Characterization of a novel member of genus *Iflavirus* in *Helicoverpa armigera.*

by Yuan, H., Xu, P., Yang, X., Graham, R.I., Wilson, K. and Wu, K.

DOI: http://dx.doi.org/10.1016/j.jip.2017.01.011

Yuan, H., Xu, P., Yang, X., Graham, R.I., Wilson, K. and Wu, K. 2017. Characterization of a novel member of genus *Iflavirus* in *Helicoverpa armigera. Journal of Invertebrate pathology*, 144, pp.65-73

Characterization of a novel member of genus *Iflavirus* in *Helicoverpa armigera*

He Yuan[1#], Pengjun Xu[1,2#], XianmingYang[1], Robert I. Graham[3], Kenneth Wilson[4] and Kongming Wu[1]*

[1]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, P.R. China

[2]Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao, P.R. China

[3]Crop and Environment Sciences, Harper Adams University, Edgmond, Shropshire TF10 8NB, UK

[4]Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

#These authors contributed equally to this work.

***Corresponding author**

Mailing address: State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, P. R. China

E-mail: kmwu@ippcaas.cn

Abstract

The cotton bollworm, *Helicoverpa armigera*, is one of the most important agricultural pests of many economic crops worldwide. Herein, we found a novel single-strand RNA virus by RNA-Seq and Polymerase Chain Reaction (PCR) method in *H. armigera* named *Helicoverpa armigera iflavirus* (HaIV), which possessed a genome with 10,017 nucleotides in length and contained a single large open reading frame (ORF) encoding a putative polyprotein of 3,021 amino acids with a predicted molecular mass of 344.16 kDa and a theoretical isoelectric point (pI) of 6.45. The deduced amino acid sequence showed highest similarity (61.0%) with the protein of *Lymantria dispar Iflavirus 1*. Phylogenetic analysis with putative RdRp amino acid sequences indicated that the virus clustered with members of the genus *Iflavirus*. The virus was mainly distributed in the fat body of its host and was found to be capable of both horizontal and vertical transmission. The efficiency of perorally horizontal transmission was dose dependent (100% infection rate with a viral dose of 10^8 copies /μl) while vertical transmission efficiency was found to be relatively low (< 28.57%). These results suggest that we have found a novel member of genus *Iflavirus* in *H. armigera*.

Keywords: *Helicoverpa armigera*, iflavirus, RNA virus, horizontal transmission, vertical transmission, tissue distribution

1. Introduction

Insects are susceptible to a variety of pathogens, which can result in chronic or lethal infections (Burden et al., 2003). Historically, viruses have been isolated and subsequently studied after the observation of overt disease symptoms in the host. However, there are also covert infections found within hosts displaying no obvious signs of disease, which could remain undetected (Burden et al., 2003; Kemp et al., 2011; Murillo et al., 2011). These covert persistent viral infections, such as baculoviruses, are almost ubiquitous in many lepidopteran insect species and their discovery is driving further research into the dynamics and behavior of covert infections and their role in the ecology of host populations, especially those of economic and agricultural importance (Graham et al., 2015).

The cotton bollworm, *Helicoverpa armigera*, is one of the most important agricultural pests of cotton and other important economic crops worldwide. The adult moth is highly migratory, and populations have been reported in Australia, Asia, Africa, Europe (Feng et al., 2007; Wu and Guo, 2005) and most recently from South America (Tay et al., 2013). Since the introduction of Bt-cotton into China in the 1990s, the *H. armigera* population has declined dramatically. However, several studies have reported that Bt-resistance has evolved in the field (Gunning et al., 2005; Zhang et al., 2012). Thus, other forms of biological pest control, including the use of host-specific viral pesticides, derived from baculovirus (Allaway and Payne, 1984; Chen et al., 2001; Fuxa, 2004; Sun et al., 2002; Sun et al., 2004), small RNA viruses (Christian et al., 2005) and densoviruses (El-Far et al., 2012), has attracted more attention from

researchers. Moreover, high-through-put analytical methods such as metagenomics and RNA sequencing provide sensitive and effective methods for the discovery of novel viruses and asymptomatic disease agents that may be useful as biological control products (Diatchenko et al., 1996; Ge et al., 2012; Marguerat and Bähler, 2009; Mokili et al., 2012; Radford et al., 2012; Roossinck et al., 2015), or conversely, may negatively or positively impact upon the biopesticide products being used. For example, the recent discovery of a novel densovirus (HaDV2) from healthy migratory cotton bollworms revealed that HaDV2 infection significantly increased host resistance to the host-specific baculovirus HaNPV and to the generalist biopesticide Bt toxin (Xu et al., 2014).The development of molecular tools and next generation sequencing technology paves the way for a greater understanding as to how we may manipulate the host-pathogen system, with the aim of reducing pest outbreaks and economic crop damage.

In this study, a novel virus infecting *H. armigera* named Helicoverpa armigera iflavirus (HaIV), was discovered by RNA-Seq. Alignment and phylogenetic analysis revealed that the virus showed a high sequence identity with members of the *Iflavirus*, which is the only genus within the family *Iflaviridae.* Members of this genus possess a single positive-strand RNA genome and share several common features, including: (1) non-enveloped icosahedral particles measuring 30 nm in diameter; (2) genome translation into a polyprotein; (3) the viral coat proteins containing three jelly-roll domains; (4) a three-domain containing a superfamily III helicase, a (cysteine) proteinase with a chymotrypsin-like fold and an RNA-dependent RNA polymerase

(RdRp) (Le Gall et al., 2008). The genome of *Iflavirus* is monocistronic with one single large open reading frame (ORF) encoding a single large polyprotein. To date, only nine species of iflaviruses have been recognized by The International Committee on Taxonomy of Viruses (ICTV), including *deformed wing virus* (Lanzi et al., 2006), *Ectropis obliqua virus* (Wang et al., 2004), *Infectious flacherie virus* (Isawa et al., 1998), *Lygus lineolaris virus 1* (Perera et al., 2012), *Nilaparvata lugens honeydew virus 1* (Murakami et al., 2013), *Perina nuda virus* (Wu et al., 2002), *Sacbrood virus* (Ghosh et al., 1999), *Slow bee paralysis virus* (de Miranda et al., 2010), and *Varroa destructor virus 1* (Ongus et al., 2004), although other iflaviruses have been reported (Silva et al., 2015; Suzuki et al., 2015). Herein, we report the nucleotide sequence, genome organization, phylogeny, transmission and tissue distribution of HaIV.

2. Materials and Methods

2.1 Insect culture

A laboratory colony of *H. armigera* was originally captured in 2005 from Langfang (Hebei province, China). *H. armigera* larvae were reared on an artificial diet (Liang et al., 2008) and adult moths were cultured with a 10% sugar and 2% vitamin mix (Liang et al., 1999) at 25 ± 1°C with a 14:10, light: dark photoperiod.

2.2 Transcriptome analysis and annotation

For transcriptome analysis, Illumina RNA-sequencing was conducted by Novogene (Beijing, China). Four 5^{th} instar larvae (one day post-ecdysis) were individually collected and total RNA isolated using the TRIzol kit (Invitrogen,

Carlsbad, CA, USA), according to the manufacturer's instructions. RNA samples were dissolved in RNase-free water and used to construct the cDNA library of *H. armigera* with suitable fragments (about 200bp). =Paired-end transcriptome sequencing was subsequently performed using an Illumina HiSeq™ 2000. Adaptor sequences and low-quality reads were trimmed and clean reads were used for *de novo* assembly using Trinity (Grabherr et al., 2011). The assembled contigs were annotated using BLASTx and BLASTn against the NCBI non-redundant nucleic acid database (NT) and the NCBI non-redundant protein database (NR), using a cut-off E-value of 10^{-5}.

2.3 Cloning the iflavirus genome of *H. armigera*

Total RNA was extracted from individual adult moths reared in laboratory using the TRIzol reagent kit as described above. Single-stranded cDNA was synthesized using the FastQuant RT Kit (Tiangen, Beijing, China), according to the manufacturer's instructions. Based on the assembled sequence from RNA-seq, nine pairs of primers were subsequently designed (Table S1). The genome of the iflavirus isolated from *H. armigera* was amplified and sequenced using cDNA as template, using the following PCR program: 4 min at 94 °C; 30 s at 94 ^{o}C, 30 s at 55 ^{o}C, and 2 min at 72 ^{o}C for 40 cycles. The PCR product was purified, inserted into the pEASY-T cloning vector (TransGen, Beijing, China), and sequenced.

2.4 Sequence and phylogenetic analysis

The open reading frame (ORF) of the viral genome was predicted using ORF finder at NCBI (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The amino acid

sequences encoding non-structural proteins were compared with members of the order *Picornavirales* using Clustal W (Thompson et al., 1994). The complete nucleotide sequence of the virus described in this study was submitted to GenBank under accession number KX228231. The deduced RdRp amino acid sequence of the new virus, together with that of members of the family *Iflaviridae*, was used in phylogenetic analysis. *Acyrthosiphon pisum virus* and *Spodoptera exigua virus AKJ-2014* were used as outgroup. The phylogenetic trees were constructed using the maximum likelihood method with a bootstrap of 1,000 replicates in MEGA6 (Tamura et al., 2013). Gaps were regarded as a complete deletion unless specifically noted.

2.5 Virus detection and quantification

A partial sequence exhibiting high similarities with known iflaviruses was identified from the RNA-seq data. For the detection of the novel virus, a pair of specific primers, VPF1/VPR1 (Table S1) were designed to amplify a PCR product of 593bp, according to the assembled sequence of the virus in *H. armigera*. The PCR program used was as follows: 30s at 94°C, 30s at 57°C, and 30s at 72°C for 40 cycles. To assess the detection threshold of the virus assay, a 10-fold dilution series of cDNA (3.37×10^{9} to 3.37×10^{1} copies/μl) containing the virus was made and tested using the VPF1/VPR1 primers. 10μl of each PCR product was analysed by agarose gel electrophoresis.

