

[Click here to view linked References](#)

1 **Structure and transcription of the *Helicoverpa armigera densovirus* (HaDV2)**
2 **genome and its expression strategy in LD652 cells**

3 Pengjun Xu^{1,2}, Robert I. Graham³, Kenneth Wilson⁴ and Kongming Wu^{1*}

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

6 ²Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao,
7 P.R. China

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

10 ⁴Lancaster Environment Centre, Lancaster University, Lancaster, UK

11 **Email address of all authors:** Pengjun Xu: xupengjun@163.com; Robert I. Graham:
12 rgraham@harper-adams.ac.uk; Kenneth Wilson: ken.wilson@lancaster.ac.uk;
13 Kongming Wu: kmwu@ippcaas.cn

14 ***Corresponding author: Kongming Wu**

15 State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of
16 Plant Protection, Chinese Academy of Agricultural Sciences, No. 2 West Yuan Ming
17 Yuan Road, Beijing 100193, P. R. China. E-mail: kmwu@ippcaas.cn.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31 **Abstract**

32 Background: Densoviruses (DVs) are highly pathogenic to their hosts. However, we
33 previously reported a mutualistic DV (HaDV2). Very little was known about the
34 characteristics of this virus, so herein we undertook a series of experiments to
35 explore the molecular biology of HaDV2 further.

36 Results: Phylogenetic analysis showed that HaDV2 was similar to members of the
37 genus *Iteradensovirus*. However, compared to current members of the genus
38 *Iteradensovirus*, the sequence identity of HaDV2 is less than 44% at the
39 nucleotide-level, and lower than 36%, 28% and 19% at the amino-acid-level of VP,
40 NS1 and NS2 proteins, respectively. Moreover, NS1 and NS2 proteins from HaDV2
41 were smaller than those from other iteradensoviruses due to their shorter N-terminal
42 sequences. Two transcripts of about 2.2 kb coding for the NS proteins and the VP
43 proteins were identified by Northern Blot and RACE analysis. Using specific
44 anti-NS1 and anti-NS2 antibodies, Western Blot analysis revealed a 78 kDa and a 48
45 kDa protein, respectively. Finally, the localization of both NS1 and NS2 proteins
46 within the cell nucleus was determined by using Green Fluorescent Protein (GFP)
47 labelling.

48 Conclusion: The genome organization, terminal hairpin structure, transcription and
49 expression strategies as well as the mutualistic relationship with its host, suggested
50 that HaDV2 was a novel member of the genus *Iteradensovirus* within the subfamily
51 *Densovirinae*.

52 Keywords: densovirus, HaDV2, expression, transcription
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

61 **Background**

62 The subfamily of *Densovirinae* within the family *Parvoviridae* is a group of small
63 (18–26 nm diameter), non-enveloped, icosahedral viruses containing a linear
64 single-stranded DNA genome ranging between 4–6 kb with characteristic terminal
65 hairpins [1-3]. Members of this subfamily typically produce “cellular dense
66 nucleosis” pathogenesis in their hosts, hence, they are commonly termed
67 densovirus (DVs) [4-8]. Since the first identification of a densovirus in the greater
68 wax moth *Galleria mellonella* [9], DVs have been isolated from many arthropods,
69 including species from six insect orders (Lepidoptera, Diptera, Orthoptera,
70 Dictyoptera, Odonata and Hemiptera) and decapod crustaceans (shrimps and crabs)
71 [10-12].

72 To date, many DVs have been identified and sequenced. Unlike vertebrate
73 parvoviruses, which all exhibit a monosense organization of their genome with
74 nonstructural protein (NS) and structural protein (VP) open reading frames (ORFs)
75 located on the same strand, arthropod DVs possess two types of genomes:
76 monosense and ambisense [13-18]. Previously, the taxonomy of DVs was ambiguous,
77 which was based on the organization of coding sequences, as well as genome size,
78 terminal hairpin structure, gene expression strategy and host range [19]. Under the
79 proposal of the International Committee on Taxonomy of Viruses (ICTV), Cotmore
80 et al. [2] reconstructed the taxonomy of the family *Parvoviridae* in which DVs were
81 classified into five distinct genera: *Ambidensovirus*, *Brevidensovirus*,
82 *Iteradensovirus*, *Hepandensovirus* and *Penstyldensovirus* according to phylogenetic
83 analysis and sequence homology.

84 DVs are highly pathogenic viruses to their hosts, and have been documented as
85 being transmitted both horizontally and vertically [7, 9, 16]. Traditionally, these
86 properties have captured the interest of many researchers investigating the potential
87 application of DVs as biopesticides for biological control of insect pests or vectors
88 for transgenic insects [20-26]. However, we previously reported a novel DV
89 displaying a mutualistic interaction with its host (*Helicoverpa armigera*), and named
90 this virus HaDV2 (previously named HaDENV-1) to distinguish it from the HaDV1

1 reported by El-Far *et al.* [27-29]. In this current study, we report the genome
2 organization, transcription and expression strategies of the virus HaDV2.
3

4 **Methods**

5 **Insect cell culture and transfection**

6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
95 *Lymantria dispar* LD652 cells, a gift from Central China Normal University in
96 Wuhan (China) [30], were cultured in Grace's insect medium containing 10% fetal
97 bovine serum (FBS) and 1% Penicillin-Streptomycin (Invitrogen, Grand Island, NY,
98 USA) at 28 °C. Purified plasmid (100 ng) containing the ORFs of NS1 and NS2 with
99 TIANpure Mini Plasmid Kit (TIANGEN, Beijing, China) was transfected into cells
100 using Cellfectin® II Reagent as recommended (Invitrogen). The luciferase activities
101 were determined using Luciferase Assay System (Promega, Madison, WI, USA).

