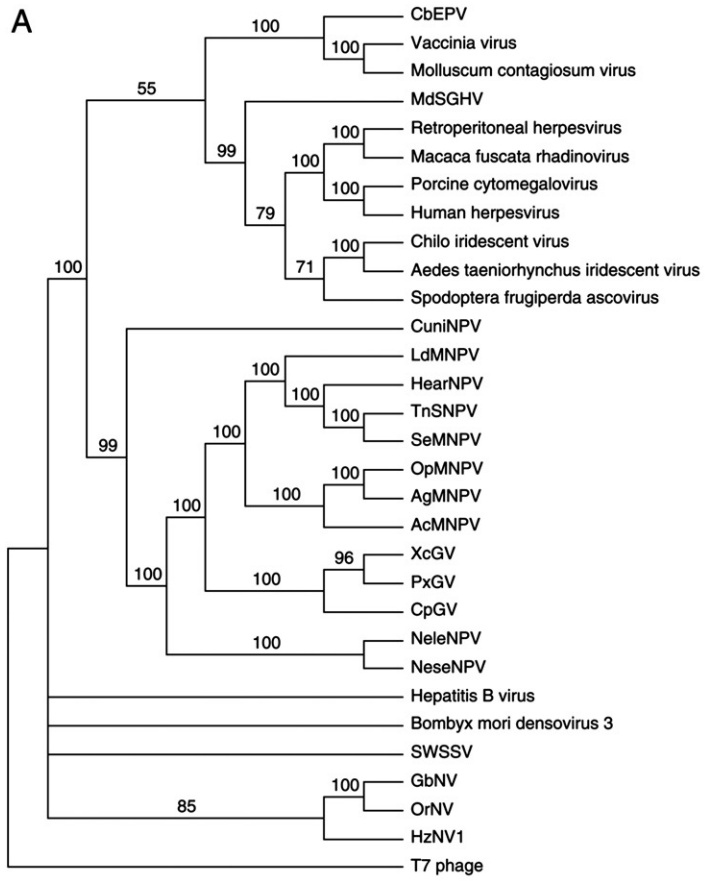
MdSGHV does not produce occlusion bodies, so there is no ODV form of the virus. Although six putative genes were found in the MdSGHV genome that had significant amino acid identity to known structural genes (Table 1), the mass spectrometry analysis of the structural proteins separated by an acrylamide gel found only two genes coding for the p74 (MdSGHV039) and odv-e66 (MdSGHV047) proteins associated with the enveloped MdSGHV (Table 2). Possibly, the MdSGHV pif proteins either could be produced in nondetectable quantities as reported for *Spodoptera littoralis* NPV (SpliNPV) pif-1 (Gutierrez et al., 2004), or may not be structural proteins. If pif-1, pif-2, pif-3 and Ac150-like proteins are expressed and are important in the oral infection of MdSGHV they may be delivered in the salivary secretions to the next host. The predicted transmembrane domains of the MdSGHV pif-1, pif-2, pif-3 and odv-e66 were localized in the N-terminal region, while the p74 transmembrane motif was found in the C-terminus of the putative peptides, similar to those proteins in the baculoviruses.

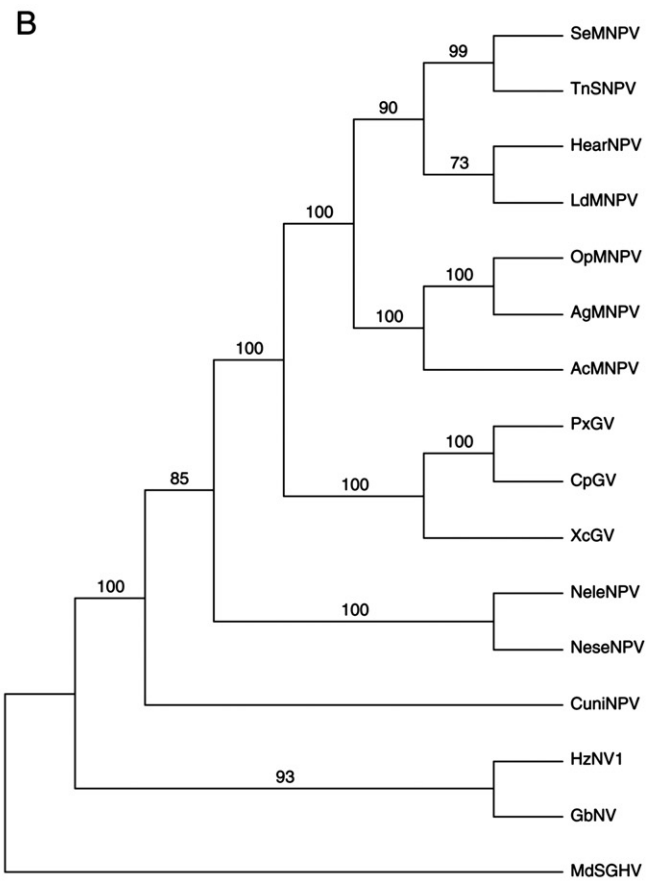
Although pif-1, pif-2, pif-3 and p74 have been described in the genomes of other nudiviruses (Wang et al., 2007a), there was no significant amino acid homology between MdSGHV pif-3 and the nudiviruses pif-3. MdSGHV odv-e66 had the best amino acid homology (35% identity) with GpSGHV. This gene has also been found in the genome of OrNV (Wang et al., 2007c), but not in the other two nudiviruses (HzNV-1 and GbNV).