Copy numbers of the novel virus were quantified using a standard curve by an absolute quantification qPCR method (Wong and Medrano, 2005). For quantification, the primers and probes of the virus (VPF/VPR, Table S1) were designed. A fragment

was amplified using the primers and cloned into the pEASY-T Cloning Vector (TransGen, Beijing, China) and sequenced. The PCR program was as follows: 30s at 94°C, 30s at 54°C, and 30s at 72°C for 40 cycles. The plasmid was used for the construction of the standard curve. Virus quantification was conducted with TaqMan in 20μl reaction agent, which comprised 1μl of template DNA, 2×SuperReal PreMix (Probe, Tiangen, Beijing, China), 50×ROX Reference Dye, 0.2mM of each primer and 0.4mM probe. The thermal cycling conditions on a 7500 Real-time PCR System (Applied Biosystems) were as follows: 40 cycles of 95°C for 3s, 60°C for 30s. The standard curve equation of y=-0.9990x+41.6662 (y=the logarithm of plasmid copy number to base 2, x=Ct value, R^2=0.9998) was used to calculate the copy number of the virus (Fig. S1).

2.6 Transmission of the virus

Filtered liquid containing an unpurified form of the iflavirus was prepared (Xu et al., 2014) . RNA was isolated and the concentration of the virus was quantified. A NONINF strain was established from a single breeding pair that was not infected with the iflavirus, according to the method described by Xu et al. (2014). An infected line (INF strain) of *H. armigera* was established by orally infecting NONINF strain larvae with the filtered liquid. Fifth-instar larvae were randomly selected to detect the infection rate of the virus.

For quantification of vertical transmission rates, four pair-types (♀+/♂+, ♀+/♂-, ♀-/♂+, and ♀-/♂-) were mated. Positive and negative individuals were from INF and NONINF culture strains respectively. The infection rate of the virus in offspring was

154 detected from 3rd instar offspring larvae.

155 For quantification of horizontal transmission rates, non-infected NONINF strain
156 neonates were provided artificial diet with different concentrations of the virus: 10^8,
157 10^7, 10^6, 10^5, 10^4, 0 (virus-free control) copies /μl for 2 days, then transferred to a
158 24-well plate and reared in glass tubes until eclosion. Newly eclosed adult moths were
159 sampled to determine the infection rate of the virus under different titer regimes. This
160 diet contamination assay (mimicking a similar mechanism in wild populations) was
161 conducted according to the method described by Xu et al. (2014). Infection rate was
162 detected by PCR and larval frass from the INF strain was used to quantify the virus
163 copy number by qPCR.

164 2.7 Quantification of the virus in eggs and different tissues

165 To test whether vertical transmission was transovum or transovarial, we
166 quantified the copy number of the virus in three groups of both non-treated (control)
167 and sodium hypochlorite treated eggs (n=50 eggs per group) of *H. armigera*. The eggs
168 were from INF strain breeding pairs, that is both females and males were infected
169 with the virus and treated according to the method described by Xu *et al.* (2014).
170 Third instar larvae originating from these eggs were also sampled to quantify the copy
171 number of the virus.

172 To examine the virus infection in different body tissues of the host, infected
173 fifth-instar larvae were dissected into foregut, midgut, hindgut, hemolymph,
174 malpighian tubules and fat body, using a fresh scalpel for every cut to avoid virus
175 cross contamination. In addition, infected females and males were dissected into brain,

muscle, wing, malpighian tubule, fat body, ovary/testis and gut as described above. Total RNA was extracted from the different body parts (both larval and adult stages) and cDNA was used as a template to quantify the copy numbers of the virus by qPCR. The cDNA sample of each body part was replicated three times. The copy number of the virus in each tissue was calculated, and these were summed to determine total copy numbers of the virus in each individual. The percentage of virus in each tissue in the same individual was statistically analyzed (larvae: n =9; adult males: n= 6; adult females: n= 9) (Xu et al., 2014).

To examine virus replication in different developmental stages of the host, larvae 24h to 240h post-infection (1^{st}-5^{th} instar) and newly eclosed adults were sampled. Absolute quantification qPCR and the comparative $2^{-\triangle\triangle Ct}$ method (Livak and Schmittgen, 2001) were used, respectively. *β-actin* (GenBank EU527017) was used as the reference gene to normalize the virus expression. The reaction was conducted in 20μl reaction mixtures, containing 10μl of SuperReal PreMix Plus (TIANGEN, Beijing, China), 0.6μl of primers (10μM), 1μl of sample cDNA, 0.5μl of Rox Reference Dye and 8.3μl of RNase-free ddH_2O. The cycling parameters were: 95℃ for 15min, followed by 40 cycles of 95℃ for 10s and 62℃ for 32s. To ensure reproducibility, each sample was carried out in three biological replicates and in three technical replicates.

2.8 Electron microscopy

Adult moths from the INF culture strain were collected. Then the virus were isolated and purified using the method of Sucrose Density Gradient Centrifugation

198 according to the method described by La Fauce et al (2007). Purified particles were
199 negatively stained with 2% sodium phosphotungstate at pH 6.8 and observed with a
200 transmission electron microscope.

201 **3. Results**

202 3.1 Transcriptome analysis and annotation

203 Using RNA-Seq we sequenced the transcriptome of *H. armigera*, obtaining
204 about 5 gigabases (in-depth) for each sample, and a total of 106,785 assembled
205 contigs. Using BLASTx and BLASTn, 45609 genes (42.71% of transcripts) and
206 34383 genes (32.20% of transcripts), respectively, were obtained through BLAST hits,
207 using the E-value cutoff. Because of the relatively short length of transcripts (mean
208 size of 758.65 bp) and lack of genome reference, most of the assembled sequences
209 could not be matched to any known genes. The E-value distribution of the best hit in
210 the nr database showed that 57% of the mapped sequences had strong homology
211 (smaller than 1.0E-50), whereas 43% of the homologous sequences ranged from
212 1.0E-5 to 1.0E-50 (Fig. S2A). Homologous genes came from several species, with
213 66% of the unigenes having the highest homology to genes from *Danaus plexippus*,
214 followed by *Bombyx mori* (5%), *Tribolium castaneum* (3%), and *Helicoverpa*
215 *armigera* (2%) (Fig. S2B). The RNA-seq original datasets generated in this study are
216 available in the NCBI GEO database (accession number: GSE86914)
217 (https://www.ncbi.nlm.nih.gov/geo/).

218 3.2 The genome sequence of HaIV

219 The assembled contig of 1495 nucleotides (nt) in length from 103,935 reads

encoded actin of *H. armigera*, and one contig of 10008 nt in length assembled using 41719 reads showed high identity with *Lymantria dispar* iflavirus 1. According to the reference sequence, we designed specific primers to amplify the viral genome sequence of HaIV containing the whole ORF, which was 10,017 nt in length, containing a single large ORF (between nt 733 and 9798) encoding a polyprotein of 3021 amino acids. It has a predicted molecular mass of 344.16 kDa and a theoretical isoelectric point (pI) of 6.45. The coding sequence was flanked by a 732bp 5' untranslated region (UTR) and a 219bp 3' UTR. The nucleotide base composition of the genome was 30.6%A, 35.6%U, 13.6%C and 20.2%G. The total A+U and G+C content were 66.2% and 33.8% respectively. Alignment analysis with putative amino acid sequences showed high identities with the members of genus *Iflavirus*, in which the highest identity was 61.0% with *L. dispar* iflavirus 1 (Table S2). It contained all three conserved domains, including the three conserved domains in helicase sequences (Fig. 1A), the GXCG and GXHXXG conserved motifs in the protease sequences (Fig. 1B) and the eight conserved domains in the RdRp amino acid sequences (Fig. 1C). The conservation of HaIV polyprotein and the various segments were compared with those of *Heliconius erato virus* (HeIV) (Smith et al., 2014) and the results are summarized in Table 1.

3.3 Phylogenetic analysis

Neighbor-joining trees with Poisson model were constructed for the putative RdRp amino acid sequences (Fig. 2) of the family *Iflaviridae* using *Acyrthosiphon pisum* virus and *Spodoptera exigua* virus AKJ-2014 as the outgroup. The result

242 indicated that the HaIV clustered with members of the genus *Iflavirus*.

243 3.4 Sensitivity of detection

244 An amplification product could be visualized by ethidium bromide staining when
245 as little as 3.37×10^3 copies/ul of HaIV in cDNA were used as template (Fig. S3).

246 3.5 Transmission of HaIV

247 The virus was capable of being vertically transmitted from both infected males
248 and females, but the transmission efficiency was relatively low (< 28.6%) and that
249 from infected females was higher than that from infected males (Table 2). To test
250 whether vertical transmission of the virus was due to transovum or transovarial,
251 quantification of INF eggs was conducted and indicated that the virus titers were 1000
252 times greater in non-treated eggs than in eggs treated with sodium hypochlorite
253 solution (paired t-test: t = 15.649, d.f. = 5, $P < 0.001$) (Fig. 3). 8 out of 17 individuals
254 (47.06%) hatching from non-treated eggs contained more than 1.0×10^7 copies/mg of
255 HaIV, whereas larvae hatching from eggs treated with sodium hypochlorite solution
256 contained no more than 3.71×10^2 copies/mg of HaIV (n=10), suggesting that
257 transovum transmission was occurring via the surface of eggs (Fig. 4).

