

1 Submitted to the Journal of General Virology, May 18, 2010 VIR/2010/023671

2 Submitted in revised form August 16, 2010

3

4 **Proteomic analysis of *Glossina pallidipes* Salivary Gland Hypertrophy**

5 **Virus virions for immune intervention in tsetse fly colonies**

6

7 Henry M. Kariithi^{1,3 ‡}, Ikbal Agah Ince^{1,4 ‡}, Sjef Boeren², Jacques Vervoort²,

8 Max Bergoin⁵, Monique M. van Oers¹, Adly M. M. Abd-Alla³, Just M. Vlak^{1*}

9

10 ¹Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen,
11 The Netherlands, ²Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703 HA
12 Wageningen, The Netherlands, ³Insect Pest Control Laboratory, Joint FAO/IAEA Programme of
13 Nuclear Techniques in Food and Agriculture, Vienna, Austria ⁴Department of Biology, Faculty
14 of Arts and Sciences, Giresun University, 28049, Giresun, Turkey, ⁵Laboratoire de Pathologie
15 Comparée, Université Montpellier 2, Place Eugene Bataillon, 34095 Montpellier, France.

16

17 [‡]Authors contributed equally to this work

18 ^{*}Corresponding author: Tel: +31 317 483090, Fax: +31 317 484820, E-mail: just.vlak@wur.nl

19

20 **Running title:** Proteomics of GpSGHV

21 **Key words:** Tsetse; Salivary gland hypertrophy; *Glossina pallidipes*; Sterile insect technique;

22 Antibodies; Viral Proteins; Mass spectrometry; Proteome.

23 **Number of words:** Abstract: 231; Text: 4309

24 **Figures and Tables:** 5 figures, 2 tables and one supplementary table.

25 **Abstract:**

26 Many species of tsetse flies (Diptera: Glossinidae) can be infected by a virus that causes
27 salivary gland hypertrophy (SGH). The genomes of viruses isolated from *Glossina*
28 *pallidipes* (GpSGHV) and *Musca domestica* (MdSGHV) have recently been sequenced.
29 Tsetse flies with SGH have a reduced fecundity and fertility which cause a serious
30 problem for mass rearing in the frame of sterile insect technique (SIT) programs to
31 control and eradicate tsetse populations in the wild. A potential intervention strategy to
32 mitigate viral infections in fly colonies is neutralizing of the GpSGHV infection with
33 specific antibodies against virion proteins. Two major GpSGHV virion proteins of about
34 130 kDa and 50 kDa, respectively, were identified by Western analysis using polyclonal
35 rabbit antibody raised against whole GpSHGV virions. The proteome of GpSGHV,
36 containing the antigens responsible for the immune-response, was investigated by liquid
37 chromatography tandem mass spectrometry (LC-MS/MS) and 61 virion proteins were
38 identified by comparison with the genome sequence. Specific antibodies were produced
39 in rabbits against seven candidate proteins including the ORF10 / C-terminal fragment,
40 ORF47 and ORF96 as well as proteins involved in peroral infectivity PIF-1 (ORF102),
41 PIF-2 (ORF53), PIF-3 (ORF76) and P74 (ORF1). Antiserum against ORF10 specifically
42 reacted to the 130 kDa protein in a Western blot analysis and to the envelope of
43 GpSGHV using immunogold-EM. This result suggests that immune intervention of viral
44 infections in colonies of *G. pallidipes* is a realistic option.

45 **INTRODUCTION**

46 Tsetse flies (Diptera: Glossinidae) are the vectors of sleeping sickness in humans (or
47 human African trypanosomosis, HAT) and the livestock disease nagana (or African
48 animal trypanosomosis, AAT) (Steelman, 1976; WHO, 2001). The continuing presence
49 and advancement of tsetse flies prevents the development of sustainable and profitable
50 livestock production systems, thus greatly influencing food production, natural-resource
51 utilization and human settlement in almost two thirds of sub-Saharan Africa (Jordan,
52 1986). There are no effective vaccines against trypanosomosis and the disease is mainly
53 managed by the prophylactic and curative treatment with trypanocidal drugs. However,
54 there have been reports of development of resistance to the available trypanocidal drugs
55 (Aksoy & Rio, 2005). It is generally accepted that vector control remains the most
56 effective way of managing the disease and the sterile insect technique (SIT) that relies on
57 the sequential release of sterile male flies in the wild has proven to be a robust technique
58 for use in an area-wide integrated pest management (AW-IPM) approach (Hendrichs *et*
59 *al.*, 2007; Vreysen *et al.*, 2000). Upon mating of the sterile males with virgin female wild
60 flies there is no offspring which leads to a reduction in tsetse fly population density.
61 *Glossina austeni* has been successfully eradicated from the Island of Unguja, United
62 Republic of Tanzania using insecticide impregnated targets, insecticide “pour on” on
63 livestock and the release of sterile insects, and efforts are being made to do the same in
64 the Southern Rift Valley of Ethiopia (Feldmann, 2005).

65 Sterile males for AW-IPM programmes with an SIT component are produced in
66 mass rearing facilities and sterilized with ionizing radiation (usually ^{60}Co or ^{137}Ce).
67 However, the production of *Glossina pallidipes* flies is hampered by the fact that the fly

68 colonies are contaminated by a Salivary Gland Hypertrophy Virus (GpSGHV) (Ellis and
69 Maudlin, 1987; Jaenson, 1978 (Jura *et al.*, 1988; Jura *et al.*, 1989) which affects the
70 productivity and fecundity of these colonies. The low productivity of these colonies
71 makes the rearing very cumbersome and often leads to colony collapse.

72 GpSGHV is a rod-shaped, double-stranded, circular double-stranded DNA virus
73 with a genome of 190 kilobase (kb) pairs and averaging 70 x 640 nm in size (Fig. 1)
74 (Abd-Alla *et al.*, 2008; Odindo *et al.*, 1986). The presence of the virus in the salivary
75 glands of male and female flies explains the hypertrophied appearance (hyperplasia) of
76 these glands. The virus has also been associated with testicular degeneration and ovarian
77 abnormalities (Kokwaro *et al.*, 1990; Sang *et al.*, 1998; Sang *et al.*, 1999). Although it is
78 not yet clear how the fly gets infected with the virus and how exactly the virus affects the
79 mating and feeding behavior of the fly, it has been demonstrated recently that in mass
80 rearing facilities the virus is released from the infected fly with saliva upon blood
81 ingestion and transmitted through the blood to other flies (Abd-Alla *et al.*, 2010). In
82 nature, the virus is likely transmitted vertically from mother to offspring, either trans-
83 ovum or through infected milk glands (Jura *et al.*, 1989; Sang *et al.*, 1998; Sang *et al.*,
84 1996) and the infection is largely asymptomatic. Similar viruses have been described
85 from the bulb fly *Merodon equestris* (Amargier *et al.*, 1979) and the house fly *Musca*
86 *domestica* (Coler *et al.*, 1993).