102 **Sequence analysis**

103 Identity and alignment of the nucleotide and amino acid sequences was calculated
104 using CLUSTAL W software [31]. The ORFs were identified using the ORF Finder
105 (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Neighbor-joining trees with
106 Poisson-corrected distances for the DV nucleotide sequences and the amino acid
107 sequences of (NS1, NS2 and VP ORFs) were constructed using CLUSTAL W
108 software and MEGA6.0 software [32].

109 **Amplification of the stem-loop structure of HaDV2 genome by inverse PCR**

110 The viral DNA was extracted from purified virus particles using TIANamp Genomic
111 DNA Kit (TIANGEN). According to the reported genome sequence of HaDV2
112 (GenBank accession No.: HQ613271), three forward primers near the 3' end (DVF1
113 [nt 4576-4595], DVF2 [nt 3891-3910], DVF3 [nt 4343-4362]), and two reverse
114 primers near the 5' end (DVR1 [nt 1038-1057] and DVR2 [nt 832-889]) were
115 designed according to the genome sequence of HaDV2 (Table S1). PCR reactions
116 were performed using TransTaq DNA Polymerase High Fidelity (TransGen, Beijing,
117 China) and extracted viral DNA as a template. The PCR program was as follows: 30
118 s at 94 °C, 30 s at 57 °C, and 60 s at 72 °C for 40 cycles.

119 **Mapping of the transcripts by 5'/3' RACE and Northern blot**

120 The 5' and 3' ends of the HaDV2 transcripts were amplified using the SMART

1 121 RACE cDNA Amplification Kit (Clontech, CA, USA), according to the
2 122 manufacturer's instructions. cDNA was synthesized by RT-PCR from the total RNA
3 123 of migrating cotton bollworms infected by HaDV2 using primers NS3F1/UPM for
4 124 the 3' end, NS5R1/UPM and NS5R2/UPM for the 5' end of the NS genes, 3F1/UPM
5 125 for the 3' end, VP5R1/UPM and VP5R2/UPM for the 5' end of the VP genes,
6 126 respectively (Table S1). RNA (30 µg total) from insects infected by HaDV2 were
7 127 separated on 1.1% formaldehyde agarose gels using MOPS buffer and blotted onto a
8 128 positively charged nylon membrane (Roche, USA). Northern blot hybridization was
9 129 performed using DIG-labeled probes (DIG DNA Labeling and Detection Kit, Roche,
10 130 USA), according to the manufacturer's instruction. A 543 bp NS probe (1073-1615
11 131 nt) and a 420 bp VP probe (4081-4500 nt) were amplified by PCR with primers pairs
12 132 NSF/NSR and VPF/VPR (Table S1) using 30 cycles on a thermocycler as follows:
13 133 30 s at 95 °C, 30 s at 50 °C, and 30 s at 72 °C.

134 **Antibody production**

135 Using the predicted amino acid sequences gained from the earlier experiments in this
136 study, two polypeptides were synthesized to raise polyclonal antisera in rabbits:
137 CWDRAEFLRKYRKKVN and CDIGKSELWAPSVNPT for NS1 and NS2 proteins,
138 respectively. The polypeptide of NS1, NS2 or VP was each emulsified with an equal
139 volume of Freund's complete adjuvant for the first injection and incomplete adjuvant
140 for subsequent injections. Antisera were obtained by injecting an adult rabbit
141 subcutaneously with 500 µg polypeptide, followed by three additional injections of
142 300 µg polypeptide at 20 days intervals. The serum was purified and stored at -70 °C.
143 The titer of the antisera was measured using ELISA as described by Liu et al. [33].

144 **NS protein expression and subcellular localization**

145 To characterize the expression of HaDV2 NS proteins in insect cells, we constructed
146 two plasmids. Firstly, the HaDV2 NS promoter was amplified by primers
147 NSPF/NSPR (Table S1), digested with restriction endonuclease *KpnI/HindII* and
148 cloned into a luciferase reporter vector pGL-3 Basic (Promega). This created the
149 pNSP-Luc plasmid, in which the luciferase gene was under the control of the
150 HaDV2 promoter. Secondly, the complete ORFs of HaDV2 were amplified and

151 cloned into the pEASY-T Cloning Vector (TransGen) to create pHaDENV-T with the
152 primers HDVF1/HDVR1, HDVF2/HDVR2, HDVF3/HDVR3 (Table S1), and using
153 restriction endonucleases *Sac*II, *Afl*III and *Bsp*1407I, The resulting plasmid
154 pHaDENV-T contained the whole genome of HaDV2 except for the hairpin structure.
155 LD652 cells transfected with pHaDENV-T plasmid were then analyzed using 12%
156 SDS-PAGE and transferred onto PVDF membranes. The membranes were blocked
157 with dry skimmed milk (5%) and incubated in PBST buffer (PBS containing 0.1%
158 Tween-20) containing polyclonal antibodies (anti-VP, anti-NS1 or anti-NS2) (1:5000,
159 1 h) and then a horseradish peroxidase (HRP)-conjugated secondary antibody
160 (ZSGB-BIO, China) (1:20000, 1 h). The blots were revealed using the Easysee
161 Western Blot Kit (Transgen).