258 To examine horizontal transmission, NONINF strain neonates were exposed to
259 different concentrations of virus. The results showed that the infection rate of
260 NONINF strain larvae was dose-dependent, with 100% infection rates at a dose of 10^8
261 copies /μl (Table 3). To examine the transmission efficiency through frass of larvae,
262 we placed NONINF strain neonates in diet cells which had previously housed infected
263 insects (n=12) and quantified the copy number of the virus in frass of larvae. The

264 results showed that the frass contained no more than 10^6 copies /mg (Fig. 5). However,
265 75% (9 of 12 samples) of NONINF strain individuals could be infected via horizontal
266 transmission by frass (Fig. 6).

267 3.6 Host tissue distribution

268 The copy number of the virus in different body tissues of *H. armigera* was
269 quantified by qPCR. In both larvae and adults, the virus titers in the fat body were
270 significantly higher than in other tissues: larvae: $F = 11.32$, df = 5,12, $P < 0.001$ (Fig.
271 7A); adult females: $F = 11.57$, df = 6,56, $P < 0.001$ (Fig. 7B); adult males: $F = 2.89$,
272 df = 6,21, $P=0.033$ (Fig. 7C). The virus was also detected in female ovaries and male
273 testes, but at lower titers than in fat body.

274 3.7 Total levels of HaIV

275 Absolute quantification qPCR results showed that the copy number of HaIV
276 increased over time, reaching the highest infection load (1.36×10^7 copies/mg) at
277 144h post infection and remaining stable after this time (Fig. 8A). In undertaking the
278 relative quantification assay, all of the expression levels of HaIV were compared with
279 those at 24h. The expression levels of HaIV increased over time and reached 10,000
280 fold at 240h, with an addition slight increase in expression during the adult stage (Fig.
281 8B).

282 3.8 Virus morphology

283 The virus particles purified from *H. armigera* were observed by using an
284 electron microscope. They had an isometric appearance and an approximate diameter
285 of 30nm (Fig. 9).

4. Discussion

Recently, next-generation sequencing technology has provided a rapid approach to high-throughput sequence determination and allowed a wide diversity of novel viruses to be discovered (Ansorge, 2009; Mokili et al., 2012; Roossinck et al., 2015). Herein, a novel virus named *H. armigera iflavirus* (HaIV) was detected in larvae of *H. armigera* by RNA-seq method. The genome organization of HaIV has the common features of other members within the family *Iflaviridae*. Briefly, the virus is monocistronic, with a single-stranded RNA genome, and contains a single large ORF encoding both structural and non-structural proteins. It also possesses the functional motifs of RNA helicase, protease and RdRp conserved in members of the order *Picornavirales* (van Oers, 2010). As members of the genus *Iflavirus*, the genomic sequence of HaIV is A/U rich (> 60% A/U). Based on the genomic nucleotide sequences and the amino acid sequences, phylogenetic analysis indicated HaIV clustered with the members of *Iflavirus*, suggesting that HaIV was a novel member of genus *Iflavirus* (Carrillo-Tripp et al., 2014; Silva et al., 2015; van Oers, 2010).

Viruses distribution in host tissues were diverse according to different virus species, such as HaDV2, a beneficial densovirus which could accelerate the growth rate of its insect host, was mainly distributed in fat body of *H. armigera* (Xu et al., 2014) and HaNPV, a lethal baculovirus, viral loads of which were higher in the head, legs and wings than in the abdomen and thorax (Graham et al., 2015). Previously, some members of *Iflavirus* don’t appear to be lethal to their hosts, such as Kakugo virus (Fujiyuki et al., 2009) and Brevicoryne brassicae virus (Ryabov, 2007). And,

some members of *Iflavirus* are known to be harmful to their hosts by lead to developmental problems and host mortality (van Oers, 2010). For example, sacbrood virus (SBV) is mainly found in larval fat body cells of the honeybee, which may impact metabolic function of these cells, resulting in failure to pupate and ultimately causing death (Bitondi et al., 2006; Park et al., 2016). Interestingly, HaIV was also mainly distributed in the fat body both in larvae and adults while with much lower titers in both the ovary and testis, suggesting a similar function with other members of genus *Iflavirus*.

As beneficial virus, HaDV2 could be high efficiently transmitted by both the horizontal and vertical mode (Xu et al., 2014). However, the harmful virus (eg. HaNPV) could infect host through high horizontal transmission rate (86.2% mortality at a dose of 10^7 OBs/ml) whereas quite low vertical transmission rate (12.6% ± 2.0%) (Zhou et al., 2005). Members of *Iflavirus* can transmit horizontally and vertically. For example, DWV could vertically transmit from queens to both worker and drone offspring (Yue and Genersch, 2005), and horizontally transmit through *Varroa* mites (Bowen-Walker et al., 1999; Wilfert et al., 2016). Transovarial and horizontal transmission of SBV have also been found to occur (Shen et al., 2005). Herein, we also found that HaIV could be both horizontally and vertically transmitted, which the horizontally transmitted efficiency was dose-dependent and the vertically transmitted efficiency was quite low (< 28.57%). To examine the possibility of horizontal transmission on natural conditions, we performed diet contamination assay and the results suggested that although the dose of HaIV in frass was no more than 10^6/mg,

the infection rate was 75%. The copy number of HaIV associated with eggs was significantly decreased by washing with sodium hypochlorite solution, suggesting that transovum transmission was occurring via the surface of eggs.

In conclusion, we report a novel virus isolated from the host lepidopteran *H. armigera* named *H. armigera* iflavirus. Molecular characterization and phylogenetic analysis indicated that HaIV was a novel member of the genus *Iflavirus*. HaIV was found to be mainly distributed in the fat body of its host, and could be both horizontally and vertically transmitted but with low efficiency, suggesting a harmful factor to its host.

Acknowledgments

This work was supported by Science Fund for Creative Research Groups of the National Science Foundation of China (No. 31321004) and the National Natural Science Foundation of China (Grant No. 31401752). KW was supported by a Biotechnology and Biological Sciences Research Council UK-China Partnering Award (ref: BB/L026821/1).

Reference

Allaway, G. P., Payne, C. C., 1984. Host range and virulence of five baculoviruses from lepidopterous hosts. Ann. Appl. Biol. 105, 29-37.

Ansorge, W. J., 2009. Next-generation DNA sequencing techniques. N. Biotechnol. 25, 195-203.

Bitondi, M. M., Nascimento, A. M., Cunha, A. D., Guidugli, K. R., Nunes, F. M., Simões, Z. L., 2006. Characterization and expression of the Hex 110 gene encoding a glutamine-rich hexamerin in the honey bee, *Apis mellifera*. Arch. insect biochem. 63, 57-72.

Bowen-Walker, P., Martin, S., Gunn, A., 1999. The transmission of deformed wing virus between honeybees (*Apis mellifera* L.) by the ectoparasitic mite*Varroa jacobsoni* Oud. J. Invertebr. Pathol. 73, 101-106.

Burden, J. P., Nixon, C. P., Hodgkinson, A. E., Possee, R. D., Sait, S. M., King, L. A., Hails, R. S., 2003. Covert infections as a mechanism for long-term persistence of baculoviruses. Ecol. Lett. 6, 524-531.

Carrillo-Tripp, J., Krueger, E. N., Harrison, R. L., Toth, A. L., Miller, W. A., Bonning, B. C., 2014. *Lymantria dispar iflavirus 1* (LdIV1), a new model to study iflaviral persistence in lepidopterans. J. Gen. Virol. 95, 2285-2296.

Chen, X., IJkel, W. F., Tarchini, R., Sun, X., Sandbrink, H., Wang, H., Peters, S., Zuidema, D., Lankhorst, R. K., Vlak, J. M., 2001. The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. J. Gen. Virol. 82, 241-257.

Christian, P. D., Murray, D., Powell, R., Hopkinson, J., Gibb, N. N., Hanzlik, T. N., 2005. Effective control of a field population of *Helicoverpa armigera* by using the small RNA virus *Helicoverpa armigera stunt virus* (*Tetraviridae: Omegatetravirus*). J. econ. ent. 98,

374 1839-1847.
375 de Miranda, J. R., Dainat, B., Locke, B., Cordoni, G., Berthoud, H., Gauthier, L., Neumann, P., Budge,
376 G. E., Ball, B. V., Stoltz, D. B., 2010. Genetic characterization of slow bee paralysis virus of
377 the honeybee (*Apis mellifera* L.). J. Gen. Virol. 91, 2524-2530.
378 Diatchenko, L., Lau, Y., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S.,
379 Lukyanov, K., Gurskaya, N., Sverdlov, E. D., 1996. Suppression subtractive hybridization: a
380 method for generating differentially regulated or tissue-specific cDNA probes and libraries.
381 Proc. Natl. Acad. Sci. U S A 93, 6025-6030.
382 El-Far, M., Szelei, J., Yu, Q., Fediere, G., Bergoin, M., Tijssen, P., 2012. Organization of the
383 ambisense genome of the *Helicoverpa armigera densovirus*. J. Virol. 86, 7024.
384 Feng, H., Wu, K., Ni, Y. X., Cheng, D., Guo, Y., 2007. Return migration of *Helicoverpa armigera*
385 (Lepidoptera: Noctuidae) during autumn in northern China. B. Entomol. Res. 95, 361-370.
386 Fujiyuki, T., Matsuzaka, E., Nakaoka, T., Takeuchi, H., Wakamoto, A., Ohka, S., Sekimizu, K.,
387 Nomoto, A., Kubo, T., 2009. Distribution of Kakugo virus and its effects on the gene
388 expression profile in the brain of the worker honeybee *Apis mellifera* L. J. Virol. 83,
389 11560-11568.
390 Fuxa, J. R., 2004. Ecology of insect nucleopolyhedroviruses. Agr. Ecosyst. Environ. 103, 27-43.
391 Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y., Shi, Z., 2012. Metagenomic analysis of
392 viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. J.
393 Virol. 86, 4620-4630.
394 Ghosh, R., Ball, B., Willcocks, M., Carter, M., 1999. The nucleotide sequence of sacbrood virus of the
395 honey bee: an insect picorna-like virus. J. Gen. Virol. 80, 1541-1549.