87 The entire GpSGHV genome has been sequenced (Abd-Alla *et al.*, 2008) and 160
88 open reading frames (ORF) have been identified. Putative functions could be assigned to
89 only very few ORFs by blasting against databases. Most notable is the presence of
90 homologues of the *per os* infectivity factor (PIF) genes of baculoviruses. These factors

91 are involved in the oral transmission of baculoviruses from insect to insect. Their
92 presence in GpSGHV suggests a similar mechanism of infection, which is in fact
93 compatible with GpSGHV transmission via the blood meal.

94 One strategy to control the GpSGHV infections of *G. pallipides* flies in mass-
95 rearing facilities is to prevent horizontal GpSGHV transmission by immune-complexing
96 the virus in the blood meal and/or in the saliva. In this paper, we report the proteomic
97 analysis of the GpSGHV proteome and the identification of the immune-responsive,
98 virus-encoded proteins for development of antibodies to be used in immune intervention
99 in order to prevent SGHV infections in *G. pallidipes* colonies.

100

101 **RESULTS AND DISCUSSION**

102

103 **Analysis of the GpSGHV proteome**

104 The recent sequencing of the Uganda isolate of GpSGHV (Abd-Alla *et al.*, 2008; Abd-
105 Alla *et al.*, 2009) has provided information that greatly facilitated the assignment of
106 ORFs for virion proteins on the genome. The major disadvantage in our study is the lack
107 of a cell culture system for the production of GpSGHV virions which made it rather
108 difficult to purify a large amount of GpSGHV virions with high quality. GpSGHV virions
109 were therefore, purified from hypertrophied salivary gland cells dissected from infected
110 tsetse flies. In addition, to minimize the disadvantage of losing the envelope protein as
111 observed after purification of GpSGHV using sucrose gradients previously (Abd-Alla *et*
112 *al.*, 2007), the virus particles were purified over a nycodenz-gradient, resulting in only a
113 single viral band at a density of 1.153g cm^{-3} (Fig. 1A). The integrity of the virions after

114 Nycodenz purification was checked by the transmission electron microscopy, but they
115 appeared to be fragile (Fig. 1B and C). It is possible that some host proteins have been
116 co-purified, but also that they are intricately associated with virions, such as actin (Lanier
117 & Volkman, 1998; Wang *et al.*, 2010). Separation of the purified GpSGHV proteins by
118 gradient SDS-PAGE revealed at least 35 proteins ranging in size from 10 kDa to > 130
119 kDa (Fig. 2A). The most abundant proteins run at about 43 to 50 kDa (multiple bands)
120 and 130 kDa.

121 The gel lane was divided into 7 slices containing proteins with a molecular mass
122 ranging from lower than 26 kDa, 26-34 kDa, 34-43 kDa, 43-55 kDa, 55-95 kDa, 95-130
123 kDa to > 130 kDa, respectively (Fig. 2B) and protein extracts of each of the slices were
124 subjected to LC/MSMS analysis. A decoy database strategy was used (Elias & Gygi,
125 2007) which, after applying the appropriate filters, resulted in 90 protein hits: 61 viral
126 proteins, 28 contaminants and 1 decoy hit giving a False Discovery Rate of 1.1%. The
127 LC-MS/MS method allowed the detection of GpSGHV proteins that were present in
128 relatively low quantities. Fifty seven GpSGHV proteins were identified with 2 or more
129 peptides (Table 1). Manual verification of the 4 peptides with a single identified peptide
130 not only revealed a sufficient overlap between virtual and measured MS/MS spectra but
131 also showed that no other peptides present in the database can explain the measured
132 spectra.

133 The ORFs corresponding to 61 virion proteins were superposed on the physical
134 map of GpSGHV (Fig. 3). The virion protein genes were almost equally distributed over
135 both strands of the genome. Within the genome two segments encoding ORF11 to 26 and
136 ORF114 to 139 were devoid of virion protein genes. In contrast, other segments (ORF78

137 to 113) were densely populated with virion protein genes (22 out of 29 = 76%). Also
138 ORF62, a giant ORF was identified as encoding a constituent of GpSGHV virions
139 representing a protein of about 511 kDa. Such a large virion protein is not unusual for
140 large double stranded DNA viruses, as white spot syndrome virus has a 664 kDa virion
141 protein which is a major nucleocapsid protein (van Hulten *et al.*, 2001; Leu *et al.*, 2005).
142 Whereas the 511 kDa protein is probably a minor component (1.8% peptide coverage),
143 another large protein of 127 kDa (ORF10) was found in high abundance (with 16.9%
144 coverage) and probably represents the 130 kDa protein seen in SDS-PAGE (Fig. 2A).

145

146 **Gene homology and domains analyses**

147 Homology searches with the identified GpSGHV proteins performed against
148 GeneBank/EMBL, Swisprot, and PIR databases, revealed that of the 61 proteins
149 identified, 4 were unique to GpSGHV (encoded by ORFs 2, 39, 47 and 49). Neither the
150 nucleotide sequences nor the deduced protein sequences of these genes shared significant
151 homology to other genes/proteins identified so far. In addition, among the GpSGHV
152 proteins identified by LC-MS/MS, five proteins had homologs in baculoviruses and
153 nudiviruses (ORFs 1, 53, 76, 102 and 110) and one had homology to a nimavirus ORF
154 (ORF36) (Table 1).

155 Among the GpSGHV virion proteins identified (Table 1) there are homologs of
156 all four baculovirus *per os* infectivity factors (PIF-1, PIF-2, PIF-3 and PIF-0/P74). These
157 factors, encoded by ORFs 1, 53, 76 and 102, are thought to be involved in the oral
158 infection process of baculovirus occlusion derived virions by binding to midgut epithelial
159 cells (Slack & Arif, 2007). PIF-1, PIF-2, and P74 have been shown to mediate specific

160 binding of ODVs to midgut cells suggesting that they are directly involved in the virus
161 host cell interaction at an initial step in infection (Ohkawa *et al.*, 2005; Kikhno *et al.*,
162 2002; Slack *et al.*, 2010). Although PIF-3 appears to be an ODV-associated protein (Li *et*
163 *al.*, 2007), it does not appear to be involved in specific binding and its function is not
164 known yet. The PIF proteins are absolutely essential for baculovirus oral infectivity
165 (Kuzio *et al.*, 1989; Kikhno *et al.*, 2002; Pijlman *et al.*, 2003; Song *et al.*, 2008; Slack *et*
166 *al.*, 2010; Peng *et al.*, 2010). The presence of these PIF proteins in the GpSGHV
167 proteome is a strong indication that they might have a similar function in GpSGHV
168 infection following uptake via a blood meal.