162 To investigate the subcellular localization of these proteins, the NS1 and NS2
163 ORFs were amplified with primers NS1LF/NS1LR and NS2LF/NS2LR (Table S1),
164 then cloned into plasmid pIE-Atg6-GFP [34] by exchanging BmAtg6 with these two
165 ORFs to create NS1-GFP and NS2-GFP constructs, respectively. The constructs were
166 then transfected into LD652 cells and the NS-GFP fusion proteins were examined by
167 fluorescent microscopy 24 h post-transfection. The IE2-GFP plasmid which was
168 constructed by inserting the OpIE2 promoter of the pIZ-V5/His (Invitrogen) into the
169 plasmid pEGFP-N1 digested by restriction enzymes *Bgl* II and *Sac* I was used as
170 control and obtained as a gift from Dr. Liu's lab in Central China Normal University
171 (China).

172 **Results**

173 **Nucleotide sequences based analysis of HaDV2 genome organization**

174 The size of virus particles of HaDV2 is about 20 nm in diameter and it possesses a
175 monosence genome about 5 kb. The sequencing analysis of the HaDV2 genome
176 reported in our previous study showed that it differed considerably from other
177 known DVs (Fig. S1) [28]. HaDV2 contained three large ORFs on the same strand
178 ORF1 (1260 nt in length) encoded the putative NS2 polypeptide of 419 amino acids
179 with a predicted molecular mass of 48 kDa and a theoretical isoelectric point (pI) of
180 7.10. ORF2 (2010 nt in length) encoded the putative NS1 protein of 669 amino acids

181 with a predicted molecular mass of 78 kDa and a pI of 5.87. Alignment of the
182 HaDV2 NS1 amino acid sequence with three other iteradensoviruses indicated that
183 HaDV2 NS1 protein shared two functional domains. Firstly, amino acids between aa
184 258-311 contained the two highly conserved replication initiator motifs involved in
185 initiation and termination of rolling circle replication; and secondly, amino acids
186 from aa 514-635 contained the NTP-binding and helicase domains typical of the
187 NS1 polypeptide. Alignment analysis indicated that the amino acid sequences of
188 NS1 and NS2 from the HaDV2 were shorter than those from closely related
189 members of the genus *Iteradensovirus* (Fig. S2). ORF3 (1980 nt in length) encoded a
190 putative VP protein of 659 amino acids with a molecular mass of 75 kDa and a pI of
191 7.13. The highly conserved phospholipase A2 domain located at amino acid
192 positions aa 2-56 was also found in the VP ORF of HaDV2.

193 **Tree-based analysis**

194 Neighbor-joining trees were constructed using the whole genome sequence of
195 HaDV2 as well as the putative amino acid sequences of NS1, NS2 and VP ORFs.
196 Thirty-one DVs from other arthropods that had either the complete genome or the
197 full coding sequence were included in the analysis (Fig. 1). The trees revealed five
198 main branches: branch 1 included all known DV members with an ambisense
199 genome in the genus *Ambidensovirus*; branch 2 included members of the genus
200 *Iteradensovirus* and HaDV2; branch 3 included members of the genus
201 *Hepandensovirus*; branch 4 included members of the genus *Penstyldensovirus*; and
202 branch 5 consisted of members of the genus *Brevidensovirus* (Fig. 1a). The branches
203 of the tree constructed with amino acid sequences of VP was similar to that of the
204 nucleotide sequence tree (Fig. 1b). The trees constructed with amino acid sequences
205 of NS1 and NS2 differed considerably from the trees described above (see Fig. 1c,
206 1d). Although the four trees were not identical, they all indicated that the HaDV2
207 was most closely related to the members of the genus *Iteradensovirus*. We also
208 reconstructed the trees using Maximum likelihood method and the trees showed
209 similar topology with the NJ trees except for NS2-tree (Fig. S3). Alignment of the
210 nucleotide sequence and amino acid sequences (of VP, NS1 and NS2) indicated that

1 211 sequence identities between viruses within the genus *Iteradensovirus* exceed 58%,
2 212 71%, 35% and 28%, respectively; and that the identities between HaDV2 and
3 213 members of the genus *Iteradensovirus* are no more than 44%, 36%, 28% and 19%,
4 214 respectively.

5 215 **Determination of the stem-loop structure of HaDV2 genome by inverse PCR**

6 216 Sequence analysis indicated that the HaDV2 terminal-ends could form a stem-loop
7 217 structure by a reverse complementary sequence (ITRs) near the two ends (Fig. 2a).
8 218 To further confirm this prediction, PCR successfully amplified segments with
9 219 forward primers near the 3' end and reverse primers near the 5' end. Sequence
10 220 alignment indicated that all the amplified sequences were consistent with those of
11 221 the HaDV2 except for the stem region located at ITRs (Fig. 2b).