396 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan,
397 L., Raychowdhury, R., Zeng, Q., 2011. Full-length transcriptome assembly from RNA-Seq
398 data without a reference genome. Nature biotechnology. 29, 644-652.
399 Graham, R. I., Tummala, Y., Rhodes, G., Cory, J. S., Shirras, A., Grzywacz, D., Wilson, K., 2015.
400 Development of a real-time qPCR assay for quantification of covert baculovirus infections in
401 a major African crop pest. Insects 6, 746-759.
402 Gunning, R. V., Dang, H. T., Kemp, F. C., Nicholson, I. C., Moores, G. D., 2005. New resistance
403 mechanism in *Helicoverpa armigera* threatens transgenic crops expressing *Bacillus*
404 *thuringiensis* Cry1Ac toxin. Appl. Environ. Microb. 71, 2558-2563.
405 Isawa, H., Asano, S., Sahara, K., Iizuka, T., Bando, H., 1998. Analysis of genetic information of an
406 insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary
407 relationships among insect, mammalian and plant picorna (-like) viruses. Arch. Virol. 143,
408 127-143.
409 Kemp, E. M., Woodward, D. T., Cory, J. S., 2011. Detection of single and mixed covert baculovirus
410 infections in eastern spruce budworm, *Choristoneura fumiferana* populations. J. Invertebr.
411 Pathol. 107, 202-205.
412 Koonin, E. V., Dolja, V. V., Morris, T. J., 1993. Evolution and taxonomy of positive-strand RNA
413 viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem.
414 Mol. Biology 28, 375-430.
415 Lanzi, G., de Miranda, J. R., Boniotti, M. B., Cameron, C. E., Lavazza, A., Capucci, L., Camazine, S.
416 M., Rossi, C., 2006. Molecular and biological characterization of deformed wing virus of
417 honeybees (*Apis mellifera* L.). J. Virol. 80, 4998-5009.

418 Le Gall, O., Christian, P., Fauquet, C. M., King, A. M., Knowles, N. J., Nakashima, N., Stanway, G.,
419 Gorbalenya, A. E., 2008. Picornavirales, a proposed order of positive-sense single-stranded
420 RNA viruses with a pseudo-T = 3 virion architecture. Arch. Virol. 153, 715-727.
421 Liang, G., Tan, W., Guo, Y., 1999. An improvement in the technique of artificial rearing cotton
422 bollworm. Plant Protec. 25, 15-17.
423 Liang, G., Wu, K., Yu, H., Li, K., Feng, X., Guo, Y., 2008. Changes of inheritance mode and fitness in
424 *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae) along with its resistance evolution
425 to Cry1Ac toxin. J. Invertebr. Pathol. 97, 142-149.
426 Livak, K. J., T. D. Schmittgen. 2001. Analysis of relative gene expression data using real-time
427 quantitative PCR and the 2 (-Delta Delta C(T)) Method. Methods. 25: 402-408. Marguerat, S.,
428 Bähler, J., 2009. RNA-seq: from technology to biology. Cell. Mol. Life Sci. 67, 569-579.
429 Mokili, J. L., Rohwer, F., Dutilh, B. E., 2012. Metagenomics and future perspectives in virus
430 discovery. Curr. Opin. Virol. 2, 63-77.
431 Murakami, R., Suetsugu, Y., Kobayashi, T., Nakashima, N., 2013. The genome sequence and
432 transmission of an iflavirus from the brown planthopper, *Nilaparvata lugens*. Virus Res. 176,
433 179-187.
434 Murillo, R., Hussey, M. S., Possee, R. D., 2011. Evidence for covert baculovirus infections in a
435 *Spodoptera exigua* laboratory culture. J. Gen. Virol. 92, 1061-1070.
436 Ongus, J. R., Peters, D., Bonmatin, J. M., Bengsch, E., Vlak, J. M., van Oers, M. M., 2004. Complete
437 sequence of a picorna-like virus of the genus Iflavirus replicating in the mite *Varroa*
438 *destructor*. J. Gen. Virol. 85, 3747-3755.
439 Park, C., Kang, H., Jeong, J., Kang, I., Choi, K., Yoo, M.S., Kim, Y.H., Kang, S.W., Lim, H.Y., Yoon,

440 B.S., 2016. In-situ Hybridization for the Detection of Sacbrood Virus in Infected Larvae of the
441 Honey Bee (*Apis cerana*). J. Comp. Pathol. 154, 258-262.
442 Perera, O. P., Snodgrass, G. L., Allen, K. C., Jackson, R. E., Becnel, J. J., O'Leary, P. F., Luttrell, R.
443 G., 2012. The complete genome sequence of a single-stranded RNA virus from the tarnished
444 plant bug, *Lygus lineolaris* (Palisot de Beauvois). J. Invertebr. Pathol. 109, 11-19.
445 Radford, A. D., Chapman, D., Dixon, L., Chantrey, J., Darby, A. C., Hall, N., 2012. Application of
446 next-generation sequencing technologies in virology. J. Gen. Virol. 93, 1853-1868.
447 Roossinck, M. J., Martin, D. P., Roumagnac, P., 2015. Plant virus metagenomics: Advances in virus
448 discovery. Phytopathology 105, 716-727.
449 Ryabov, E. V., 2007. A novel virus isolated from the aphid *Brevicoryne brassicae* with similarity to
450 Hymenoptera picorna-like viruses. J. Gen. Virol. 88, 2590-2595.
451 Shen, M., Cui, L., Ostiguy, N., Cox-Foster, D., 2005. Intricate transmission routes and interactions
452 between picorna-like viruses (Kashmir bee virus and sacbrood virus) with the honeybee host
453 and the parasitic *varroa* mite. J. Gen. Virol. 86, 2281-2289.
454 Silva, L. A., Ardisson-Araujo, D. M., Tinoco, R. S., Fernandes, O. A., Melo, F. L., Ribeiro, B. M.,
455 2015. Complete genome sequence and structural characterization of a novel iflavirus isolated
456 from *Opsiphanes invirae* (Lepidoptera: Nymphalidae). J. Invertebr. Pathol. 130, 136-140.
457 Smith, G., Macias-Muñoz, A., Briscoe, A. D., 2014. Genome sequence of a novel iflavirus from
458 mRNA sequencing of the butterfly *Heliconius erato*. Genome Announc. 2: e00398-14.
459 Sun, X., Chen, X., Zhang, Z., Wang, H., Bianchi, F. J., Peng, H., Vlak, J. M., Hu, Z., 2002. Bollworm
460 responses to release of genetically modified *Helicoverpa armigera* nucleopolyhedroviruses in
461 cotton. J. Invertebr. Pathol. 81, 63-69.

462 Sun, X., Wang, H., Sun, X., Chen, X., Peng, C., Pan, D., Jehle, J. A., van der Werf, W., Vlak, J. M.,
463 Hu, Z., 2004. Biological activity and field efficacy of a genetically modified *Helicoverpa*
464 *armigera* single-nucleocapsid nucleopolyhedrovirus expressing an insect-selective toxin from
465 a chimeric promoter. Biol. Control 29, 124-137.
466 Suzuki, T., Takeshima, Y., Mikamoto, T., Saeki, J. D., Kato, T., Park, E. Y., Kawagishi, H., Dohra, H.,
467 2015. Genome sequence of a novel iflavirus from mRNA sequencing of the pupa of *Bombyx*
468 *mori* inoculated with *Cordyceps militaris*. Genome Announc. 3, e01039-15.
469 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary
470 genetics analysis version 6.0. Mol. Biol. evol. 30, 2725-2729.
471 Tay, W. T., Soria, M. F., Walsh, T., Thomazoni, D., Silvie, P., Behere, G. T., Anderson, C., Downes,
472 S., 2013. A brave new world for an old world pest: *Helicoverpa armigera* (Lepidoptera:
473 Noctuidae) in Brazil. PLoS One. 8**,** e80134.
474 Thompson, J. D., Higgins, D. G., Gibson, T. J., 1994. CLUSTAL W: improving the sensitivity of
475 progressive multiple sequence alignment through sequence weighting, position-specific gap
476 penalties and weight matrix choice. Nucleic acids res. 22, 4673-4680.
477 van Oers, M. M., 2010. Genomics and biology of Iflaviruses. Insect Virology. Caister Academic Press,
478 Norfolk. 231-250.
479 Wang, X., Zhang, J., Lu, J., Yi, F., Liu, C., Hu, Y., 2004. Sequence analysis and genomic organization
480 of a new insect picorna-like virus, Ectropis obliqua picorna-like virus, isolated from *Ectropis*
481 *obliqua*. J. Gen. Virol. 85, 1145-1151.
482 Wong, M. L., Medrano, J. F., 2005. Real-time PCR for mRNA quantitation. Biotechniques 39, 75-85.
483 Wu, C., Yang, H., Lo, C., Wang, C., 2002. A *Perina nuda* cell line (NTU-Pn-HF) from pupal ovary

484 that is persistently infected with a picorna-like virus (PnPV). Appl. Entomol. zool. 37,
485 171-179.
486 Wu, K., Xu, G., Guo, Y., 1997. Observations on migratory activity of cotton bollworm woths across
487 the Bohai Gulf in China. Acta. Phytophy. Sin. 25, 337-340.
488 Xu, P., Liu, Y., Graham, R. I., Wilson, K., Wu, K., 2014. Densovirus is a mutualistic symbiont of a
489 global crop pest (*Helicoverpa armigera*) and protects against a baculovirus and Bt
490 biopesticide. PLoS Pathog. 10, e1004490.
491 Yue, C., Genersch, E., 2005. RT-PCR analysis of deformed wing virus in honeybees (*Apis mellifera*)
492 and mites (*Varroa destructor*). J. Gen. Virol. 86, 3419-3424.
493 Zhang, H., Tian, W., Zhao, J., Jin, L., Yang, J., Liu, C., Yang, Y., Wu, S., Wu, K., Cui, J., Tabashnik,
494 B. E., Wu, Y., 2012. Diverse genetic basis of field-evolved resistance to Bt cotton in cotton
495 bollworm from China. Proc. Natl. Acad. Sci. U S A 109, 10275-10280.
496 Zhou, M., Sun, X., Sun, X., Vlak, J. M., Hu, Z., van der Werf, W., 2005. Horizontal and vertical
497 transmission of wild-type and recombinant *Helicoverpa armigera* single-nucleocapsid
498 nucleopolyhedrovirus. J. Invertebr. Pathol. 89, 165-175.
499

Supporting information

Table S1 Primers used in this study.