169

170 **Comparative proteomics of GpSGHV and MdSGHV**

171 Recently a limited proteomic analysis has been carried out for an SGHV from the house
172 fly *Musca domestica* (MdSGHV), the genome of which is also entirely sequenced
173 (Garcia-Maruniak *et al.*, 2008). This virus is highly related to GpSGHV and is proposed
174 to be accommodated in a newly proposed virus family *Hytrosaviridae* (Garcia-Maruniak
175 *et al.*, 2009; Abd-Alla *et al.*, 2009). GpSGHV has 42 ORFs with homologs in MdSGHV
176 (Garcia-Maruniak *et al.*, 2009). Comparative analysis of the GpSGHV virion proteins
177 with all MdSGHV ORFs (Garcia-Maruniak *et al.*, 2008) showed that 33 out of the 61
178 identified GpSGHV virion proteins were homologous to 28 MdSGHV ORFs (Table 2).
179 The difference in ORF numbers (32 compared to 28) follows from the notion that four
180 MsSGHV ORFs have pairs of homologs in GpSGHV (see below). Of the 28 MdGSHV
181 ORFs homologous to the GpSGHV proteome ORFs, only 12 were actually identified in
182 the MdSGHV proteome (Garcia *et al.*, 2008).

183 The homology of the virion proteins and the presence of a number of proteins
184 shared between GpSGHV and MdSGHV virions suggests that the two viruses have
185 further properties in common, such as their virion structure and assembly mechanisms
186 and give further credence to the presence of a common ancestor of these two viruses
187 (Abd-Alla *et al.*, 2009). The notion that out of four putative SGHV *per os* infectivity
188 factors only P74 was identified in the MdSGHV virion, but not PIF-1, 2 and 3, may
189 suggest that some virion proteins may have been missed in the MdSGHV analysis
190 (Garcia-Maruniak *et al.*, 2008).

191 The comparative proteomics of GpSGHV and MdSGHV showed four instances of
192 gene duplications in GpSGHV (Table 2). GpSGHV ORF30 (133 aa) and ORF31 (285 aa)
193 are both homologous to MdSGHV ORF82 (359 aa). The GpSGHV ORFs ORF32/33,
194 ORF96/97 and ORF107/108 are homologous to MdSGHV ORFs 74, 25 and 33,
195 respectively. Gene duplication is one important mechanism by which large DNA viruses
196 increase their genome size and is a major mechanism for the acquisition of new gene
197 functions (Long *et al.*, 2001). ORF107 and ORF108 were shown before to be the likely
198 result of gene duplication (Garcia-Maruniak *et al.*, 2009). On the other hand, in 2 of the 4
199 cases (GpSGHV ORF30/31 and ORF32/33) the two GpSGHV ORFs in these pairs are
200 homologous to different parts of the corresponding MdSGHV ORF and show no
201 significant mutual homology within the pairs. Hence, these pairs are most likely the
202 consequence of extra stop codons in GpSGHV or, alternatively, the result of sequencing
203 artifacts. For ORF96 and ORF97 the situation is less clear as these ORFs show mutual
204 similarities over 276 amino acids, but only for 41% and only when several gaps are
205 introduced. In addition GpSGHV ORF110 has a homolog in ORF111 as a result of gene

206 duplication (Garcia-Maruniak *et al.*, 2009), but in the current analysis only peptides
207 derived from ORF110 were detected.

208

209 **Identification of immunodominant proteins**

210 To identify those GpSGHV virion proteins that give an immune response in rabbits,
211 antiserum was raised against purified GpSGHV virions and tested against GpSGHV
212 proteins in a Western analysis (Fig. 4). The antiserum reacted with several GpSGHV
213 proteins, more specifically to a 130 kDa protein, to proteins in the 50 kDa range and to an
214 array of proteins in the 30 kDa range. On the basis of the proteomic LC-MS/MS and the
215 Western analyses (Table 1 and Fig. 4) ORF10 (approximately 127 kDa) and six ORFs
216 with molecular sizes ranging between 43-50 kDa (ORFs 41, 47, 70, 96, 97 and 140) were
217 tested as candidates for the observed immunodominant proteins in GpSGHV (Fig. 4).
218 These ORFs were expressed in *E. coli* BL21 cells. Due to its large size, ORF10 was split
219 into two fragments (N- and C- terminal fragments with calculated molecular sizes of
220 approximately 69 kDa and 66 kDa, respectively). In addition, due to their assumed
221 important role in oral infectivity and potential target for immune intervention, all the four
222 PIF genes were selected as candidates for protein expression in bacteria. ORF1 (*p74*),
223 ORF41, ORF70, ORF97 and the N-terminal fragment of ORF10 could not be
224 successfully expressed in *E. coli*. Of the six proteins that were successfully made and that
225 reacted with the rabbit antiserum against GpSGHV virions (PIF-1, PIF-2, PIF-3, ORF47,
226 ORF96 and the C-terminal fragment of ORF10), two ORFs, ORF10 (C-terminal) and
227 ORF96, were selected to generate mono-specific polyclonal antibodies based on the fact
228 that these two ORFs were found to be the major viral proteins (Table 1). Antibodies

229 raised against the proteins encoded by ORF10 (C-terminal) and ORF96 were tested
230 against GpSGHV and homogenates of hypertrophied salivary glands of infected *G.*
231 *pallipides* flies (Fig. 5A). As expected these antibodies reacted against the 130 kDa and
232 50 kDa band, respectively.

233

234 **Immunolocalization of SGHV proteins**

235 Transmission electron microscopy (TEM) using the antibody against the C-terminal
236 fragment of ORF10 gave an indication that the protein is likely to be a component of the
237 envelope of the virus (Fig. 5B-b). Immunolocalization studies of ORF96 using the
238 specific rabbit antibody against this protein did not give conclusive evidence whether this
239 protein is part of the virus envelope as there were no gold particles observed on the
240 virions under the conditions used in the TEM studies. This could mean that either the
241 ORF96 protein could have hidden epitopes or the antiserum was not suited for immuno
242 EM. Furthermore it could be seen in the TEM that only a few GpSGHV virion rods
243 remained entirely intact after Nycodenz-preparation. This confirms the fragile nature of
244 GpSGHV virus and further work need to be done to study its stability under different
245 conditions such as temperature, and especially the effects of virus handling on its
246 infectivity to the tsetse fly. It is to be noted that a high density of gold particles were
247 observed when antiserum against ORF10 was used, most likely on debris of the viral
248 envelope. In the control experiments (preimmune serum), no gold particles could be seen
249 for either the GpSGHV virion or the nucleocapsid. These studies may be direct evidence
250 that the ORF10 could be involved functionally in the formation and/or the assembly of
251 the GpSGHV envelope.

252

253 **CONCLUDING REMARKS**

254 Current proteomic analysis of GpSGHV allowed us to determine a total of 61 proteins.

255 The identities of the proteins within the virion proteome revealed many candidates which

256 provide a basis for further studies focussing on the virulence and pathogenesis of

257 GpSGHV as well as on mechanisms of virus infection and transmission in tsetse flies.

258 Comparison of the 61 identified GpSGHV ORFs with double-stranded DNA viruses of

259 other virus families showed only a few homologies (5), more specifically with

260 baculoviruses and nudiviruses. These involve the PIF proteins that are essential for oral

261 infectivity of baculoviruses in insects. Whether they are also functional PIF proteins,

262 remains to be investigated. The proteomic data also clearly indicated that GpSGHV has a

263 total of 32 structural ORFs that encode proteins not encoded in MdSGHV. Twenty-nine

264 GpSGHV ORFs have homologs in MdSGHV, and so far twelve of these have also been

265 identified in the MdSGHV proteome. In addition MdSGHV has a number of virion

266 proteins that do not have homologs in GpSGHV. Therefore, this analysis further supports

267 the placement of these two members of the newly-proposed *Hytrosaviridae* family in two

268 separate genera (proposed names *Glossinavirus* and *Muscavirus*) (Abd-Alla *et al.*, 2009).