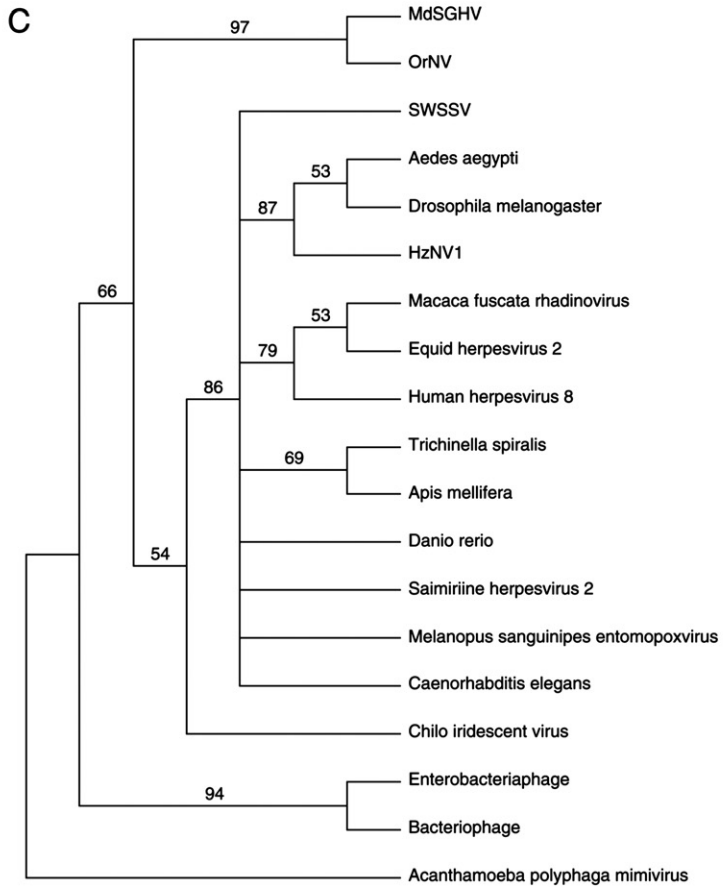
MdSGHV proteins homologous to auxiliary and undefined baculovirus proteins

MdSGHV036 is homologous to the zinc-dependant matrix metalloproteinase (MMP) of several viruses including the granuloviruses, HzNV-1, entomopoxvirus and the insect iridescent virus. MMPs are responsible for pericellular proteolysis of extracellular matrix and cell surface molecules and, in many instances, they are anchored to cell

Fig. 4. Phylogenetic analysis of selected MdSGHV putative proteins: A. DNA Polymerase; B. combined analysis of p74, pif-1, pif-2 and pif-3; C. thymidylate synthase. The trees were inferred using neighbor-joining in PAUP*, and a bootstrap analysis with 1000 replicates was performed to measure support for the branches. Organisms included in this analysis with abbreviated names: *Anticarsia gemmatalis* nucleopolyhedrovirus (AgMNPV), *Autographa californica* MNPV (AcMNPV), *Choristoneura biennis* entomopoxvirus (CbEPV), *Culex nigripalpus* NPV (CuniNPV), *Cydia pomonella* granulovirus (CpGV), *Gryllus bimaculatus* nudivirus (GbNV), *Helicoverpa armigera* NPV (HearNPV), *Heliothis zea* nudivirus 1 (HzNV1), *Lymantria dispar* MNPV (LdMNPV), *Musca domestica* salivary gland hypertrophy virus (MdSGHV), *Neodiprion lecontei* NPV (NeleNPV), *Neodiprion sertifer* NPV (NeseNPV), *Orgyia pseudotsugata* MNPV (OpMNPV), *Oryctes rhinoceros* nudivirus (OrNV), *Plutella xylostella* GV (PxGV), Shrimp white spot syndrome virus (SWSSV), *Spodoptera exigua* MNPV (SeMNPV), *Trichoplusia ni* SNPV (TnSNPV), *Xestia c-nigrum* GV (XcGV).


A
CbEPV
Vaccinia virus
Molluscum contagiosum virus
MdSGHV
Retroperitoneal herpesvirus
Macaca fuscata rhadinovirus
Porcine cytomegalovirus
Human herpesvirus
Chilo iridescent virus
Aedes taeniorhynchus iridescent virus
Spodoptera frugiperda ascovirus
CuniNPV
LdMNPV
HearNPV
TnSNPV
SeMNPV
OpMNPV
AgMNPV
AcMNPV
XcGV
PxGV
CpGV
NeleNPV
NeseNPV
Hepatitis B virus
Bombyx mori densovirus 3
SWSSV
GbNV
OrNV
HzNV1
T7 phage


B
SeMNPV
TnSNPV
HearNPV
LdMNPV
OpMNPV
AgMNPV
AcMNPV
PxGV
CpGV
XcGV
NeleNPV
NeseNPV
CuniNPV
HzNV1
GbNV
MdSGHV


C
MdSGHV
OrNV
SWSSV
Aedes aegypti
Drosophila melanogaster
HzNV1
Macaca fuscata rhadinovirus
Equid herpesvirus 2
Human herpesvirus 8
Trichinella spiralis
Apis mellifera
Danio rerio
Saimiriine herpesvirus 2
Melanopus sanguinipes entomopoxvirus
Caenorhabditis elegans
Chilo iridescent virus
Enterobacteriaphage
Bacteriophage
Acanthamoeba polyphaga mimivirus

membranes via *trans*-membrane domains (Baumann et al., 1993). MdSGHV036 had a transmembrane domain in its N-terminus (Table 1), although it had not been detected by the mass spectrum analysis of MdSGHV as a structural protein. In granuloviruses this protein is believed to disrupt the peritrophic matrix or the basement membranes that help hold tissues together thus assisting in virus dissemination (Ko et al., 2000).

MdSGHV052 had a high homology to a dUTP pyrophosphatase (dUTPase) of the iridovirus (51% aa identity). This gene, also present in several baculoviruses and in HzNV-1, is involved with nucleotide transport and metabolism. UTPase hydrolyzes the triphosphate DNA polymerase substrate dUTP to non-substrate monophosphate (dUMP) to prevent incorporation of uracil into the DNA while providing the substrate for thymidine synthesis (Harris et al., 1999). The protein coded by MdSGHV052 was detected by mass spectrometry to be a structural protein (Table 2).