Table S2 Percentage amino acid differences between accepted and proposed iflavirus species.

Table S3 Full names and GenBank accession numbers of the viruses used in this paper.

Figure S1 The standard curve for quantifying the number of virus particles using Taqman method.

Figure S2 (A) The E-value distribution of the top matches in the nr database. (B) The species distribution of best hits in the nr database.

Figure S3 Demonstration of the detection threshold of HaIV by agarose gel electrophoresis using the VPF1/VPR1 primers.

Figure legends

Figure 1. Protein domain alignments. Shown are alignments of the helicase (A), protease (B) and RdRp sequences (C) domains of the novel virus (HaIV) with those of the other *Iflaviridae* (ApIV, DWV, EoV, HeIV, IFV, LyLV-1, PnV, SBV and VDV-1), some *Dicistroviridae* (ABPV, BQCV, CrPV and DCV) and one *Picornaviridae* (EMCV). Conserved regions corresponding to those recognized by Koonin et al. (1993) are indicated by bars above the protein alignment. Black shading indicates 100% sequence identity and other residues that are also conserved among these sequences are shaded in gray. The full names and the accession numbers of the virus are given in Table S3.

Figure 2. Maximum likelihood phylogenetic analysis of the family *Iflaviridae*, with *Acyrthosiphon pisum* virus and *Spodoptera exigua* virus AKJ-2014 used as the outgroup. The phylograms are based on the putative RdRp amino acid sequence (Fig. 2). The Genbank accession numbers are listed behind each virus species. Numbers at nodes represent bootstrap values as percentages estimated by 1000 replicates in an analysis using Clustal_X software. Branches with less than 50% bootstrap support were collapsed.

Figure 3. Viral load of HaIV in cotton bollworm eggs. Absolute quantification of HaIV copy number per ug of host RNA in eggs washed or non-washed in 1% sodium hypochlorite (n=6).

Figure 4. Viral load of HaIV in larvae hatching from non-treated eggs (C1-C17) and eggs treated with sodium hypochlorite solution (T1-T10).

Figure 5. Absolute quantification of HaIV copy number per mg of frass of larvae. LF: larval frass.

Figure 6. The detection of HaIV in sample used in diet contamination assay. (a) PCR detection of HaIV in larvae which were reared on contaminated diet for 6 days; (b) PCR detection of HaIV in frass of larvae; (c) β-actin was used as internal reference gene to test the integrity of each cDNA templates.

Figure 7. Tissue distribution of the HaIV in A) larvae, B) adult females and C) adult male cotton bollworms. Within each figure, significant differences ascribed using Tukey tests are shown using different letters. Percentage (%) = the ratio of HaIV in different tissues (per mg), as described by Xu et al. (2014) (larvae: n = 9; adult males: n = 6; adult females: n = 9). Means ± SE.

Figure 8. HaIV levels in different developmental stages of host were tested using two methods. (A) absolute quantification qPCR method. (B) the comparative $2^{-\triangle\triangle Ct}$ method.

Figure 9. Electron microscopy image showing HaIV particles purified from an extract from INF strain (100000X).

Table 1. Sequence identity and similarity in an alignment between nucleotide and amino acid of HaIV and HeIV.

HaIV sequence	Sequence identity (similarity) (%)
Nucleotide sequence	
3'NTR	20.91
5'UTR	35.77
Amino acid sequence	
Entire polyprotein	58.78
RdRp	80.22
Protease	67.50
Helicase	92.17
Rhv_like	58.92
Rhv_like	77.95
CRPV_capsid	52.78

Table 2. Vertical transmission efficiency of HaIV.

Individuals	Number testing + ve	Number testing - ve	Transmission efficiency (%)
Female+/Male+	6	17	26.09
Female+/Male-	4	10	28.57
Female-/Male+	1	23	4.17
Female-/Male-	0	15	0.00

Infected individuals = "+", uninfected individuals = "-"

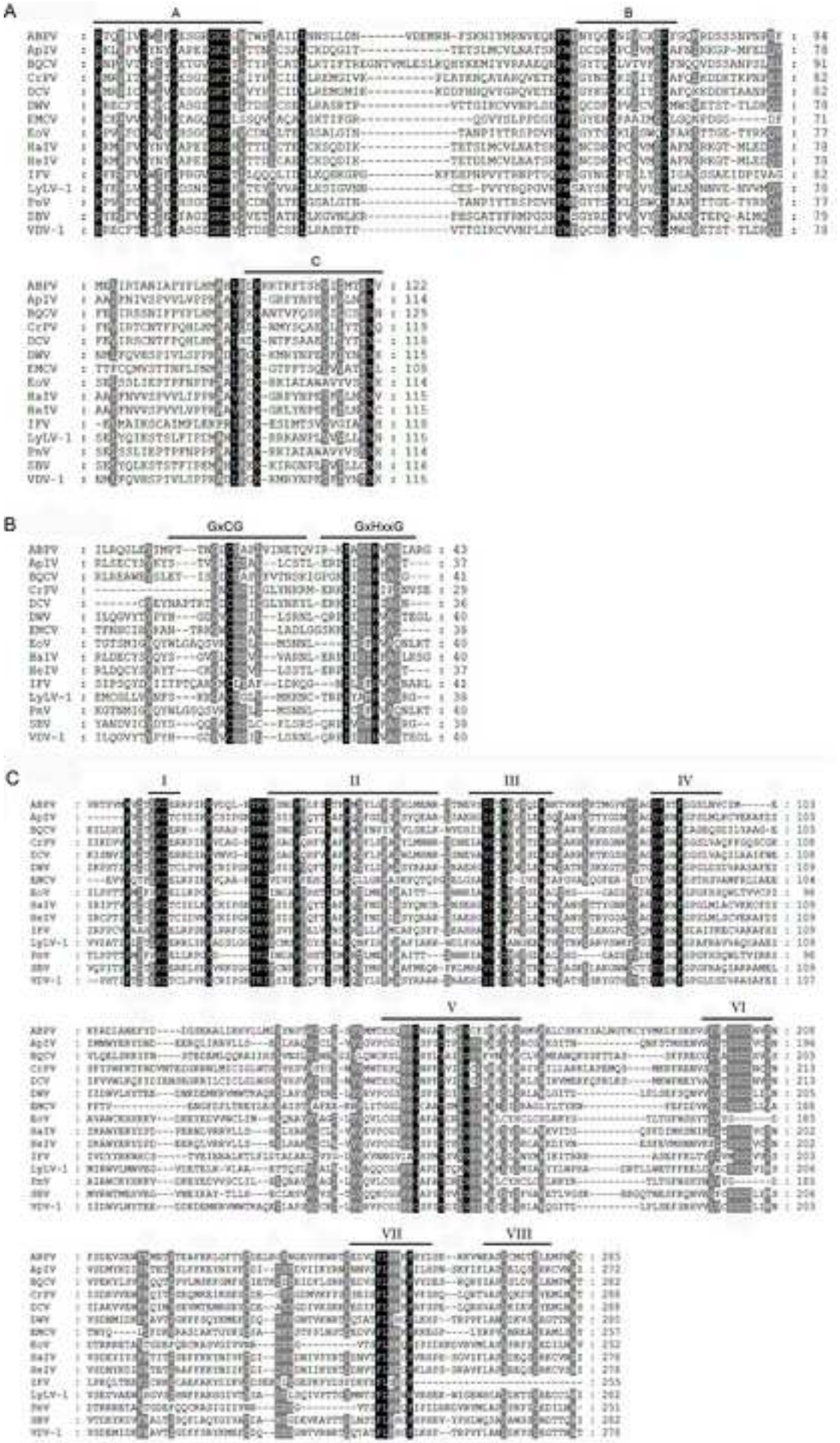
Table 3. Detection of HaIV infecting larvae dosed at a range of concentrations.

Concentrations (copy number/ul)	Number testing + ve	Number testing - ve	Infection rate (%)
10^8	11	0	100.00
10^7	10	3	76.92
10^6	9	3	75.00
10^5	8	4	66.67
10^4	4	8	33.33
0	0	12	0

Infected individuals = "+ve", uninfected individuals = "-ve".