269 The role of these virion proteins in virion structure and infectivity will be the subject of

270 future investigations.

271 Proteomic and immunolocalization data indicated that the ORF10 protein is

272 abundant and present on the virion envelope. In the light of these findings, the ORF10

273 protein would be a good target for studies to mitigate infections of tsetse colonies by the

274 SGHV. The approach would be to supplement the blood meal with the ORF10 antibody
275 or by immunizing cattle who are the blood donors in tsetse fly rearing systems.

276

277 **MATERIAL AND METHODS**

278

279 **Preparation of virus particles and analysis by LC-MS/MS**

280 GpSGHV virions were purified from hypertrophied salivary glands collected from
281 a *G. pallidipes* colony maintained at the Entomology Unit of the FAO/IAEA Agriculture
282 & Biotechnology Laboratory, Seibersdorf, Austria. Hypertrophied salivary glands were
283 collected from dissected flies, homogenized in Tris buffer (50 mM, pH 7.8), and clarified
284 twice by centrifugation for 10 min at 3000 g. The supernatant was layered on a 10-60%
285 linear Nycodenz gradient and centrifuged for 1h at 27,000 g. The viral band was taken
286 and washed in Tris buffer and centrifuged for 1h at 150,000 g. The viral pellet was
287 resuspended in Tris buffer.

288 The purified virus particles were solubilized in 2x concentrated Laemmli buffer,
289 and fractionated by SDS-PAGE (12%). Fermentas PageRuler Prestained Marker proteins
290 were used. The gel was stained with colloidal blue and the gel lane containing the virion
291 proteins was excised into seven contiguous sections spanning the complete gel lane based
292 on a comparison with molecular markers.

293 In-gel protein digestions and peptide extractions were performed at 25°C according to a
294 method described (Ince *et al.*, 2010). The peptides resulting from this digestion were
295 analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) by injecting
296 18 µl of sample on a 0.10 x 32-mm Prontosil 300-5-C18H pre-concentration column

297 (Bischoff, Germany) at a flow of 6 μ l/min for 5 min. Peptides were eluted from the pre-
298 concentration column onto a 0.10 x 200-mm Prontosil analytical column 300-3-C18H
299 (Bischoff) with an acetonitrile (ACN; HPLC grade) gradient at a flow of 0.5 μ l/min for
300 50 min. The gradient consisted of 10 to 35% (v/v) acetonitril increased in water with
301 1 ml/L formic acid in 50 min. As a subsequent cleaning step, the ACN concentration was
302 increased to 80% (v/v) in 3 min (with 20% water and 1 ml/L formic acid in both the ACN
303 and the water). Between the pre-concentration and analytical column, an electrospray
304 potential of 3.5 kV was applied directly to the eluent via a solid 0.5 mm platina electrode
305 fitted into a P875 Upchurch microT. Full scan positive mode Fourier transform mass
306 spectra (FTMS) were measured between mass-to-charge ratios of 380 and 1400 with a
307 LTQ-Orbitrap spectrometer (Thermo electron, San Jose, CA, USA).

308 Tandem mass spectrometry (MS/MS) scans of the four most abundant doubly and
309 triply charged peaks in the FTMS scan were recorded in a data dependent mode in the
310 linear trap (MS/MS threshold = 10.000). All MS/MS spectra obtained with each run
311 were analyzed with Bioworks 3.1.1 software (Thermo Fisher Scientific, Inc.). A
312 maximum of a single differential modification allowed per peptide was set for oxidation
313 of methionines and de-amidation of asparagine and glutamine residues.
314 Carboxamidomethylation of cysteines was set as a fixed modification. Trypsin specificity
315 was set to fully enzymatic and a maximum of 3 missed cleavages with monoisotopic
316 precursor and fragment ions. The mass tolerance for peptide precursor ions was set to 10
317 parts per million and for MS/MS fragment ions to 0.5 Da.

318 A *Glossina pallidipes* salivary gland hypertrophy virus protein database was used
319 for the analysis (EF568108; created February 25, 2008; downloaded from

320 www.ncbi.nlm.nih.gov/sites/entrez) after adding a list of commonly observed
321 contaminants like: BSA (P02769, bovine serum albumin precursor), trypsin (P00760,
322 bovine), trypsin (P00761, porcine), keratin K22E (P35908, human), keratin K1C9
323 (P35527, human), keratin K2C1 (P04264, human) and keratin K1CI (P35527, human). A
324 decoy database was created by adding the reversed sequences using the program
325 SequenceReverser from the MaxQuant package (Cox *et al.*, 2008).

326 To identify the proteins in the GpSGHV virions, the spectra obtained from the
327 LC-MS/MS were searched against the GpSGHV ORF database using Bioworks 3.3.1.
328 The peptide identifications obtained were filtered in Bioworks with the following filter
329 criteria: $\Delta Cn > 0.08$, $Xcorr > 1.5$ for charge state 2+, $Xcorr > 3.3$ for charge state 3+ and
330 $Xcorr > 3.5$ for charge state 4+ (Peng *et al.*, 2003). Only those proteins that showed a
331 Bioworks Score factor (Sf) larger than 0.9 were considered.

332

333 **Detection of immunodominant protein candidates**

334 Following separation by 12% SDS-PAGE, proteins of purified GpSGHV particles were
335 transferred to Immobilon-P membranes. Membranes were blocked overnight by
336 incubation with 1% low fat milk and 0.05% Tween-20 in PBS (137 mM NaCl, 2.7 mM
337 KCl, 10 mM Na_2HPO_4 , and 1.76 mM KH_2PO_4 pH 7.4) at room temperature. Membranes
338 were washed once with 0.2% low fat milk in PBS-tween-20 for 5 min, incubated with
339 rabbit primary antibody (anti-GpSGHV; diluted 1/500, see below) at room temperature
340 for 30 min, washed three times with PBS-Tween-20, and further incubated for 30 min in
341 goat anti-rabbit IgG-Alkaline phosphatase conjugate (Promega) diluted 1:3000. Blots
342 were washed three times with alkaline phosphatase buffer (0.1 M Tris-HCl, 5mM MgCl_2 ,

343 pH 9.5) and stained with 1% of Nitro-Blue Tetrazolium Chloride/5-Bromo-4-Chloro-3'-
344 Indolyphosphate p-Toluidine Salt (NBT/BCIP) in alkaline phosphatase buffer.