12 222 **Transcript analysis of HaDV2**

13 223 Using 5' and 3' RACE primers, two transcription initiation sites (TISs) (at positions
14 224 nt 307 (Figure 3b) and nt 2516 (Fig. 3c)) and two transcription termination sites
15 225 (TTSs) (at positions nt 2516 (Figure 3d) and nt 4688 (Fig. 3e)) were determined. The
16 226 TTS for the NS genes occurred at nt 2516, with canonical poly(A) addition sites
17 227 (AATAAA) located 16 nucleotides upstream from the TTS (Fig. 4d). The VP TTS
18 228 was located at nt 4688, with canonical poly(A) addition sites situated 18 nucleotides
19 229 upstream from the TTS (Fig. 4e). The transcript of 2209 bp ending at nt 2516 may
20 230 encode the NS1 and NS2 proteins. The transcripts of 2173 bp ending at nt 4688 may
21 231 encode the VP proteins (Fig. 4a).

22 232 Northern Blot analysis was undertaken to determine the size and relative
23 233 abundance of the transcripts for each of the three viral ORFs. Hybridization of total
24 234 RNA isolated from infected insects revealed one band of 2.2 kb when using the NS
25 235 probe; and one band of 2.2 kb when using the VP probe (Fig. S4a, S4b).

26 236 **The expression and subcellular localization of NS1 and NS2 proteins of HaDV2** 27 237 **in LD652 cells**

28 238 The expression of HaDV2 NS proteins was undertaken in *Lymantria dispar* LD652
29 239 cells. To determine the functionality of the NS promoter, the NS promoter-luciferase
30 240 construct pNSP-Luc was transfected into LD652 cells and luciferase activity was

1 241 measured 24h post-transfection. Our results showed that the luciferase activity
2 242 driven by the NS promoter was approximately 225 times higher than that of the
3 243 promoterless control vector pGL3-Basic (Fig. 4a), indicating that the transcription
4 244 machinery of the LD652 cells recognized the NS promoter leading to expression of
5 245 HaDV2 NS proteins. Western blot analyses of pHaDV-T transfected LD652 cells
6 246 using antisera prepared against NS1 and NS2 proteins (Fig. S5) revealed two
7 247 proteins of 78 kDa (NS1) and 48 kDa (NS2) (Fig. 4b). Both the NS1 and NS2
8 248 GFP-fused proteins were exclusively present within the nucleus of the host cells,
9 249 suggesting NS1 and NS2 might localize within the nucleus (Fig. 5).

10 250 **Discussion**

11 251 DVs are a group of viruses usually associated with causing high pathogenicity to
12 252 their hosts [7, 9, 12]. However, we previously reported a novel DV (HaDV2) which
13 253 was found to be beneficial to its host by increasing larval and pupal developmental
14 254 rate, fertility, adult female lifespan and enhancing host resistance to both a
15 255 baculovirus and low doses of the Bt toxin [28, 29]. This suggested a virus with quite
16 256 different characteristics to the other previously described members within subfamily
17 257 Densovirinae. In this current study, we determined the molecular biology of the
18 258 HaDV2 virus, namely through examining its genome structure and ORF
19 259 transcription and expression strategy. Based on our results, HaDV2 was a novel
20 260 member of genus *Iteradensovirus*, with new features differing from other members
21 261 from this genus, such as an ITR of 101 nt at both termini, a single 90 nts hairpin
22 262 structure at the 3' end and the first ORF encoding NS2 protein [17, 28, 35-40].

23 263 Phylogenetic analysis using both nucleotide and amino acid sequences showed
24 264 that HaDV2 was clustered within the genus *Iteradensovirus*. The sequence identities
25 265 of the viral DNA and the amino acid identities for VP, NS1 and NS2 ORFs among
26 266 members of the genus *Iteradensovirus* exceed 58%, 71%, 35% and 28%,
27 267 respectively. However, the sequence identities between HaDV2 and the current
28 268 members of the genus *Iteradensovirus* are no more than 44%, 36%, 28% and 19%,
29 269 respectively. Thus, although the HaDV2 was clustered with *Iteradensovirus*, it
30 270 differs considerably from the other iteradensoviruses and appears to have a different

1 271 function as described previously [29].

2 272 Although phylogenetic analysis indicated that HaDV2 was clustered with
3
4 273 members of the genus *Iteradensovirus*, the NS1 and NS2 proteins of the HaDV2 are
5
6 274 smaller than those of other *Iteradensovirus* (more than 753 and 451 amino acids,
7
8 275 respectively) [17, 28, 35-40]. Are they functionally expressed as predicted? We used
9
10 276 Western Blot analysis of transfected LD652 cells using anti-NS1 and anti-NS2 to
11
12 277 show that the NS1 and NS2 proteins were 78 kDa and 48 kDa, respectively,
13
14 278 consistent with the predicted size of the protein. NS proteins are a pivotal factor for
15
16 279 viral transcription and replication as well as pathogenicity. The replication of DVs
17
18 280 occurred in the nucleus of their hosts [19]. Therefore, the NS proteins of DVs should
19
20 281 be located in the nucleus by nuclear localization signal (NLS) as reported by Yu *et al.*
21
22 282 [40]. To further investigate whether the NS proteins of the HaDV2 localized within
23
24 283 the nucleus of their hosts (as those of other DVs), NS1 and NS2 proteins were
25
26 284 expressed in LD652 cells using the recombinant plasmid NS1-GFP and NS2-GFP.
27
28 285 The result indicated that the NS1 and NS2 proteins were completely located in the
29
30 286 nucleus, suggesting that they possess a common function and could possibly play a
31
32 287 role in the novel interactions between HaDV2 and its host. The experiments with the
33
34 288 NS proteins were carried out by transient expression in LD652 cells, which were not
35
36 289 the virus's original host. It is acknowledged that this expression may not reflect the
37
38 290 real role of HaDV2-NS promoter and how it works in the natural host.