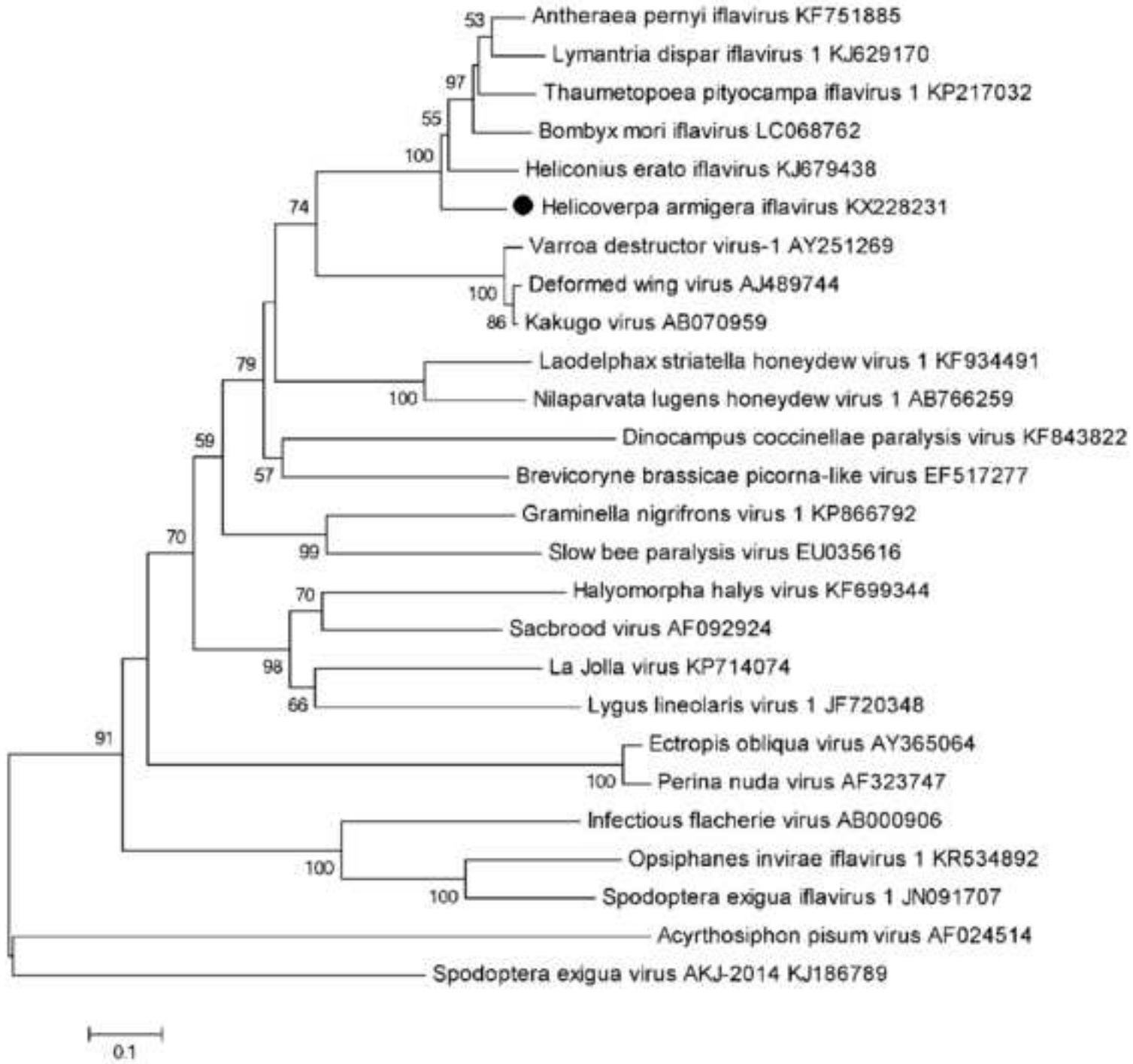
Figure

Click here to download high resolution image



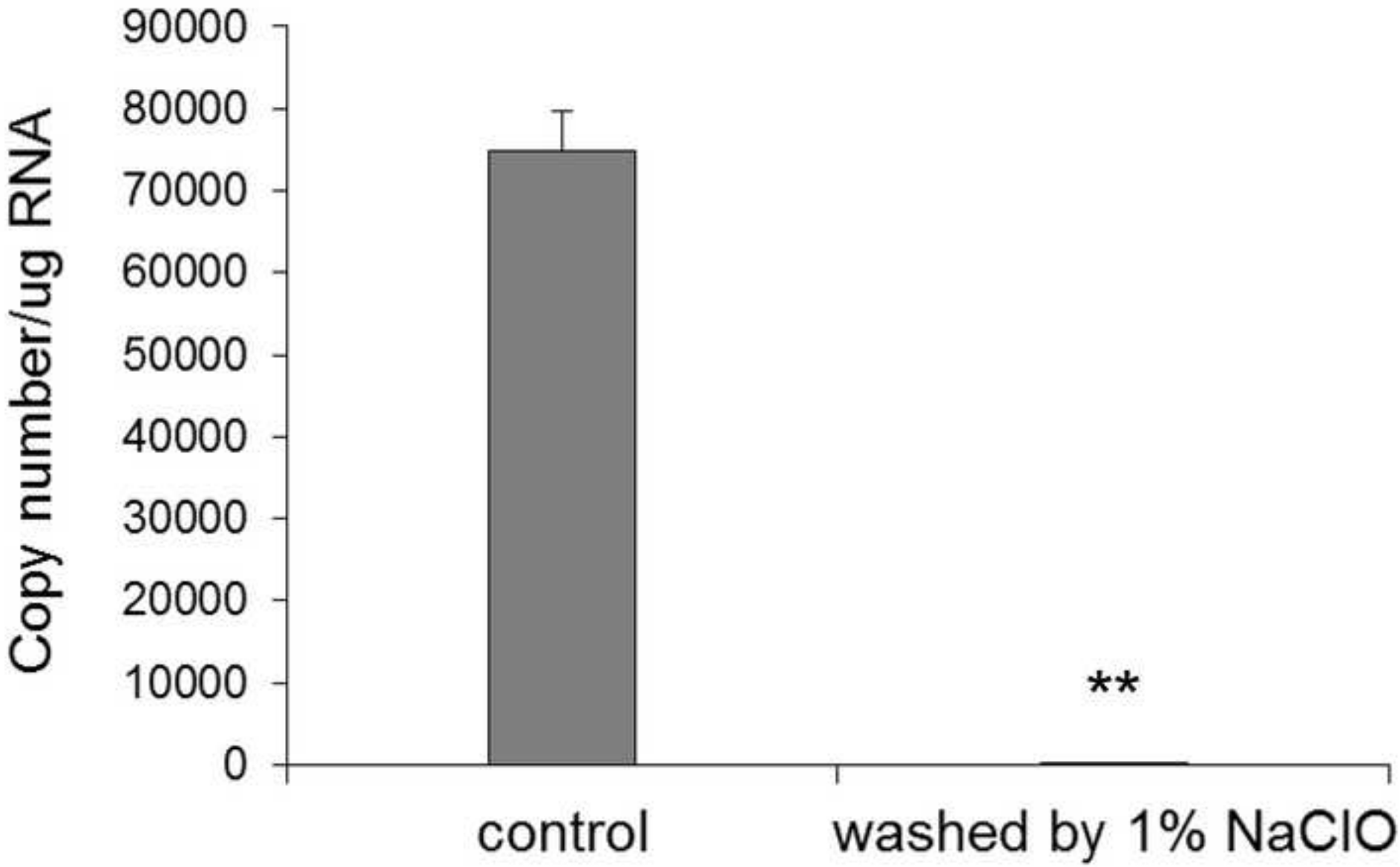
Figure

Click here to download high resolution image



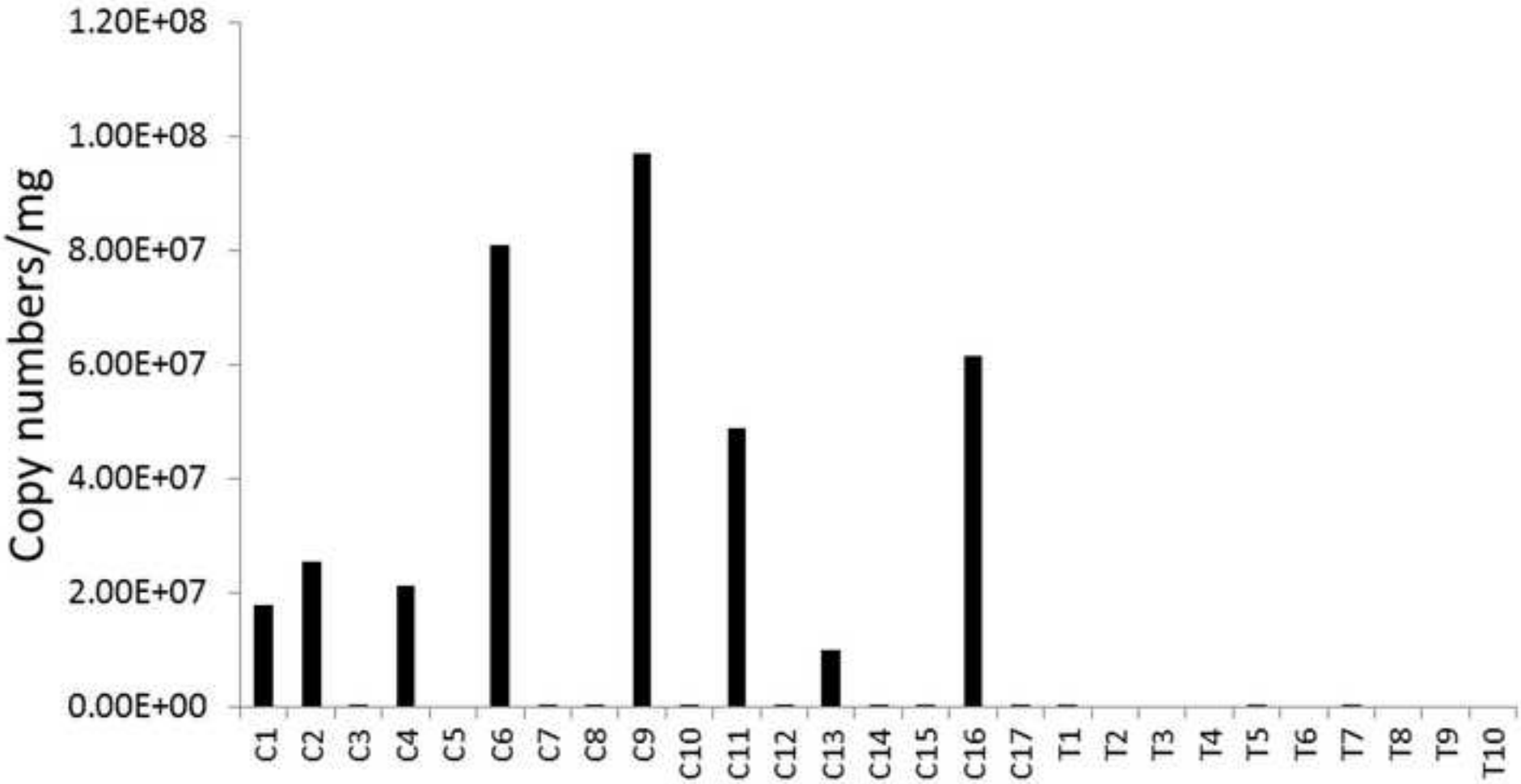
Figure

Click here to download high resolution image



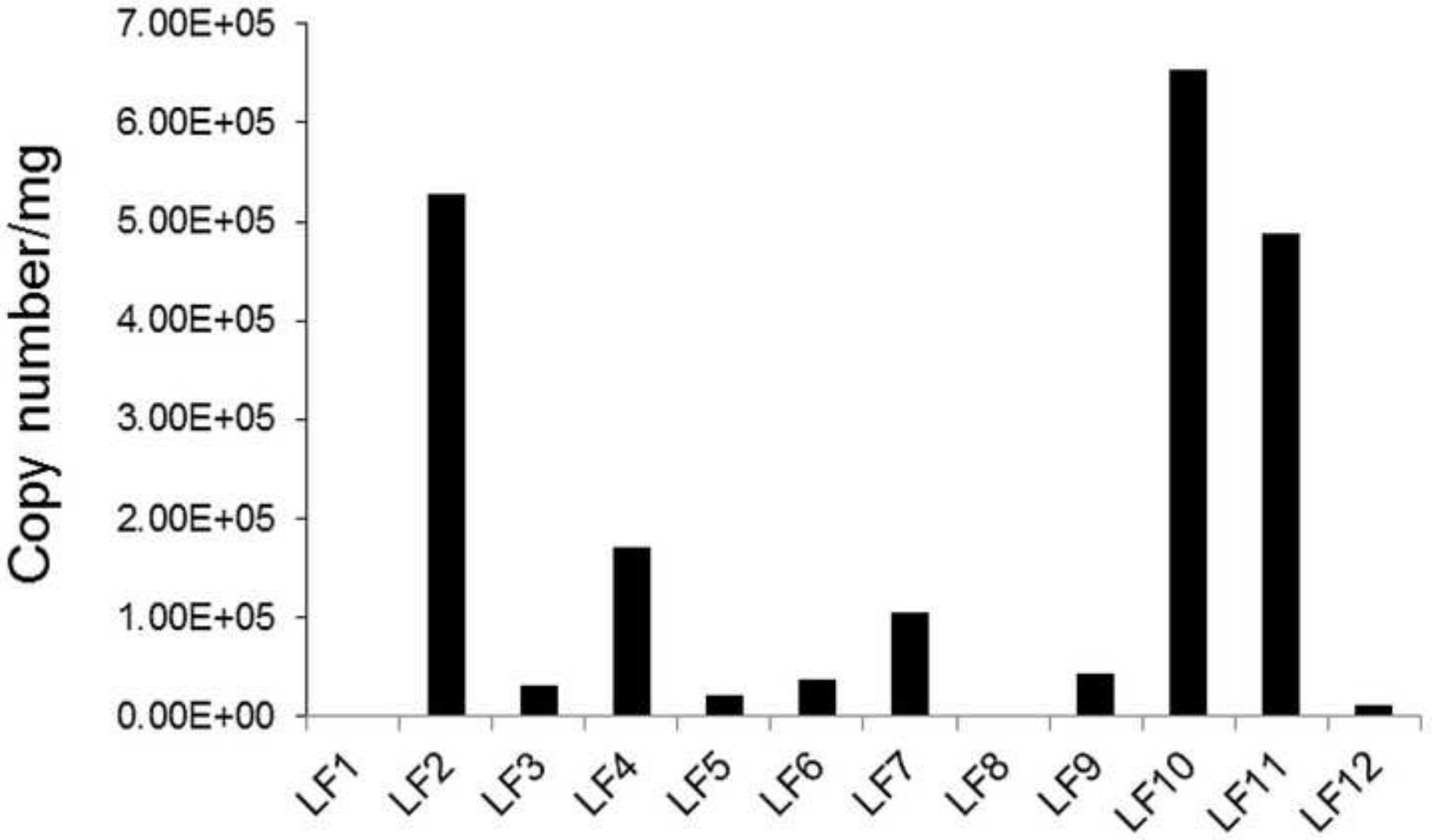