The highest amino acid homology was found for MdSGHV062 and MdSGHV065 to *Lymantria dispar* NPV (LdMNPV) and *Spodoptera litura* NPV (SpltNPV) ribonucleotide reductase small (rr2) and large (rr1) subunits, with 72% and 61% aa identities, respectively (Table 1). These two MdSGHV ORFs were separated by almost 3 kbp and had opposite transcriptional orientations. The genes coding for rr1 and rr2 are present in viruses including nudiviruses and baculoviruses, as well as eukaryotes and many prokaryotes.

MdSGHV078 encoded an inhibitor of apoptosis (iap) homologue. The iaps were first discovered in baculoviruses and are present in various animal species (Clem, 1997; Uren et al., 1998). They have one to three consensus amino-terminal regions named BIR (baculoviral iap repeat) domains, essential to inhibit apoptosis by acting as a direct inhibitor of the caspase family of protease enzymes (Miller, 1999). The iaps also have a carboxy-terminal zinc-binding motif named ring finger, which appears to be involved in ubiquitination of the iap itself and potentially, any caspase bound to it (Wilson et al., 2002). The MdSGHV078 encodes a 142aa peptide that is smaller than the majority of the baculovirus iaps, but very similar in size to the *Melanoplus sanguinipes* entomopoxvirus iap which had the highest homology to MdSGHV078, although baculoviruses and nudiviruses iap-3 also had significant homology. MdSGHV078 had only one BIR motif in its N-terminus with the three conserved cysteines (aa position 46, 49 and 73) and one histidine (aa 66) in the expected conserved locations, but did not have the zinc finger domain at its C-terminus. Other baculoviruses have iap proteins with only one BIR motif, and since not all iap are capable of preventing cell death (Maguire et al., 2000), the function of this ORF will need to be confirmed experimentally.

A homologue of AcMNPV ORF 81(Ac81) present in all the baculoviruses and nudiviruses was found in the MdSGHV genome. The best homology was found with the *Neodiprion lecontei* NPV Ac-81 like protein. MdSGHV108 has two transmembrane domains at its C-terminus. Although this gene belongs to the core of baculovirus conserved genes, its functions has not been defined (Slack and Arif, 2007).

MdSGHV proteins homologous to nudivirus proteins

Seven ORFs (MdSGHV ORFs 10, 11, 12, 26, 48, 69 and 76) had homology to nudivirus putative proteins not found in any baculovirus genome (Fig. 3, Table 1). Five of those are cellular genes, the majority of which have been well studied in eucaryotes (see below). MdSGHV010 has homology with a mitochondrial carrier protein present in OrNV (ORF C7). The mitochondrial carrier protein family comprises a variety of small proteins (28 to 34kDa) that catalyze the exchange of substrates across the inner mitochondrial membrane (IMM), and have three tandemly repeated mitochiondria carrier domains of approximately 100aa each (Walker and Runswick, 1993). MdSGHV010 had a predicted molecular weight of 29kDa and two mitochondria carrier domains at its C terminus (between aa 83–179 and 176–256, respectively).

MdSGHV011 was homologous to HzNV-1 dihydrofolate reductase (DHFR). DHFR is a ubiquitous enzyme that plays a key role in DNA replication and cell division, since the enzyme is essential in providing purines and pyrimidine precursors for the biosynthesis of DNA, RNA and amino acids (Miller and Benkovic, 1998).

MdSGHV012 was homologous to a putative thymidylate synthase (TS) found in the genomes of OrNV, HzNV-1, HzNV-2, SWSSV, herpesviruses and other viruses. The TS is involved in the biosynthesis of DNA precursors and, hence, it is one of the most conserved enzymes across species and phyla (Perryman et al., 1986). The phylogenetic analysis of several viral and cellular TS grouped MdSGHV012 closer to OrNV than to HzNV-1 TS (Fig. 4C). The lack of monophyletic structure in the tree suggested that this gene was acquired from the host at multiple independent occasions rather than originating from a common DNA virus ancestor. This is further corroborated by the tree positions found between HzNV-1 and insects, as well as an entomopoxvirus and herpesviruses basal with insects and a nematode. In several protozoa, a single polypeptide chain codes for both, DHFR and TS, forming a bifunctional enzyme (DHFR-TS), possibly through gene fusion at a single evolutionary point. Interestingly, MdSGHV ORFs 11 and 12 were adjacent to each other although with opposite transcriptional directions, and their functions (if confirmed to be active proteins) are interconnected. The size of both putative proteins was similar to those of active proteins. As mentioned before, the DNA polymerase and helicase genes from MdSGHV are distantly related to those of the baculoviruses and other insect viruses. Potentially, different genes, not even present in the baculoviruses are key players in the DNA replication of MdSGHV.

MdSGHV069 was a 203 aa putative protein homologous to the deoxynucleoside kinase (dNK) family protein. The best match was found to *Spodoptera frugiperda* ascovirus 1a thymidine kinase (tk). This protein is produced by several cellular organisms and viruses such as HzNV-1 (ORF 70). The are key enzymes in the salvage of deoxyribonucleosides originating from extra- or intracellular breakdown of DNA (Johansson et al., 2001). MdSGHV069 contained the conserved substrate-binding sites of this protein family (cd01673).

MdSGHV026 was a 60 aa long, serine-rich putative protein. This ORF had homology only to GbNV ORF 73 (Wang et al., 2007b). MdSGHV048 presented a low homology to a cellular hypothetical gene and also to HzNV-1 ORF 153. Lastly, MdSGHV076 was homologous to unknown proteins detected in nudivirus genes including OrNV-C3, GbNV ORF 65, HzNV-1 ORF 68 and HzNV-2 ORF 70 (Cheng et al., 2002; Wang et al., 2007a,b,c).