41 291 Transcriptional patterns are diverse among the DVs. For example, JcDV,
42
43 292 *Galleria mellonella densovirus* (GmDV) and *Mythimna loreyi densovirus* (MIDV)
44
45 293 all have one transcript for the VP gene and two transcripts for the NS genes (the
46
47 294 larger one for NS1 and the smaller one for NS2), in which the ORFs of NS1 and
48
49 295 NS2 share a common TTS [15, 30, 41]. Meanwhile, the transcripts of CpDV,
50
51 296 *Periplaneta fuliginosa densovirus* (PfdV) and *Myzus persicae nicotianae densovirus*
52
53 297 (MpnDV) arise from alternative splicing [13, 42, 43]. The first ORFs of all known
54
55 298 iteradensoviruses encode NS1 protein and the ORFs of NS2 are completely included
56
57 299 in the ORFs of NS1 [40]. However, the first ORF of HaDV2 encodes NS2, which
58
59 300 may impact gene expression of NS2 compared to NS1. In addition, the NS1 and NS2
60
61

1 301 of other iteradensoviruses were translated from different transcripts and the TIS of
2 302 NS1 was found to start 2-26 nt upstream of the start codon [40]. Unexpectedly, our
3 303 results suggested that the NS1 and NS2 of HaDV2 translated from the same
4 304 transcript which started 63 nt upstream of the start codon of NS2. Surprisingly,
5 305 although we provide evidence of the activity of the NS promoter, we failed to find
6 306 the TATA-box upstream of the TIS of NS. Two TATA-box like sequences were
7 307 located at nts 313 and 335 upstream of the start codon of NS1 and NS2; suggesting
8 308 HaDV2, maybe like brevidensoviruses, has overlapping NS gene promoters
9 309 responsible for different transcript starts and dictating the relative transcription rates
10 310 of these transcripts. However, one of the two transcripts was in great excess, making
11 311 it difficult to detect both transcripts by RACE. Like other DVs, the VP transcripts of
12 312 HaDV2 had short-untranslated regions, located at 5 nts upstream of the start codon
13 313 of VP.

27 314 **Conclusion**

28 315 We report a novel densovirus, assigned as HaDV2, which differs from the other DVs
29 316 in its genome organization, terminal hairpin structure, and transcription and
30 317 expression strategies. Taken together with the unique mutualistic relationship
31 318 previously described between HaDV2 and its host [29], this strongly indicates that
32 319 HaDV2 is a novel member within the genus *Iteradensovirus*.

41 321 **Abbreviation**

42 322 DV: densovirus

43 323 ORF: open reading frame

44 324 ICTV: the International Committee on Taxonomy of Viruses

45 325 pI: theoretical isoelectric point

46 326 NS: nonstructural protein

47 327 VP: structural protein

58 329 **Ethics approval and consent to participate**

59 330 No permit was required to collect the tested the virus in insect. Sampling did not involve regulated,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

331 endangered or protected species.

332

333 **Consent for publication**

334 Not applicable.

335

336 **Competing interests**

337 The authors declare that they have no competing interests.

338

339 **Funding**

340 This research was supported by Science Fund for Creative Research Groups of the National Science
341 Foundation of China (Grant No. 31321004), the National Natural Science Foundation of China (Grant
342 No. 31401752).

343

344 **Authors' contributions**

345 KWu and PX conceived the study. PX performed the experiments. KWu, PX, RIG and KWi wrote the
346 manuscript. All of the authors critically reviewed and approved the final manuscript.

347

348 **Acknowledgements**

349 We would like to thank Dr. Jie Wang (Tobacco Research Institute, Chinese Academy of Agricultural
350 Sciences, Qingdao, P.R. China) and Dr. Peng Xu (Central China Normal University, Wuhan, P.R. China)
351 for providing suggestions and repairing figures.

352

353 **Author information**

354 ¹State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection,
355 Chinese Academy of Agricultural Sciences, No. 2 West Yuan Ming Yuan Road, Beijing 100193, P.R.
356 China. ²Tobacco Research Institute, Chinese Academy of Agricultural Sciences, No. 11 Ke Yuan Jing
357 Si Road, Qingdao 266101, P.R. China. ³Crop and Environment Sciences, Harper Adams University,
358 Edgmond, Shropshire TF10 8NB, UK. ⁴Lancaster Environment Centre, Lancaster University,
359 Lancaster, LA1 4YQ, UK

360

1 361

2 362

3
4 363 **References**

5
6 364 1. Berns KI, Bergoin M, Bloom M, Lederman M, Muzyczka N, Siegl G et al.