Figure

Click here to download high resolution image



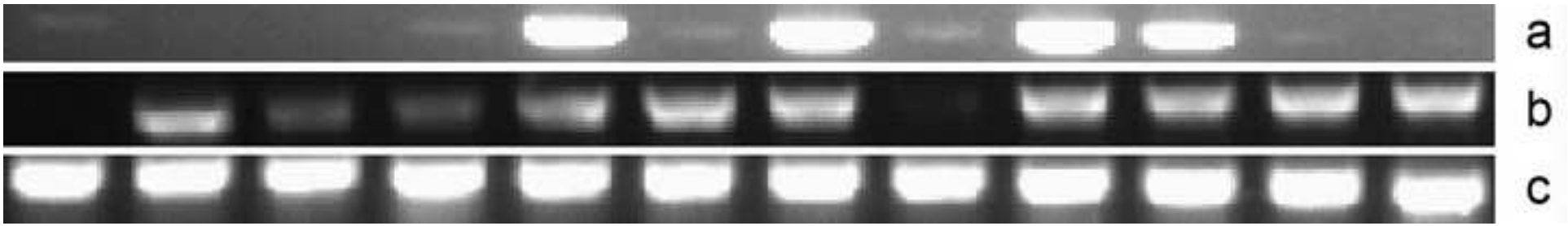
Figure

Click here to download high resolution image



Figure

Click here to download high resolution image



Figure

Click here to download high resolution image

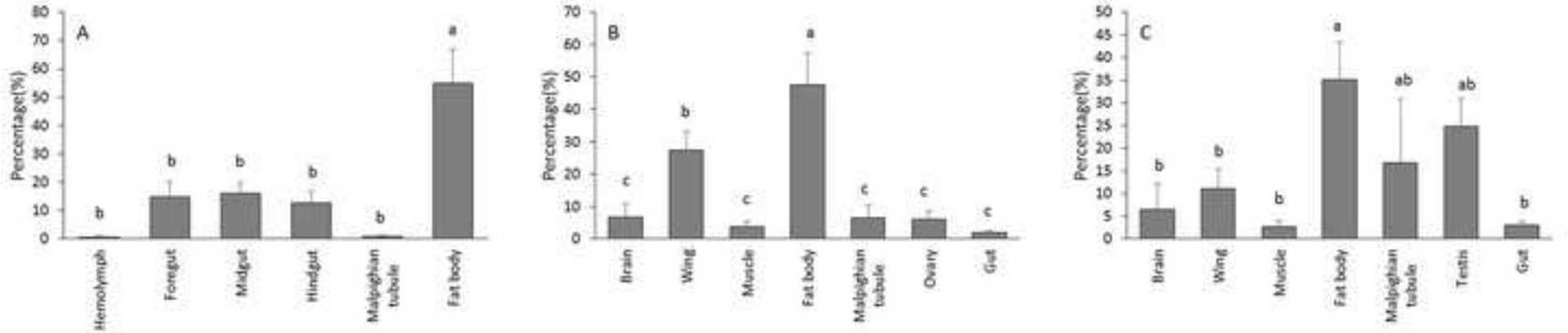