MdSGHV ORFs with homology to cellular genes

Ten ORFs had homology to cellular genes (Fig. 3, Table 1). MdSGHV014 had a low homology to a molybdopterin oxidoreductase from Mycobacterium. MdSGHV033, was similar to the putative vacuolar sorting ATPase from *Candida albicans.* It had an AAA-ATPase motif that belongs to the AAA-superfamily of ATPases, associated with a wide variety of cellular activities, including membrane fusion, proteolysis, and DNA replication (Patel and Latterich, 1998). MdSGHV050 was similar to a cellular aminoacylase-1. Mass spectral analysis showed that these three ORFs code for structural proteins of MdSGHV (Table 2). MdSGHV064 had a RING-, and a zinc-finger (C3HC4) motifs. The RING-finger domain is a specialized type of a Zinc-finger of 40 to 60 residues that binds two atoms of zinc; it is defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48)C-X2-C. This C3HC4 motif was present between aa 214–251 of MdSGHV064. Proteins carrying this motif have been identified with a wide range of functions such as viral replication, signal transduction, and probably are involved in mediating protein–protein interactions (Saurin et al., 1996). MdSGHV066 is a 579aa putative sodium dependent nucleoside transporter protein. The conserved motif of the nucleoside transporter was in the C-terminal

Table 3
MdSGH direct repeat (dr) sequences

Name	Repeat size (bp)×number of copies	dr sequence and genomic position	AT%	Total size (bp)
dr1	21×3.3	4559 CGAGTTTGTCTTTTAATATGAAC 4581 4582 CGAGTTTG-CTTT-GATGTGTAC 4602 4603 CGAGTTTG-TTTT-AATATGAAC 4623 4624 CGAGTTT 4630	65	72
dr2	64×4.3	5933 TGTTTGTGTGTCACTCCGATTTTA--T--TTAATAGACTGTTCCGCCTAGT-CCCATAAGAAATATACA 5996 5997 TGTTTGTGTGTCACTCCGATTTTATTTGGTTGGTATGGACTATTAGTGGGT-CCCATAAGAAATATACA 6064 6065 TGTTTGTGTGTCACTCCGATTTCA--T--TTAATAGACTGTTTCGTTTGGT-CCCATATAAAATATACA 6128 6129 TGTTTGTCTGTCACTCCGATTTTAA-T---TAGTGGACCGTTTCGTTTGGTCCCCATAAGAAATGTACA 6193 6194 TGTTTGTGTGTCACTCCG 6211	63	279
dr3	53×2.3	6912 TGTAATTTGGGAGCAGAACTTGACCGACAGAACTCAGCCTGTATAGTTGTTAG 6964 6965 TGTAATTTGGGAGCAGAATTTAACCGACAGAACTCAGCCTGCATTGTTGTTAG 7017 7018 TGTAATTTGTGGAAGCAGAA 7037	59	126
dr4	26×7.4	7173 TTGTTTTTGTAATCCAG-CCTGCATAG 7198 7199 TTGTTTTTGTACTCTAG-CCTGTATAG 7224 7225 TTG-TTTTGTACTCGAG-CCTGTATAG 7249 7250 TTGTTTTTGTACTCGAG-CCTGCATAG 7275 7276 TTGTTTTTGTACCCATG-TCTGTATAG 7301 7302 TTGTTAGTGTAATTC-G-CTACATGAAA 7327 7328 TTGTTTTTGTAATC-AGGCCTGCATAG 7353 7354 TTGTTTTTGTA 7364	64	192
dr5	24×2.7	35734 GTGGTGCCT-GGGACGGAGCTTCCT 35757 35758 GTGGTGGTGGGGATGGAGG-TTCCT 35781 35782 GTGGTGCTGGGGATGG 35797	34	64
dr6	30×2.6	51725 TTGATCTCCTGGGGCTTCGGTTGGGGCAGA 51754 51755 TTGATCTCCTGGGGCTTCGGTTGGGGCAGA 51784 51785 TTGATCTTCTGAGGCTT 51801	44	77
dr7	54×2.7	51918 TGATCTTCTGGGACTTCGTTAACGGACTGATCTCCTTGGGCCTCGTAAACGGAC 51971 51972 TGATCTCCTGGGTTTTCGTTAACGGATTGATCTCCTTGGGCCTCGTCAACGGAC 52025 52026 TGATCTCCTGGGTTTTCGTTAACGGATTGATCTCCTTGGG 52065	49	148
dr8	43×2.9	60302 GACGGTCGACAGAGAAAACACTAACAACAAGTACAGGG-TATAT 60344 60345 GACGGTCGGCAGAGAAAACACTAACAA-AA-TACATGGGGCTTAT 60387 60388 GGTGG-GGGCAGCGAAAACACTAACAACAAGTACAGG 60423	54	122
dr9	44×3.4	60422 GGCTTGCGAATATCGACAGCGGGTTACAAAAACAAAATACATGG 60465 60466 GGCTTGTGGATATCGACAGAGGGTTACAAAAACAAAATATACAG 60509 60510 GCCTTATGAATATCGACAGAGGGTTACAAAAACAAAATATACAG 60553 60554 ACCTTATGAATATCGACAG 60572	62	151
dr10	18×2	62619 AGGAACCAAGTATCCAGG 62636 62637 AGGAACCAAGTATCCAGG 62654	50	36
dr11	59×1.9	64420 TATTTTTCATTTTCATGTTAGTGTATATTTGGACTTAGTATTTTTCAAGTTCGCGAGAA 64478 64479 TATTTTTCATTTTCATCTTCGCATATATTGGGACTTAGTATTTTTCAAGTTCGCGA 64534	71	115
dr12	60×3.0	64704 TTTTCATTTCAGACGAGTAGGATATTTTTT-ATTATCAAGTCCGCATAT-GAC--TTGTGC 64760 64761 TTTTCATTTTCAGATGAGTAGGGTATTTTTCATTTTCAAGACCACATATTGAGACTTTGTA 64821 64822 TTTTCGTTTCAG-TCAAGTAGGGTATTTTTCATTTTCAAGTCCGCATATTGAGACTTGTG 64880	67	177
dr13	23×2.0	70445 ACCCAAGTCCCCATACGAAGAGC 70467 70468 ACCCAAGTCCCCATACAAAGAGC 70490	46	46
dr14	17×2.0	71349 TTTTGGTTTTTGGTTGG 71365 71366 TTTTGATTTTTGGTTGG 71382	68	34
dr15*	149×2.6	81117 TTCA-AACCAAGTTGAAACCTGGCCAGCGAAAAATGAAAAATTGTCTAGCAAAAATTTCTGAAAA 81180 81265 TTCACAACCAAGTTGAAACCCGGCCAGCGAAAAATGAAAAATTG-CAGAGTAAAAATTTCTGAAA 81328 81409 TTCACAACCAAGTTGAAACCCACCCAGTGAAAAATGAAAAATTGTCGAGCAAAAAATTTCTGAAA 81473 81181 AGCAAGCATGATCGGGCATGCACACTCTCTCGTTAGGTATCAGAGGGAGAGGGGTGATGGGGGTG 81245 81329 AACC-GGACATGATCGAGCATGCACACT-T-ACGATA-G-ACTCA-AGGGAG-GGGATGATGGGA 81386 81474 AATCAAGCATGATCGAGCATGCA 81496 81246 AGGAGGGA--T-CGATCTACAA 81264 81387 GAGCGGGGGTTCCGATCCACAA 81408	56	380
dr16	12×2.5	97194 TCATCGTCATCA 97205 97206 TCATCGTCATCA 97217 97218 TCATCG 97223	57	30
dr17	15×2.0	104632 GGTGGAAGTGGTGGA 104646 104647 GGTGGAAGTGATGGA 104661	43	30
TCA repeat	3×15	112083 TCA TCA TCA TCA TCG TCA TCA TCC TCA TCC TCA TCA CCA TCA TCA 112128	58	45
dr18	9×3.8	112519 CATCACCAT 112527 112528 CATCAACAC 112536 112537 CATCACCAC 112545 112546 CATCACC 112552	50	34