7
8 365 Parvoviridae: Vith report of international committee on taxonomy of viruses.

9
10 366 Arch Virol 1995;10 Suppl:169–178.

11
12 367 2. Cotmore SF, Agbandje-McKenna M, Chiorini JA, Mukha DV, Pintel DJ, Qiu J,

13
14 368 et al. The family Parvoviridae. Arch Virol. 2014;159:1239–1247.

15
16 369 3. Dhar AK, Robles-Sikisaka R, Saksmerprome V, Lakshman DK. Biology,

17
18 370 genome organization, and evolution of parvoviruses in marine shrimp. Adv Virus

19
20 371 Res. 2014; 89:85–139.

21
22 372 4. Amargier A, Vago C, Meynadier G. Histopathological study of a new type of

23
24 373 viral disease demonstrated in the Lepidoptera *Galleria mellonella*. Arch Gesamte

25
26 374 Virusforsch. 1965;15:659–667.

27
28 375 5. Hu Y, Zheng J, Lizuka T, Bando H. A densovirus newly isolated from the

29
30 376 smoky-brown cockroach *Periplaneta fuliginosa*. Arch Virol.

31
32 377 1994;138:365–372.

33
34 378 6. O'Neill SL, Kittayapong P, Braig HR, Andreadis TG, Gonzalez JP, Tesh RB.

35
36 379 Insect densoviruses may be widespread in mosquito cell lines. J Gen Virol.

37
38 380 1995;76:2067–2074.

39
40 381 7. Rivers CF, Longworth JF. A nonoccluded virus of *Junonia coenia* (Nymphalidae:

41
42 382 Lepidoptera). J Invertebr Pathol. 1972; 20:369–370.

43
44 383 8. Vago C, Quiot JM, Luciani J. Infection of lepidopteran tissue culture with the

45
46 384 purified virus of the denonucleus. C R Acad Sci Hebd Seances Acad Sci D.

47
48 385 1966;263:799–800.

49
50 386 9. Meynardier G, Vago C, Atger P. Virose d'un type inhabituel chez le lepidoptere

51
52 387 *Galleria mellonella* L. Rev Zool Agric Appl. 1964 ;63:207–209.

53
54 388 10. Fediere G. Epidemiology and pathology of *Densovirinae*. Contrib Microbiol.

55
56 389 2000;4:1–11.

57
58 390 11. Safeena MP, Tyagi A, Rai P, Karunasagar I. Complete nucleic acid sequence of