345

346 **Selection of open reading frames, PCR amplification and gene cloning**

347 Candidate open reading frames (ORFs) for protein expression were selected based on the
348 molecular masses of LC-MS/MS identified proteins and the immunoblot analyses. Viral
349 DNA was extracted from purified virus as previously reported (Abd-Alla *et al.*, 2007),
350 and approximately 5 ng of the DNA was used as template. PCR amplifications were
351 performed with HF Phusion *Taq* DNA polymerase (Finnzymes), using the reaction
352 mixture recommended by the supplier. The primers were designed to amplify the
353 hydrophilic regions of the selected ORFs (Supplementary Table S1) and were used at a
354 final concentration of 0.2 mM. The PCR conditions were 98°C for 30 sec; 98°C for 10s,
355 59°C for 20s, and 72°C for 30 s/kbp for 25 to 30 cycles; and finally 72°C for 5 min. The
356 PCR products were individually inserted into pJET1.2/blunt cloning vector (CloneJET™
357 PCR Cloning kit, Fermentas). The resulting recombinant plasmids were purified with
358 homemade GF/F columns as described in Borodina *et al.*, (2003) and the inserts were
359 sequenced to confirm the sequences. The inserted DNA fragments were re-cloned into
360 pET28a(+) (Sambrook *et al.*, 1989) at the multiple cloning site.

361

362

363 **GpSGHV proteins: expression, purification and production of antisera**

364 *Escherichia coli* BL21 (DE3) cells were transformed with the pET28-derived plasmids to
365 express the cloned genes according to the pET system manual (Novagen). The bacteria

366 expressing the viral genes were sonicated in Laemmli sample buffer (Laemmli, 1970) and
367 purified using preparative SDS-PAGE (Model 491 Prep Cell, Bio-Rad Laboratories)
368 according to manufacturer's instructions. Ten μ l of each recombinant protein was
369 analyzed on 12 % SDS-PAGE followed by silver staining according to standard
370 protocols. The purity and quantity were verified with Coomassie Blue staining and with
371 Western blotting, using specific immune serum directed against His-tag. To reduce the
372 amount of SDS in the samples, each protein fraction was concentrated with
373 CENTRIPREP[®] YM-10 centrifugal membranes (Amicon bioseparations).

374 Antisera were prepared against the purified proteins (proteins encoded by ORF96
375 and the C-terminal fragment of ORF10) by injecting rabbits with 0.4-0.8 mg of the
376 recombinant protein emulsified in Freund incomplete adjuvant. Two booster injections
377 were given at 2-week intervals. Another two antisera were prepared against the P74
378 protein (ORF1) using the synthetic oligopeptides LYEHSKDEDGVYHRA-C (amino
379 acids 114 to 128) and C-SEENKIASIDDKEQF (amino acids 612-626) (Pacific
380 Immunology Corp, Ramona, CA 92065). Polyclonal antibody against the whole
381 GpSGHV particles was collected from rabbit used for several months to feed tsetse flies
382 in CIRAD, France.

383

384 **Electron microscopy and immunolocalization of authentic viral proteins**

385 Aliquots (5 μ l) of GpSGHV virion suspension were adsorbed to carbon-coated and
386 ionized nickel grids (400 mesh) for 5 min at room temperature and treated for negative
387 staining with phosphotungstic acid or for immunogold labeling. For the latter the grids
388 were then blocked with blocking buffer (5% bovine serum albumin, 5% normal serum,

389 0.1% cold water skin gelatin, 10 mM phosphate buffer, 150 mM NaCl, pH 7.4) for 30
390 min, and incubated with primary antibody or pre-immune rabbit serum (1:20 dilution in
391 incubation buffer) for 1.5 h at room temperature. After incubation, and after several
392 washes, the grids were incubated with goat anti-rabbit secondary antibody conjugated
393 with gold particles (10 nm-diameter; 1:20 dilution in incubation buffer) for 45 min at
394 room temperature. The grids were washed extensively with incubation buffer to remove
395 excess salt, and negatively stained with 2% sodium phosphotungstate, (pH 6.5) for 5-10
396 sec. The specimens examined with a transmission electron microscope (Jeol, JEM-1011,
397 100 kV EM).

398

399

400 **ACKNOWLEDGMENTS**

401 The authors would like to thank Wageningen University and Research Centre, The
402 Netherlands for awarding a Master of Science grant to Mr. Henry M. Kariithi to carry out
403 these studies. We are indebted to the Entomology Unit, FAO/IAEA Agriculture and
404 Biotechnology Laboratories, Seibersdorf Austria for providing the virus used in this
405 research. All proteomic LC-MS/MS measurements were done at Biqualyt Wageningen
406 (www.biqualyt.nl). We would like to thank Marc Vreysen for reviewing the manuscript.

407

408 **References**

409 **Abd-Alla, A., Bossin, H., Cousserans, F., Parker, A., Bergoin, M. & Robinson, A.**
410 **(2007).** Development of a non-destructive PCR method for detection of the salivary
411 gland hypertrophy virus (SGHV) in tsetse flies. *J Virol Methods* **139**, 143-149.

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

416 **Abd-Alla, A. M. M., Kariithi, H., Parker, A. G., Robinson, A. S., Kiflom, M.,**
417 **Bergoin, M. & Vreysen, M. J. B. (2010).** Dynamics of the salivary gland
418 hypertrophy virus in laboratory colonies of *Glossina pallidipes* (Diptera:
419 Glossinidae). *Virus Res* **150**, 103-110.

420 **Abd-Alla, A. M. M., Vlak, J. M., Bergoin, M., Maruniak, J. E., Parker, A. G.,**
421 **Burand, J. P., Jehle, J.A. & Boucias, D.G. 2009.** Hytrosaviridae: a proposal for
422 classification and nomenclature of a new insect virus family. *Arch Virol* **154**, 909-
423 918.

424 **Aksoy, S. & Rio, R. V. M. (2005).** Interactions among multiple genomes: Tsetse, its
425 symbionts and trypanosomes. *Insect Biochem Mol Biol* **35**, 691-698.

426 **Amargier, A., Lyon, J. P., Vago, C., Meynadier, G. & Veyrunes, J. C. (1979).**
427 Discovery and purification of a virus in gland hyperplasia of insects. Study of
428 *Merodon equestris* F. (Diptera, Syrphidae). *C R Acad Sci D* **289**, 481-484.

429 **Borodina, T. A., Lehrach, H. & Soldatov, A. V. (2003).** DNA purification on
430 homemade silica spin-columns. *Anal Biochem* **321**, 135-137.

431 **Coler, R., Boucias, D., Frank, J., Maruniak, J., Garcia-Canedo, A. & Pendland, J.**
432 **(1993).** Characterization and description of a virus causing salivary gland hyperplasia
433 in the housefly, *Musca domestica*. *Med Vet Entomol* **7**, 275-282.

434 **Cox, J. & Mann, M. (2008).** MaxQuant enables high peptide identification rates,
435 individualized p.p.b.-range mass accuracies and proteome-wide protein
436 quantification. *Nat Biotechnol* **26**, 1367-1372.