*Sequence of each 149 bp direct repeat is presented in 3 different rows.

end of MdSGHV066, and a total of 13 transmembrane domains were found in this ORF. A similar sized protein having 14 transmembrane motifs, capable of selectively transporting pyrimidine nucleoside and adenine has been studied in mammal jejunal epithelium (Huang et al., 1994). MdSGHV082 had homology to an integrase/recombinase protein that is involved in DNA rearrangements by site-specific DNA recombination (Grainge and Jayaram, 1999). The conserved integrase/recombinase motif was found at the C-terminal end of MdSGHV082. Although GbNV also has an integrase protein (ORF 57) (Wang et al., 2007b), there was not significant homology to MdSGHV082. MdSGHV035, MdSGHV046 and MdSGHV055 had homology to functionally unknown cellular proteins. Finally, MdSGHV104 with low homology to helicase gene has been discussed previously (under homologues to proteins involved in DNA replication).

Direct repeats

Eighteen tandem direct repeats (drs) were found distributed throughout the genome. The size of the repeated sequences ranged from 149bp (dr15) to only 9bp long (dr18), and the number of copies of the drs ranged from 1.9 to 7.4, making the total sizes of the drs range from 30bp to 380bp long (Table 3). Unlike the homologous regions (hrs) found in large, circular invertebrate dsDNA viruses that have imperfect palindromic sequences within the repeated region (Cochran and Faulkner, 1983; Garcia-Maruniak et al., 1996), no palindromes were found in the MdSGHV direct repeats. The MdSGHV drs appeared similar to those found in HzNV-1, GbNV, and OrNV and *Neodiprion sertifer* NPV (NeseNPV) (Cheng et al., 2002; Garcia-Maruniak et al., 2004; Wang et al., 2007b,c), however the AT percentage was not as high as GbNV drs. Seven drs (dr5, 6, 7, 10, 16, 17 and 18) were located inside putative ORFs. Baculovirus hrs are located between genes and act as transcriptional enhancers (Guarino and Summers, 1986) and/or origin of DNA replication (Pearson et al., 1992). Another feature of the hrs is that their sequences have homology within the same genome. This feature was not present in the MdSGHV drs, as no significant homology was observed between the 18 drs. A 45bp TCA repeat was located inside ORF 101, similar to the 3 base repeats reported in GbNV (Wang et al., 2007b).

In this study, we have determined the complete genome sequence and provided the first comprehensive analysis of the MdSGHV genome. Completion of the genome sequence has provided a foundation to examine the biology of a unique insect virus. This virus, with the ability to sterilize female flies, has the potential of being developed into an adult biopesticide. The sequence data will play an important role in elucidating the mechanisms underlying its ability to shut down reproductive behaviors and will provide a platform to engineer improved biopesticides against pest fly populations. The sequence data will also be used to characterize genetic variants isolated from feral housefly populations and determine which polymorphisms are associated with the epizootic or the enzootic profiles exhibited by this virus in fly populations. Importantly, related SGHVs have been isolated from tsetse fly vectors of African trypanosomiasis. These SGHVs have a larger genome than MdSGHV, show wider tissue tropism, are vertically transmitted, and are known to persist asymptomatically in tsestse flies. Comparisons among the different SGHV sequences will provide opportunities to assign genotypic profiles to similar and dissimilar viral phenotypes.

Material and methods

Virus production and purification

As an alternative to cloning the virus (no cell lines are known to support MdSGHV replication), a single hypertrophic gland was dissected from an MdSGHV infected feral adult house fly. This gland was homogenized in 1ml of phosphate buffered saline (PBS), filter-sterilized (Ultrafree-MC, Millipore Corp., Billerica, CA), and injected (2μl/fly) into 200 newly emerged adult house flies. Injected house flies were placed in cages, provided with food and water, and incubated at 26°C. After 5days, the injected flies were immobilized on ice, and infected salivary glands were dissected, pooled, and frozen at − 70°C. Glands from 100 infected flies were homogenized in 5mM Tris–HCl (pH 7.5) containing 0.2M sucrose, 0.13M NaCl, 3mM KCl, and 3mM EDTA, subjected to low-speed centrifugation (1500*g*, 20min) to remove tissue debris, applied to a 10–50% isotonic linear gradient of Nycodenz (Accurate Chemicals & Scientific Corp., Westbury, NY), and centrifuged at 30,000rpm (~ 100,000*g*) for 90min. The virus band was collected and subjected to another high-speed centrifugation for 30min to remove residual gradient material. Purified virus was suspended in TE buffer (10mM TRIS, 1mM EDTA), aliquoted and stored at − 20°C. The quality of the virus fraction was determined by electron microscopy. A dilute viral preparation was applied to formvar + carbon-coated grids, fixed with 2% glutaraldehyde in 0.1M cacodylate, water washed, dried, and stained with aqueous uranyl acetate and lead citrate. Grids were examined with an Hitachi H-7000 TEM operating at 100kV and images captured with a Soft-Imaging System MegaView III with AnalySIS.