- 1 391 *Penaeus monodon densovirus* (PmDENV) from India. *Virus Res.* 2010;150:1–11.
- 2 392 12. van Munster M, Dulleman AM, Verbeek M, van den Heuvel JF, Reinbold C,
3
4 393 Brault V, et al. Characterization of a new densovirus infecting the green peach
5
6 394 aphid *Myzus persicae*. *J Invertebr Pathol.* 2003;84:6–14.
- 7
8 395 13. Baquerizo-Audiot E, Abd-Alla A, Jousset FX, Cousserans F, Tijssen P, Bergoin
9
10 396 M. Structure and expression strategy of the genome of *Culex pipiens densovirus*,
11
12 397 a mosquito densovirus with an ambisense organization. *J Virol.*
13
14 398 2009;83:6863–6873.
- 15
16 399 14. Guo H, Zhang J, Hu Y. Complete sequence and organization of *Periplaneta*
17
18 400 *fuliginosa densovirus* genome. *Acta Virol.* 2000;44:315–322.
- 19
20 401 15. Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zadori Z. Organization and
21
22 402 expression strategy of the ambisense genome of denonucleosis virus of *Galleria*
23
24 403 *mellonella*. *J Virol.* 2003;77:10357–10365.
- 25
26 404 16. van Munster M, Dulleman AM, Verbeek M, van den Heuvel JF, Reinbold C,
27
28 405 Brault V, et al. A new virus infecting *Myzus persicae* has a genome organization
29
30 406 similar to the species of the genus *Densovirus*. *J Gen Virol.* 2003;84:165–172.
- 31
32 407 17. Wang J, Zhang J, Jiang H, Liu C, Yi F, Hu Y. Nucleotide sequence and genomic
33
34 408 organization of a newly isolated densovirus infecting *Dendrolimus punctatus*. *J*
35
36 409 *Gen Virol.* 2005;86:2169–2173.
- 37
38 410 18. Zhai YG, Lv XJ, Sun XH, Fu SH, Gong ZD, Fen Y, et al. Isolation and
39
40 411 characterization of the full coding sequence of a novel densovirus from the
41
42 412 mosquito *Culex pipiens pallens*. *J Gen Virol.* 2008;89:195–199.
- 43
44 413 19. Bergoin M, Tijssen P. Molecular biology of *Densovirinae*. *Contrib Microbiol.*
45
46 414 2000;4:12–32.
- 47
48 415 20. Bossin H, Fournier P, Royer C, Barry P, Cerutti P, Gimenez S, et al. *Junonia*
49
50 416 *coenia* densovirus-based vectors for stable transgene expression in Sf9 cells:
51
52 417 influence of the densovirus sequences on genomic integration. *J Virol.*
53
54 418 2003;77:11060–11071.
- 55
56 419 21. Corsini J, Afanasiev B, Maxwell IH, Carlson JO. Autonomous parvovirus and
57
58 420 densovirus gene vectors. *Adv Virus Res.* 1996;47:303–351.
- 59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- 421 22. Giraud C, Devauchelle G, Bergoin M. The densovirus of *Junonia coenia* (Jc
422 DNV) as an insect cell expression vector. *Virology*. 1992;186:207–218.
- 423 23. Hu L, Zhang L, Shen C, Lu J, Zhang J, Hu Y. The densovirus of *Periplaneta*
424 *fuliginosa* (PfdNV) as an insect vector for persistent foreign gene expression in
425 vivo. *Biochem Biophys Res Commun*. 2007;358:976–982.
- 426 24. Jiang H, Zhang J, Wang J, Yang B, Liu C, Lu J, et al. Genetic engineering of
427 *Periplaneta fuliginosa* densovirus as an improved biopesticide. *Arch Virol*.
428 2007;152:383–394.
- 429 25. Mutuel D, Ravallec M, Chabi B, Multeau C, Salmon JM, Fournier P, et al.
430 Pathogenesis of *Junonia coenia* densovirus in *Spodoptera frugiperda*: a route of
431 infection that leads to hypoxia. *Virology*. 2010; 403:137–144.
- 432 26. Suzuki Y, Niu G, Hughes GL, Rasgon JL. A viral over-expression system for the
433 major malaria mosquito *Anopheles gambiae*. *Sci Rep*. 2014;4:5127.
- 434 27. El-Far M, Szelei J, Yu Q, Fediere G, Bergoin M, Tijssen P. Organization of the
435 ambisense genome of the *Helicoverpa armigera* densovirus. *J Virol*.
436 2012;86:7024.
- 437 28. Xu P, Cheng P, Liu Z, Li Y, Murphy RW, Wu K. Complete genome sequence of a
438 monosense densovirus infecting the cotton bollworm, *Helicoverpa armigera*. *J*
439 *Virol*. 2012;86:10909.
- 440 29. Xu P, Liu Y, Graham RI, Wilson K, Wu K. Densovirus is a mutualistic symbiont
441 of a global crop pest (*Helicoverpa armigera*) and protects against a baculovirus
442 and Bt biopesticide. *PLOS Pathog*. 2014;10:e1004490.
- 443 30. Wang Y, Abd-Alla AM, Bossin H, Li Y, Bergoin M. Analysis of the transcription
444 strategy of the *Junonia coenia* densovirus (JcDNV) genome. *Virus Res*.
445 2013;174:101–107.
- 446 31. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity
447 of progressive multiple sequence alignment through sequence weighting,
448 position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*.
449 1994;22:4673–4680.
- 450 32. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular

- 1 451 evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30:2725–2729.
- 2 452 33. Liu C, Gao Y, Ning C, Wu K, Oppert B, Guo Y. Antisera-mediated in vivo
- 3 453 reduction of Cry1Ac toxicity in *Helicoverpa armigera*. *J Insect Physiol.*
- 4 454 2010;56:718–724.
- 5 455 34. Wei W, Gai Z, Ai H, Wu W, Yang Y, Peng J, et al. Baculovirus infection triggers
- 6 456 a shift from amino acid starvation-induced autophagy to apoptosis. *PLoS One.*
- 7 457 2012;7:e37457.
- 8 458 35. Fediere G, Li Y, Zadori Z, Szelei J, Tijssen P. Genome organization of *Casphalia*
- 9 459 *extranea densovirus*, a new iteravirus. *Virology.* 2002;292:299–308.
- 10 460 36. Li Y, Zadori Z, Bando H, Dubuc R, Fediere G, Szelei J, et al. Genome
- 11 461 organization of the densovirus from *Bombyx mori* (BmDENV-1) and enzyme
- 12 462 activity of its capsid. *J Gen Virol.* 2001;82:2821–2825.
- 13 463 37. Yu Q, Fediere G, Abd-Alla A, Bergoin M, Tijssen P. Iteravirus-like genome
- 14 464 organization of a densovirus from *Sibine fusca* Stoll. *J Virol.*
- 15 465 2012;86:8897–8898.
- 16 466 38. Yu Q, Hajek AE, Bergoin M, Tijssen P. *Papilio polyxenes densovirus* has an
- 17 467 iteravirus-like genome organization. *J Virol.* 2012;86:9534–9535.
- 18 468 39. Yu Q, Tijssen P. *Iteradensovirus* from the monarch butterfly, *Danaus plexippus*
- 19 469 *plexippus*. *Genome Announc.* 2014;2:e00321–14.
- 20 470 40. Yu Q, Tijssen P. Gene expression of five different iteradensoviruses: *Bombyx*
- 21 471 *mori densovirus*, *Casphalia extranea densovirus*, *Papilio polyxenes densovirus*,
- 22 472 *Sibine fusca densovirus*, and *Danaus plexippus densovirus*. *J Virol.*
- 23 473 2014;88:12152–12157.
- 24 474 41. Fediere G, El-Far M, Li Y, Bergoin M, Tijssen P. Expression strategy of
- 25 475 denso-nucleosis virus from *Mythimna loreyi*. *Virology.* 2004;320:181–189.
- 26 476 42. Yamagishi J, Hu Y, Zheng J, Bando H. Genome organization and mRNA
- 27 477 structure of *Periplaneta fuliginosa densovirus* imply alternative splicing
- 28 478 involvement in viral gene expression. *Arch Virol.* 1999;144:2111–2124.
- 29 479 43. Tang S, Song X, Xue L, Wang X, Wang X, Xu P, et al. Characterization and
- 30 480 distribution analysis of a densovirus infecting *Myzus persicae nicotianae*