437 **Elias, J. E. & Gygi, S. P. (2007).** Target-decoy search strategy for increased confidence
438 in large-scale protein identifications by mass spectrometry. *Nat Meth* **4** 207–214.

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

441 **Feldmann, U. (1994).** Guidelines for the rearing of tsetse flies using the membrane
442 feeding technique. In *Techniques of insect rearing for the development of integrated*
443 *pest and vector management strategies*, pp. 449-471. Edited by J. P. R. Ochieng'-
444 Odero. Nairobi, Kenya: ICIPE Science Press.

445 **Feldmann, U. (2005).** The sterile insect technique as a component of area-wide
446 integrated pest management of tsetse. In *The Trypanosomiases*, pp. 565-582. Edited
447 by I. Maudlin, P. H. Holmes & M. A. Wallingford, UK: CABI Publishing.

448 **Garcia-Maruniak, A., Abd-Alla, A. M. M., Salem, T. Z., Parker, A. G., van Oers, M.**
449 **M., Maruniak, J. E., Kim, W., Burand, J. P., Cousserans, F., Robinson, A. S.,**
450 **Vlak, J. M., Bergoin, M. & Boucias, D. G. (2009).** Two viruses that cause salivary
451 gland hypertrophy in *Glossina pallidipes* and *Musca domestica* are related and form a
452 distinct phylogenetic clade. *J Gen Virol* **90**, 334-346.

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

- 456 **Gooding, R. H., Feldmann, U. & Robinson, A. S. (1997).** Care and maintenance of
457 tsetse colonies. In *The molecular biology of insect disease vectors: a methods*
458 *manual*, pp. 41-55. Edited by J. M. Crampton, C.B. Beard, & C. Louis. London, UK:
459 Chapman & Hall Ltd.
- 460 **Hendrichs, J. P., Kenmore, P., Robinson, A. S. & Vreysen, M. J. B. (2007).** Area-
461 wide integrated pest management (AW-IPM): Principles, practice and prospects, pp.
462 3-33. In *Area-wide Control of Insect Pests: From Research to Field Implementation*.
463 Edited by M. J. B. Vreysen, A. S. Robinson & J. Hendrichs. Dordrecht, the
464 Netherlands: Springer,
- 465 **Ince, A. I., Boeren, S. , van Oers, M. M., Vervoort, J. J. M. & Vlak, J. M. (2010).**
466 Proteomic analysis of *Chilo iridescent virus*. *Virology* **405**, 253-258.
- 467 **Jaenson, T. G. T. (1978).** Virus-like rods associated with salivary gland hyperplasia in
468 tsetse, *Glossina pallidipes*. *Trans R Soc Trop Med Hyg* **72**, 234-238.
- 469 **Jordan, A. M. (1986).** Trypanosomiasis control and African rural development.
470 Longman, London (UK).
- 471 **Jura, W. G. Z. O., Odhiambo, T. R., Otieno, L. H. & Tabu, N. O. (1988).** Gonadal
472 lesions in virus-infected male and female tsetse, *Glossina pallidipes* (Diptera:
473 Glossinidae). *J Invertebr Pathol* **52**, 1-8.
- 474 **Jura, W. G. Z. O., Otieno, L. H. & Chimtawi, M .M. B. (1989).** Ultrastructural
475 evidence for trans-ovum transmission of the DNA virus of tsetse, *Glossina pallidipes*
476 (Diptera: Glossinidae). *Curr Microbiol* **18**, 1-4.

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

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

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

485 **Laemmli, K.U. (1970).** Cleavage of structural proteins during the assembly of the head
486 of bacteriophage T4. *Nature* **227**, 680-685.

487 **Lanier, L. M., & Volkman, L. E. 1998.** Actin binding and nucleation by *Autographa*
488 *californica* M nucleopolyhedrovirus. *Virology* **243**,167-177.

489 **Leu, J.-H., Tsai, J.-M., Wang, H.-C., Wang, A. H. J., Wang, C.-H., Kou, G.-H. & Lo,**
490 **C. F. (2005).** The unique stacked rings in the nucleocapsid of the white spot
491 syndrome virus virion are formed by the major structural protein VP664, the largest
492 viral structural protein ever found. *J Virol* **79**, 140-149.

493 **Li, X., Song, J., Jiang, T., Liang, C. & Chen, X. (2007).** The N-terminal hydrophobic
494 sequence of *Autographa californica* nucleopolyhedrovirus PIF-3 is essential for oral
495 infection. *Arch Virol* **152**, 1851-1858.

496 **Long, M., Thornton, K., Zhang, L., Gaut, B. S., Vision, T. J., Lynch, M. & Conery,**
497 **J. C. (2001).** Gene duplication and evolution. *Science* **293**, 1551a.

498 **Odindo, M. O., Payne, C. C., Crook, N. E. & Jarret, P. (1986).** Properties of a novel
499 DNA virus from the tsetse fly, *Glossina pallidipes*. *J Gen Virol* **67**, 527-536.

500 **Ohkawa, T., Washburn, J. O., Sitapara, R., Sid, E. & Volkman, L. E. (2005).**
501 Specific binding of *Autographa californica* M nucleopolyhedrovirus occlusion-
502 derived virus to midgut cells of *Heliothis virescens* larvae is mediated by products of
503 pif genes Ac119 and Ac022 but not by Ac115. *J Virol* **79**, 15258-15264.

504 **Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. & Gygi, S. P. (2003).** Evaluation
505 of multidimensional chromatography coupled with tandem mass spectrometry
506 spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome.
507 *J Proteome Res* **2**, 43-50.

508 **Peng, K., Van Oers, M. M., Hu, Z. H., Van Lent, J. W. M. & Vlak, J. M. (2010).**
509 Baculovirus *per os* infectivity factors form a complex on the surface of occlusion-
510 derived virus. *J Virol*, doi:10.1128/JVI.00812-10

511 **Pijlman, G. P., Pruijssers, A. J. & Vlak, J. M. (2003).** Identification of pif-2, a third
512 conserved baculovirus gene required for *per os* infection of insects. *J Gen Virol* **84**,
513 2041-2049.

514 **Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989).** Molecular cloning: a laboratory
515 manual, 2 ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

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

520 **Sang, R. C., Jura, W. G. Z. O., Otieno, L. H., Mwangi, R. W. & Ogaja, P. (1999).**
521 The effects of a tsetse DNA virus infection on the functions of the male accessory

522 reproductive gland in the host fly *Glossina morsitans centralis* (Diptera; Glossinidae).
523 *Curr Microbiol* **38**, 349-354.

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

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

529 **Slack, J. M., Lawrence, S. D., Krell, P. J. & Arif, B. M. (2010).** A soluble form of P74
530 can act as a per os infectivity factor to to the *Autographa californica* multiple
531 nucleopolyhedrovirus. *J Gen Virol* **91**, 915 - 918

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

535 **Steelman, C. D. (1976).** Effects of external and internal arthropod parasites on domestic
536 livestock production. *Annu Rev Entomol* **21**, 155-178.