Analysis and identification of viral structural proteins

An aliquot of the purified virus preparation was solubilized in SDS-β mercaptoethanol and electrophoresed onto a 5 to15% gradient polyacrylamide gel electrophoresis (PAGE) gel (Laemmli, 1970). The Coomassie R-250 blue-stained gel was recorded with the Bio-Rad gel imaging XRS system. Precision Plus Protein unstained standards (Bio-Rad Laboratories, Inc., Hercules, CA) were used to calculate a standard curve (*R*-square = 0.987) and subsequent molecular weights of viral peptide bands were calculated using the Quantity One software (Bio-Rad Laboratories, Inc.).

Coomassie R-250 blue-stained protein bands and interband gel regions were excised and digested in-gel with trypsin (Link et al., 1999). Capillary reversed-phase HPLC separation of protein digests was performed on a 15-cm-by-75-μm (inner diameter) TARGA C18 column (Higgins Analytical, Mountain View, CA) in combination with an Ultimate Capillary HPLC system (LC Packings, Sunnyvale, CA) operated at a flow rate of 200nl/min. Mass spectrometry analysis of the trypsin-digested peptides derived from the one-dimensional gel (GeLC-MS/MS) was accomplished by a hybrid quadrupole time-of-flight MS instrument (QSTAR; Applied Biosystems, Foster City, CA) equipped with a nanoelectrospray source at the Proteomics Core Facility from the Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, Florida. Amino acid sequence data generated by the QSTAR from the MS/MS were searched against the National Center for Biotechnology Information non-redundant sequence database using the Mascot (Matrix Science, Boston, MA) database search algorithm. Mascot was set up to search the MdSGHV ORF database. Scaffold (version Scaffold-01_06_19, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Proteins identified by Mascot were considered significant if individual ion scores exceeded the threshold value calculated for identity or extensive homology ($P \leq 0.05$). The percent coverage of each protein by peptide fragments identified was calculated using only the unique peptide fragments with 100% amino acid identity (i.e., amino acid sequences from recurring peptide fragments or overlapping amino acid sequences were counted only once).

MdSGHV genome purification and EcoRI library construction

Virus was treated with 1.5% sarcosyl and 250μg/ml proteinase K and the DNA was purified by phenol/ether extractions as described by Garcia-Maruniak et al. (2004). To determine if the genome was circular or linear, 400ng of undigested DNA was end-labeled with 2μCi ^{32}P-dATP and DNA polymerase I large fragment (Klenow, New England BioLabs, Beverly, MA). MdSGHV genomic DNA (200ng) digested separately with EcoRI, HindIII, BamHI and XhoI was also end-labeled and run in a 0.8% agarose gel. To construct an EcoRI library, 1μg of DNA was digested, ligated into 350ng pGEM 7Z plasmid (Promega Corp., Madison, WI), and used to transform MAX efficiency DH5-α competent cells (Invitrogen, Carlsbad, CA). Multiple colonies were screened to confirm fragment size. Fragments not cloned during by this approach were cut out of the agarose gel and individually cloned.

Genome sequencing and analysis

Both Sanger (Sanger et al., 1977) and pyrophosphate-based (454) sequencing techniques were used to sequence the MdSGHV genome.

Initially, 192 EcoRI clones were used for cycle sequencing using T7 and SP6 primers and BigDye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) and sent to the University of Florida Interdisciplinary Center for Biotechnology Research (ICBR). Secondly, an aliquot of 20μg of total DNA was submitted to 454 sequencing. This was performed as described in the supplementary material and methods to Margulies et al. (2005) with slight modifications from the specifications (454 Life Sciences, Branford, CT). Briefly, high molecular weight DNA from the RCA reactions was sheared by nebulization to a size range of 300 to 800bp. DNA fragment ends were repaired and phosphorylated using T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides "A" and "B" supplied with the 454 Life Sciences sequencing reagent kit were ligated to the DNA fragments using T4 DNA ligase. Purified DNA fragments were hybridized to DNA capture beads and clonally amplified by emulsion PCR (emPCR). DNA capture beads containing amplified DNA were deposited in individual lanes of a 40 × 75mm PicoTiter plate and DNA sequences determined using the GS-FLX instrument. DNA sequence information from the initial and supplementary runs was combined in a single assembly using Newbler sequence assembly software. Some PCR amplifications followed by sequencing of the DNA products were needed to close every gap of the genome. The contigs assembled by Newbler and the electropherograms obtained from the EcoRI clones or PCR products were further aligned using Sequencher 4.8 (Gene Codes Corp., Ann Arbor, MI). Regions of nucleotide polymorphism were mapped. Putative open reading frames containing at least 50 amino acids, ATG start sites, and presenting minimal overlapping in the genome were identified with ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and Gene Finding in Viruses program from SoftBerry (http://www.softberry.com/berry.phtml). The translated amino acid sequence of each ORF was compared to the GenBank database by protein BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) against the default and virus-only databases. A local database with only the nudivirus protein sequences was created and used to compare all MdSGHV ORFs. Significant homology was arbitrarily set at a BlastP E-value of 0.01 or less. Conserved domains and putative functional protein sites were searched for each ORF using Interproscan (http://www.ebi.ac.uk/interpro/) (Quevillon et al., 2005). Tandem-repeat sequences were found with the help of Tandem Repeat software (http://tandem.bu.edu/trf/trf.html) (Benson, 1999).