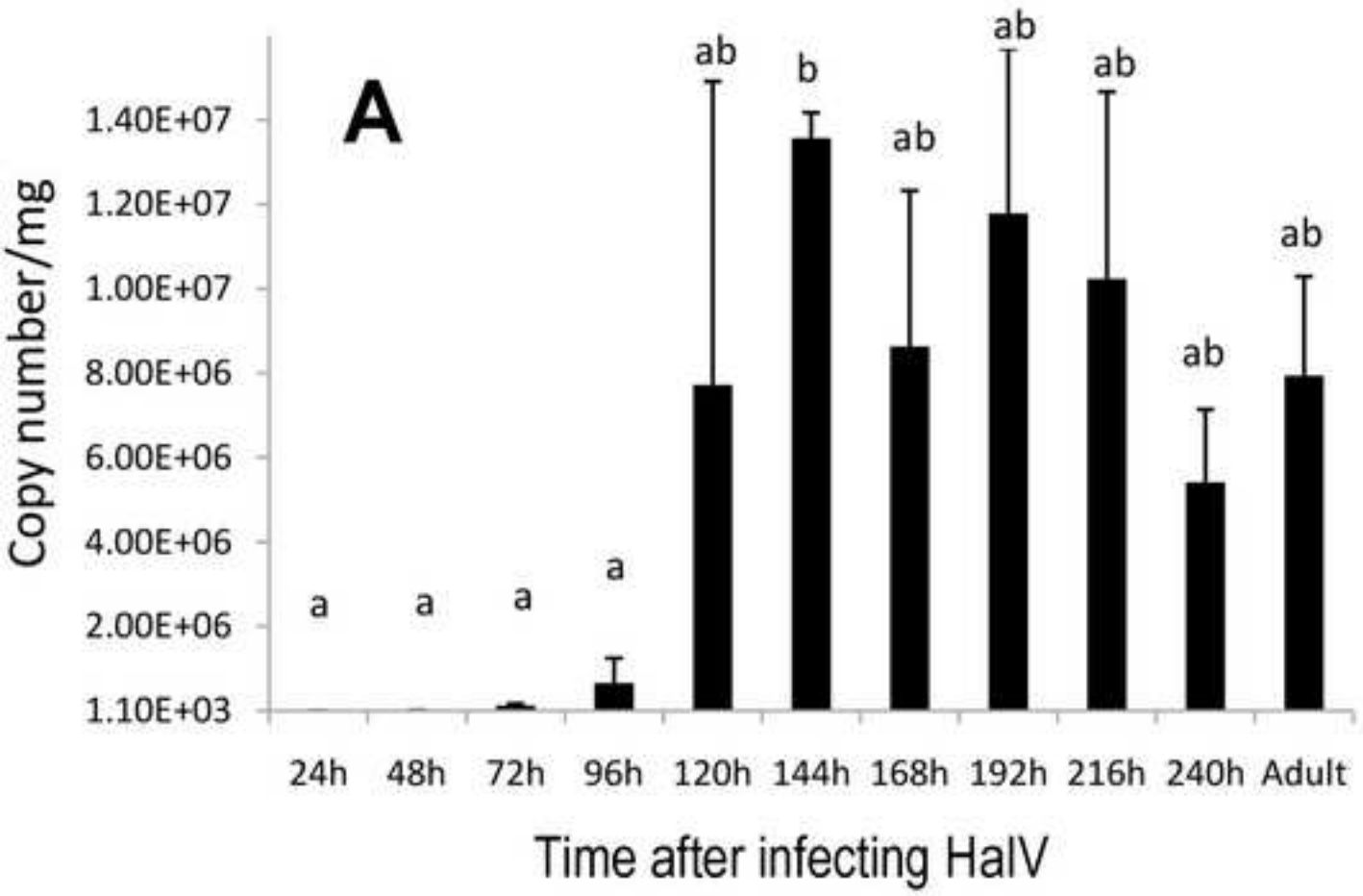
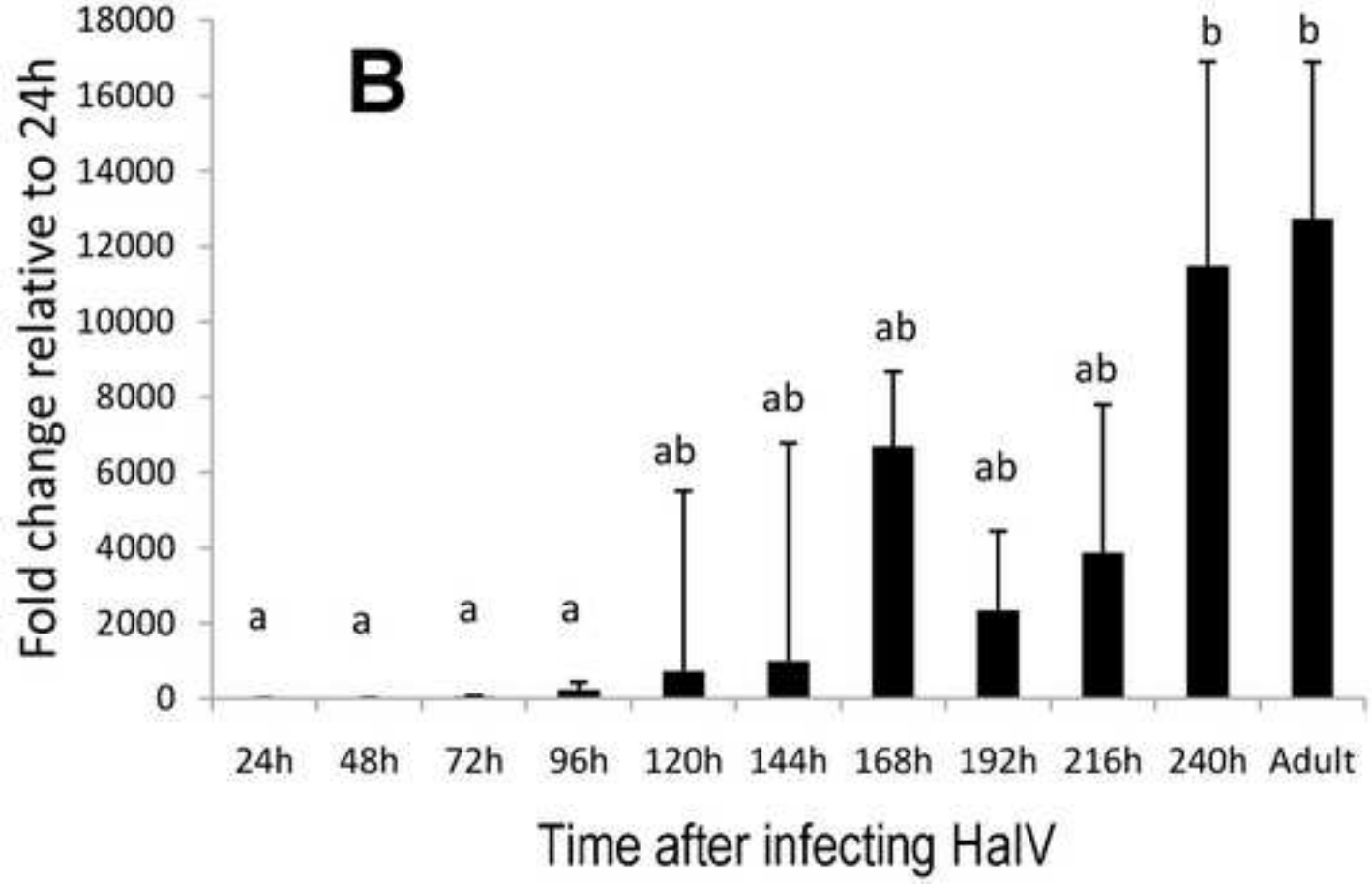


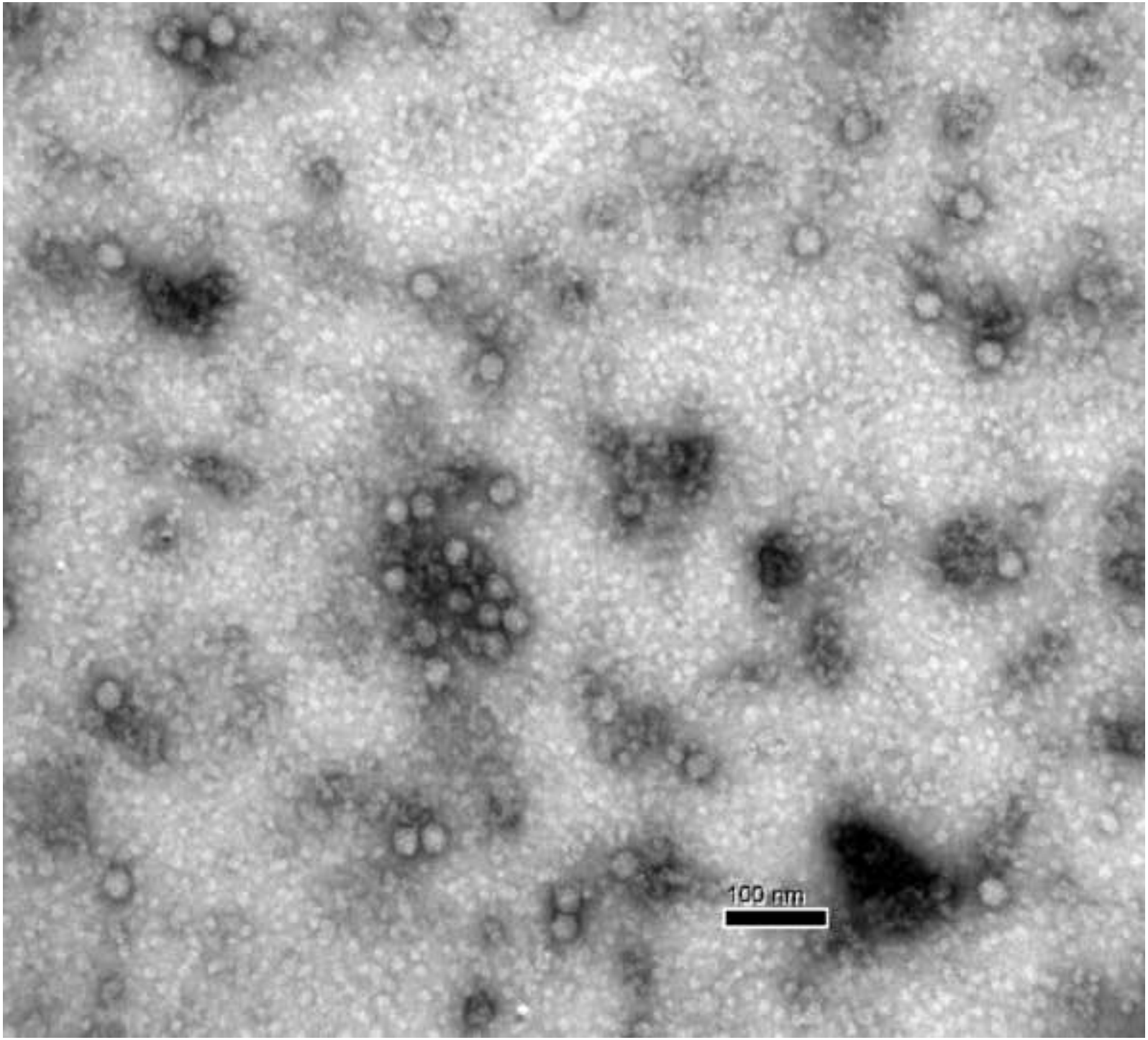
Figure
Click here to download high resolution image





Figure

Click here to download high resolution image



Supplementary Material

Click here to download Supplementary Material: Supporting information-revised.docx