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

540 **Vreysen, M. J. B., Saleh, K. M., Ali, M. Y., Abdulla, A. M., Zhu, Z.-R., Juma, K. G.,**
541 **Dyck, V. A., Msangi, A. R., Mkonyi, P. A. & Feldmann, H. U. (2000).** *Glossina*
542 *austeni* (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the
543 sterile insect technique. *J Econ Entomol* **93**, 123-135.

- 544 **Wang, R., Deng, F., Houm D., Zhao, Y., Gua, L., Wang, H. & Hu, Z. H. (2010).**
545 Proteomics of the *Autographa californica* nucleopolyhedrovirus budded virions. *J*
546 *Virolog* **84**, 7233-7242.
- 547 **WHO (2001).** Report on African trypanosomiasis (sleeping sickness), WHO/TDR
548 Committee Report TDR/SWG/01. World Health Organization, Geneva, Switzerland.
- 549 **Zelger, R. & Russ, K. (1976).** Untersuchungen über die mechanische Trennung von
550 Männchen und Weibchen der Kirschfruchtfliege, *Rhagoletis cerasi* L., (Diptera:
551 Tephritidae), im Puppenstadium. *Z Angew Zool* **63**, 257-266.

552 **Figure legends**

553

554 **Figure 1:** Purification of GpSGHV virions used for mass spectrometry analysis. Virions
555 from hypertrophied salivary glands were purified by Nycodenz gradient centrifugation (A);
556 the arrow indicates a single distinct band that was selected after centrifugation for 1 h at
557 150,000 x g. Micrographs show a negatively-stained ultra-thin section (B) and a suspension
558 of the isolated GpSGHV (C).

559

560 **Figure 2:** SDS-PAGE analysis of Nycodenz-purified GpSGHV stained with Coomassie
561 blue (A) and the relative locations of the 61 SGHV-encoded virion proteins identified by
562 LC-MS-MS analysis (B). Adjacent numbers represent the different open reading frames
563 (ORFs), that were detected in the seven gel sections (i to vii). The numbers in the
564 parentheses represent the number of unique peptides used to identify the designated ORFs.

565

566 **Figure 3:** Positioning of the 61 virion (structural) proteins encoded by GpSGHV on the
567 genomic map of GpSGHV (Abd-Alla *et al.*, 2008). The arrows indicate the positions and
568 the direction of transcription for the ORFs.

569

570 **Figure 4:** Western blot analysis of purified GpSGHV using rabbit antiserum against the
571 whole GpSGHV virion (dilution of 1: 500). The arrows indicate two bands (approximately
572 130 kDa and 50 kDa) that reacted strongly with the rabbit antiserum. Budded virus (BV)
573 and occlusion derived virus (ODV) particles from the baculovirus *Autographa californica*
574 multiple nucleopolyhedrovirus (AcMNPV) were used as negative control.

575

576 **Figure 5:** Detection of ORF10 and ORF96 with specific antibodies in SGHV virions.
577 A. Western blot analysis with specific rabbit antibodies against proteins encoded by
578 GpSGHV ORF96 and ORF10 (diluted 1: 500). Shown in the figure are Nycodenz-purified
579 GpSGHV virions (lanes 1), a homogenate of hypertrophied (lanes 2), and non-infected
580 (lanes 3) tsetse fly salivary gland cells. B. Electron micrographs of GpSGHV virions using
581 pre-immune rabbit antiserum (A) and rabbit antiserum against the C-terminal fragment of
582 the ORF10 (B). The secondary antibody was goat anti-rabbit IgG conjugated with 10 nm of
583 gold (Aurion).
584

Table 1: The GpSGHV virion proteins identified by mass spectrometry in order of descending molecular mass (511 to 10 kDa)*.

ORF	GeneBank Accession No.	Predicted mol. mass in kDa	Protein coverage in % amino acids	No. of peptides identified	pI	Characteristics of deduced proteins	Conserved domain(s)
62	YP_001687010	511.81	1.8	7	4.70		
45	YP_001686993	200.98	4.1	7	5.94		
38	YP_001686986	136.46	42.20	130	5.91		TM
10	YP_001686958	126.96	81	305	5.60	Major protein	Coiled-coil region
40	YP_001686988	104.065	3.9	3	9.16		
83	YP_001687031	81.54	13	9	5.59	(Md013)	
1	YP_001686949	81.35	13.8	12	4.73	p74 (<i>bac</i>) (Md039)	TM
104	YP_001687052	77.84	7	6	7.30		TM, coiled-coil region
88	YP_001687036	77.71	15.3	11	7.36		
102	YP_001687050	76.07	12.4	9	5.02	PIF-1 (<i>bac</i>)	TM, SP
71	YP_001687019	71.95	2	1	6.66	(Md090)	
86	YP_001687034	70.13	29.4	28	7.76		
64	YP_001687012	69.98	43.7	42	8.95	ATPase/Helicase (Md097)	
108	YP_001687056	63.90	30.5	15	5.34	Cell division 48-like protein (Md033)	
46	YP_001686994	61.50	37.7	24	6.48	(Md084)	
107	YP_001687055	59.53	45.1	31	5.39	Cell division 48-like protein (Md33)	
61	YP_001687009	57.40	38.3	14	8.68		
106	YP_001687054	55.05	31.3	15	5.31		Coiled-coil region
27	YP_001686975	53.03	5.5	3	6.12	Chitinase Chit1 precursor-like protein	TM
70	YP_001687018	50.89	23.9	13	9.44		
41	YP_001686989	48.74	24.9	15	6.78		
140	YP_001687088	48.41	10	5	9.26		Coiled-coil region
47	YP_001686995	47.16	61.9	98	4.47		
97	YP_001687045	44.37	62.9	83	8.61	(Md025)	TM
96	YP_001687044	43.50	77.2	170	6.57	Major protein (Md025)	TM, SP
44	YP_001686992	42.81	20.3	8	8.84		SP
34	YP_001686982	41.15	7.7	3	9.13		
7	YP_001686955	41.02	15.3	5	9.14		
33	YP_001686981	40.98	10.3	3	9.51		
53	YP_001687001	40.21	27.8	11	5.05	PIF-2 (<i>bac</i>)	
154	YP_001687102	40.07	8	4	7.50	(Md071)	
2	YP_001686950	38.64	31.8	17	8.99		
93	YP_001687041	38.51	57.8	39	9.53	(Md022)	TM
39	YP_001686987	37.64	4.8	1	4.47		TM, SP
52	YP_001687000	36.67	6.5	7	8.63		
105	YP_001687053	34.75	7.9	2	9.82		Coiled-coil region
31	YP_001686979	33.58	19.3	5	9.62		
113	YP_001687061	33.10	34.5	11	8.76		
50	YP_001686998	32.72	29.2	45	6.33	(Md086)	
94	YP_001687042	32.68	57.5	34	9.28		
72	YP_001687020	31.75	27.9	7	8.32	(Md102)	TM
8	YP_001686956	31.58	31.4	9	9.61		
91	YP_001687039	31.47	7.1	2	7.96		TM, SP
67	YP_001687015	30.99	47.1	30	4.55		
69	YP_001687017	30.89	41.9	49	7.41		TM, SP
32	YP_001686980	30.08	13.9	3	4.90		