Phylogenetic analysis

Amino acid sequences were aligned with ClustalX 2.0 (Thompson et al., 1997). Alignments were edited, when needed, with MacClade 4.0 PPC (Maddison and Maddison, 2000) and the phylogenetic trees were constructed using PAUP* 4.0b (Swofford, 2003). Distance analysis with neighbor joining was done. The robustness of the tree was tested using bootstrap and 1000 replicates. For the construction of the combined tree of p74, pif-1, pif-2, and pif-3, a file containing the concatamer of the four genes was created and analyzed as described before.

Acknowledgments

We acknowledge the support of Chris Geden (USDA/ARS), Hilary Lauzon (Canadian Forest Service, Ontario, Canada) and Khuong Nguyen (University of Florida, Department of Entomology) and extend our gratitude to Tamer Salem, Verena Lietze (University of Florida, Department of Entomology), Adly Abd-Alla and Andrew Parker (Entomology Unit, FAO/IAEA laboratory, Vienna, Austria) for critical reviews of the manuscript. The excellent technical support provided by Scott McClung and Sophie Alvarez (ICBR Proteomics Core), Savita Shanker and Regina Shaw (ICBR Sequencing Core), and Karen Kelly (ICBR Electron Microscopy Core). We gratefully acknowledge the financial support provided by the National Institute of Health (NIAID R21 AI073501-01) and USDA/NRI (2007-35302-18127).

References

Abd-Alla, A., Bossin, H., Cousserans, F., Parker, A., Begoin, 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., Couserans, F., Parker, A.G., Jehle, J.A., Parker, N.J., Vlack, 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.

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. Étude sur *Merodon equestris* (Diptera, Syrphidae). Note. Comp. Rend. Seanc. Acad. Sci. Ser. D Sci. Natur. 289, 481–484.

Baumann, U., Wu, S., Flaherty, K.M., McKay, D.B., 1993. Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: a two-domain protein with a calcium binding parallel beta roll motif. EMBO J. 12, 3357–3364.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580.

Boucias, D.G., Maruniak, J.E., Pendland, J.C., 1989. Characterization of a non-occluded baculovirus (subgroup C) from the field cricket, *Gryllus rubens*. Arch. Virol. 106, 93–102.

Braunegel, S.C., Russell, W.K., Rosas-Acosta, G., Russell, D.H., Summers, M.D., 2003. Determination of the protein composition of the occlusion-derived virus of *Autographa californica*. Proc. Natl. Acad. Sci. U. S. A. 100, 9797–9802.

Burtt, E., 1945. Hypertrophied salivary glands in *Glossina*: evidence that *G. pallidipes* with this abnormality is peculiarly suited to trypanosome infection. Ann. Trop. Med. Parasit. 39, 11–13.

Cheng, C.H., Liu, M.S., 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.

Clem, R.J., 1997. Regulation of programmed cell death by baculoviruses expression. In: Miller, L.K. (Ed.), The Baculoviruses. Plenum Press, New York, pp. 237–266.

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

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.

Feldmann, U., Barnor, H., Acs, R., 1992. Abweichungen in der Reproduktion von *Glossina morsitans submorsitans* Newstead (Diptera: Glossinidae): Untersuchungen zur Bestimmung eines gestörten Geschlechterverhältnisses und zum Übertragungensweg von fertilitäts-reduzierenden Viren an die Nachkommen. Mitteilungen der Deutschen Gesellschaft fuer Allgemeine und Angewandte Entomologie 8, 248–251.

Garcia-Maruniak, A., Pavan, O.H.O., Maruniak, J.E., 1996. A variable region of *Anticarsia gemmatalis* nuclear polyhedrosis virus contains tandemly repeated DNA sequences. Virus Res. 41, 123–132.

Garcia-Maruniak, A., Maruniak, J.E., Zanotto, P.M., Doumbouya, A.E., Liu, J.C., Merritt, T.M., Lanoie, J.S., 2004. Sequence analysis of the genome of the *Neodiprion sertifer* nucleopolyhedrovirus. J. Virol. 78, 7036–7051.

Geden, C.J., Lietze, V., Boucias, D.G., 2008. Seasonal prevalence and transmission of salivary gland hyperplasia virus of house flies, *Musca domestica* L. (Diptera: Muscidae). J. Med. Entomol. 45, 42–51.

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

Grainge, I., Jayaram, M., 1999. The integrase family of recombinase: organization and function of the active site. Mol. Microbiol. 33, 449–456.

Gutierrez, S., Kikhno, I., Ferber, M.L., 2004. Transcription and promoter analysis of pif, an essential but low-expressed baculovirus gene. J. Gen. Virol. 85, 331–341.

Guarino, L.A., Summers, M.D., 1986. Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. J. Virol. 60, 215–223.

Harris, J.M., McIntosh, E.M., Muscat, G.E., 1999. Structure/function analysis of a dUTPase: catalytic mechanism of a potential chemotherapeutic target. J. Mol. Biol. 288, 275–287.

Huang, Q.Q., Yao, S.Y., Ritzel, M.W., Paterson, A.R., Cass, C.E., Young, J.D., 1994. Cloning and functional expression of a complementary DNA encoding a mammalian nucleoside transport protein. J. Biol. Chem. 269, 17757–17760.

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

Jehle, J.A., Blissard, G.W., Bonning, B.C., Cory, J.S., Herniou, E.A., Rohrmann, G.F., Theilmann, D.A., Thiem, S.M., Vlak, J.M., 2006. On the classification and nomenclature of baculoviruses: a proposal for revision. Arch. Virol. 151, 1257–1266.

Johansson, K., Ramaswamy, S., Ljungcrantz, C., Knecht, W., Piskur, J., Munch-Petersen, B., Eriksson, S., Eklund, H., 2001. Structural basis for substrate specificities of cellular deoxyribonucleoside kinases. Nat. Struct. Biol. 8, 616–620.

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., 1989. Ultrastructural evidence for *trans*-ovum transmission of the DNA virus of tsetse *Glossina pallidipes* (Diptera: Glossinidae). Current Microb. 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.

Ko, E., Okano, K., Maeda, S., 2000. Structural and functional analysis of the *Xestia c-nigrum* granulovirus matrix metalloproteinase. J. Virol. 74, 11240–11246.
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.
Kokwaro, E.D., Otieno, L.H., Chimtawi, M., 1991. Salivary glands of the tsetse *Glossina pallidipes* Austen infected with *Trypanosoma brucei* and virus particles: an ultrastructural study. Insect Sci. Appl. 12, 661–669.
Kikhno, I., Gutierrez, S., Croizier, L., Croizier, G., Ferber, M.L., 2002. Characterization of pif, a gene required for the per os infectivity of *Spodoptera littoralis* nucleopolyhedrovirus. J. Gen. Virol. 83, 3013–3022.
Kuzio, J., Jaques, R., Faulkner, P., 1989. Identification of p74, a gene essential for virulence of baculovirus occlusion bodies. Virology 173, 759–763.
Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 681–685.
Lapointe, R., Popham, H.J., Straschil, U., Goulding, D., O'Reilly, D.R., Olszewski, J.A., 2004. Characterization of two *Autographa californica* nucleopolyhedrovirus proteins, Ac145 and Ac150, which affect oral infectivity in a host-dependent manner. J. Virol. 78, 6438–6448.
Lauzon, H.A., Lucarotti, C.J., Krell, P.J., Feng, Q., Retnakaran, A., Arif, B.M., 2004. Sequence and organization of the *Neodiprion lecontei* nucleopolyhedrovirus genome. J. Virol. 78, 7023–7035.
Lietze, V., Geden, C.J., Blackburn, P., Boucias, D.G., 2007. Effects of MdSGHV infection on the reproductive behavior of the house fly, *Musca domestica*. Appl. Environ. Microbiol. 73, 6811–6818.
Link, A., Eng, J.J., Schieltz, D.M., Carmack, E., Mize, E.J., Morris, D.R., Garvik, B.M., Yates, J.R., 1999. Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17, 676–682.
Maddison, D.R., Maddison, W.P., 2000. MacClade 4:0 analysis of phylogeny and character evolution, version 4.06. Sinauer Associates, Sunderland, Massachusetts, USA.
Maguire, T., Harrison, P., Hyink, O., Kalmakoff, J., Ward, V.K., 2000. The inhibitor of apoptosis of *Epiphyas postvittana* nucleopolyedrovirus. J. Gen. Virol. 81, 2803–2811.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.M., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.
Miller, L.K., 1999. An exegesis of IAPs: salvation and surprises from BIR motifs. Trends Cell Biol. 9, 323–328.
Miller, G.P., Benkovic, S.J., 1998. Stretching exercises — flexibility in dihydrofolate reductase catalysis. Chem. Biol. 5, R105–R113.
Minter-Goedbloed, E., Minter, D.M., 1989. Salivary gland hyperplasia and trypanosome infection of *Glossina* in two areas of Kenya. Trans. Royal 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.
Odindo, M.O., Payne, C.C., Crook, N.E., Jarrett, P., 1986. Properties of a novel DNA virus from the tsetse fly, *Glossina pallidipes*. J. Gen. Virol. 67, 527–536.
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, 15258–152764.
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.
Patel, S., Latterich, M., 1998. The AAA team: related ATPases with diverse functions. Trends Cell Biol. 8, 65–71.
Payne, C.C., 1974. The isolation and characterization of a virus from *Oryctes rhinoceros*. J. Gen. Virol. 25, 105–116.
Pearson, M., Bjornson, R., Pearson, G., Rohrmann, G., 1992. The *Autographa californica* baculovirus genome: evidence for multiple replication origins. Science 257, 1382–1384.
Perera, O., Green, T.B., Stevens, S.M., White, S., Becnel, J., 2007. Proteins associated with the *Culex nigripalpus* nucleopolyhedrovirus occluded virions. J. Virol. 81, 4585–4590.
Perryman, S.M., Rossana, C., Deng, T.L., Vanin, E.F., Johnson, L.F., 1986. Sequence of a cDNA for mouse thymidylate synthase reveals striking similarity with the prokaryotic enzyme. Mol. Biol. Evol. 3, 313–321.
Pijlman, G.P., Pruijsser, 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.
Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R., 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 33, W116–W120.
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.
Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.
Saurin, A.J., Borden, K.L., Boddy, M.N., Freemont, P.S., 1996. Does this have a familiar RING? Trends Biochem. Sci. 21, 208–214.
Shamoo, Y., Steitz, T.A., 1999. Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. Cell 99, 155–166.
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.M., Arif, B.M., 2007. The baculoviruses occlusion-derived virus: virion, structure and function. Adv. Virus Res. 69, 99–165.
Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland. Massachusetts, USA.
Tellam, R.L., Wijffels, G., Willadsen, P., 1999. Peritrophic matrix proteins. Insect Biochem. Mol. Biol. 29, 87–101.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.
Uren, A.G., Coulson, E.J., Vaux, D.L., 1998. Conservation of baculovirus inhibitor of apoptosis repeat proteins (BIRPs) in viruses, nematodes, vertebrates and yeasts. Trends Biochem. Sci. 23, 159–162.
Walker, J.E., Runswick, M.J., 1993. The mitochondrial transport protein superfamily. J. Bioenerg. Biomembranes 25, 435–446.
Wang, Y., Burand, J.P., Jehle, J.A., 2007a. Nudivirus genomics: diversity and classification. Virologica Sinica 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.
Whitnall, A.M.B., 1934. The trypanosome infections of *Glossina pallidipes* in Umfolosi game reserve, Zuhuland. Onderstepoot J. Anim. Ind. 11, 7–21.
Wilson, R., Goyal, L., Ditzel, M., Zachariou, A., Baker, D.A., Agapite, J., Seller, H., Meier, P., 2002. The DIAP1 RING finger mediates ubiquitination of *Dronc* and is indispensable for regulating apoptosis. Nat. Cell Biol. 4, 445–450.
Zhang, J.H., Ohkawa, T., Washburn, J.O., Volkman, L.E., 2005. Effects of Ac150 on virulence and pathogenesis of *Autographa californica* multiple nucleopolyhedrovirus in noctuid hosts. J. Gen. Virol. 86, 1619–1627.