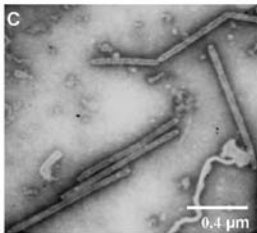
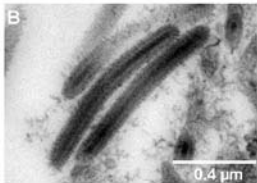
ORF	GeneBank Accession No.	Predicted mol. mass in kDa	Protein coverage in % amino acids	No. of peptides identified	pI	Characteristics of deduced proteins	Conserved domain(s)
85	YP_001687033	30.07	40	15	9.29		
78	YP_001687026	27.83	9.3	3	7.93		TM
146	YP_001687094	26.67	9	2	9.50		
84	YP_001687032	25.75	28.8	8	9.37		
76	YP_001687024	24.39	9.5	2	4.34	PIF-3 (<i>bac</i>)	TM
110	YP_001687058	23.62	27.9	4	7.71	mp-nase (<i>bac</i>)	TM
112	YP_001687060	19.05	19.2	8	4.64		TM
82	YP_001687030	18.67	23.3	3	9.84		
43	YP_001686991	16.89	69.4	28	9.91	(<i>Md072</i>)	TM
30	YP_001686978	15.91	10.5	1	10.17		
36	YP_001686984	13.79	12.2	2	5.77	Thymidylate synthase (<i>nima</i>)	
98	YP_001687046	13.52	51.3	10	10.67		
68	YP_001687016	12.64	37	6	7.63		
101	YP_001687049	12.33	62.3	37	9.56		TM, SP
49	YP_001686997	10.00	12.2	1	8.17		

* pI is isoelectric point, SP is signal peptide, TM is transmembrane domain; homologs with MdSGHV (*Md*), baculovirus (*bac*) and nimavirus (*nima*) and ascoviruses are indicated.

Table 2: Proteins represented in the proteome of GpSGHV with homolog genes in MdSGHV genome. The MdSGHV homologs that were also identified in the MdSGHV proteome are in bold*.

GpSGHV ORF		MdSGHV ORF		GpSGHV ORF		MdSGHV ORF	
ORF	Length	ORF	Length	ORF	Length	ORF	Length
1 (<i>p74</i>)	696	<u>39</u> (<i>p74</i>)	707	72	269	<u>102</u>	257
30	133			76 (<i>pif-3</i>)	211	106 (<i>pif-3</i>)	174
31	285	82	359	78	236	108	205
32	259			82	159	4	131
33	348	74	698	83	694	<u>13</u>	644
40	901	70	967	86	1779	16	1500
41	413	55	416	88	652	17	553
43	144	<u>72</u>	136	93	329	<u>22</u>	343
44	359	73	390	96	381		
45	1728	83	1780	97	394	<u>25</u>	376
46	533	<u>84</u>	509	102 (<i>pif-1</i>)	652	29 (<i>pif-1</i>)	644
50	291	<u>86</u>	381	104	660	30	692
53 (<i>pif-2</i>)	360	89 (<i>pif-2</i>)	379	107	521		
61	494	100	455	108	545	<u>33</u>	497
64	595	<u>97</u>	595	110	201	36	196
68	327	<u>94</u>	336	154	338	<u>71</u>	333
71	608	<u>90</u>	672				

* MdSGHV proteome data (Garcia et al., 2008; ibid, 2009). Four GpSGHV ORFs share a homolog in MdSGHV. ORF length is given in amino acids.



A**B****M****GpSGHV**

kDa

170

130

95

72

55

43

34

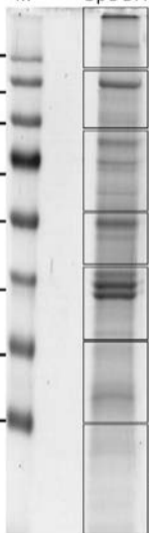
26

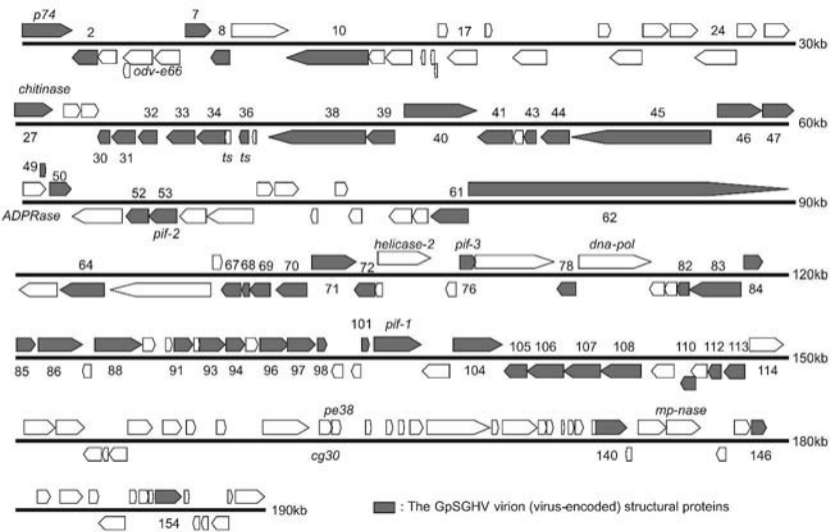
i

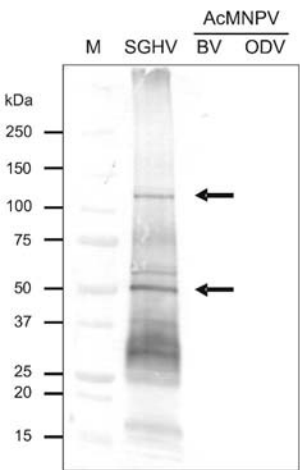
62(7); 38(130); 45(7)

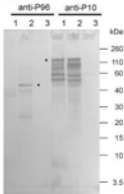
ii

10(305); 40(3)

iii1(12); 46(24); 61(14); 64(42); 71(1);
83(9); 86(28); 88(11); 102(9);
104(6); 106(15); 107(31); 108(15)**iv**27(3); 41(15); 47(98); 70(13);
96(170); 97(83); 140(5)**v**2(17); 7(5); 33(3); 34(3); 39(1);
44(8); 52(7); 53(11); 93(39);
105(2); 154(4)**vi**8(9); 31(5); 32(3); 50(45); 67(30);
69(49); 72(7); 85(15); 78(3);
91(2); 94(34); 113(11); 146(2)**vii**30(1); 36(2); 43(28); 49(1); 68(6);
76(2); 82(3); 84(8); 101(37);
110(4); 112(8); 98(10)





A**B**