481 (Hemiptera: Aphididae). J Econ Entomol. 2016;109:580–587.

482

483 **Figure legends**

484 **Fig. 1** Neighbor-joining phylogenetic trees for members of the densoviruses,
485 including (a) the genomic sequence,(b) the amino acid sequence of the VP ORF, (c)
486 the amino acid sequence of the NS1 ORF, and (d) the amino acid sequence of the
487 NS2 ORF. Other accession numbers are given after the abbreviated names. ▲ =
488 HaDV2. Bootstrap values (1000 pseudoreplicates) > 50% are indicated on the nodes.

489 **Fig. 2** Determination of the stem-loop structure with the HaDV2 sequence. (a) The
490 predicted stem-loop structure of HaDV2 genome. Nucleotides of the 3' terminal
491 hairpin and reverse complementary sequences are shown. Numbers indicate
492 locations of the nucleotides on the HaDV2 genome (5'-3'). The primers used for
493 determining the stem-loop structure were shown, including three forward primers
494 DVF1, DVF2, DVF3 and two reverse primers DVR1, DVR2. (b) Alignment of
495 terminal nucleotide sequences from different clones to confirm the stem-loop
496 structure. Clone names shown on the left indicate their amplification primers and
497 clone numbers (e.g., clone DVF1-DVR1-1 stands for one clone amplified by primers
498 DVF1/DVR1). The numbers show the nucleotide location on the genome of HaDV2.
499 "The stem region" indicates the double DNA region as show in (a). "." = identical
500 nucleotides; "-" = absence of nucleotides.

501 **Fig. 3** Transcriptional analysis of HaDV2 by RACE. (a) The putative HaDV2
502 transcripts were shown according to the results of 3' and 5' RACE. The length of
503 each transcript is indicated on the left except for the nucleotides of poly(A) (wavy
504 lines). (b) and (c) Analysis of the transcriptional initiation site (TIS) by 5' RACE
505 with NS and VP primers, respectively. (d), (e) Analysis of the transcriptional
506 terminal site (TTS). The TIS, TTS and adaptor sequence are indicated by arrows.
507 The number of TISs, TTSs and the consensus sequence are shown (HaDV2). "." =
508 identical nucleotides; "-" = absence of nucleotides.

509 **Fig. 4** The expression of NS proteins in LD652 cells. (a) Analysis of the
510 transcriptional activity of the NS promoter of HaDV2 in LD652 cells.(b) Western

1 511 blot analysis of the proteins expressed by HaDV2 with anti-NS1 and anti- NS2
2 512 antibodies using LD652 cells transfected with the pHaDENV-T plasmid.

3
4 513 **Fig. 5** Subcellular localization of IE2-GFP, NS1-GFP and NS2-GFP fusion proteins
5 514 in LD652 cells (x400), indicating the proteins of NS1 and NS2 were located to the
6 515 nucleus.

7
8
9 516

10 11 12 517 **Additional file**

13 518 **Table S1** Primers used in this study. **Fig. S1** The virus particles and its genome organization. (a)

14 519 Electron micrograph of HaDENV-1 viruses purified from adult *Helicoverpa armigera* negatively

15 520 stained with uranyl acetate (×200000). Bar, 100nm. (b) Agarose gel electrophoresis (1%) of the

16 521 extracted HaDENV-1 DNA. Lane 1=DNA from the HaDENV-1, Lane 2= Marker. (c) The putative ORFs

17 522 of HaDENV-1. The plus strand contains three large ORFs: ORF1, ORF2 and ORF3, which encode NS2,

18 523 NS1 and VP proteins, respectively. (d) Hairpin structure in the 3' terminus of HaDENV-1 predicted by

19 524 the QuickFOLD program. The numbers in bracket stand for the start and stop nucleotides of the

20 525 hairpin on the HaDENV-1 genome. **Fig. S2** Alignment of amino acid sequences of NS1 (a) and NS2 (b)

21 526 of HaDENV-1 with the ones of members from *Iteravirusdensovirus*. HaDENV=*Helicoverpa armigera*

22 527 densovirus 1 (accession number: HQ613271), BmDENV=*Bombyx mori* densovirus 1 (AY033435),

23 528 CeDENV= *Casphalia extranea* densovirus (AF375296), DpDENV= *Dendrolimus punctatus* densovirus

24 529 (NC_006555). **Fig. S3** The maximum-likelihood tree for members of the densoviruses, including (a)

25 530 the genomic sequence with GTR+G+I model, (b) the amino acid sequence of the VP ORF with LG+G

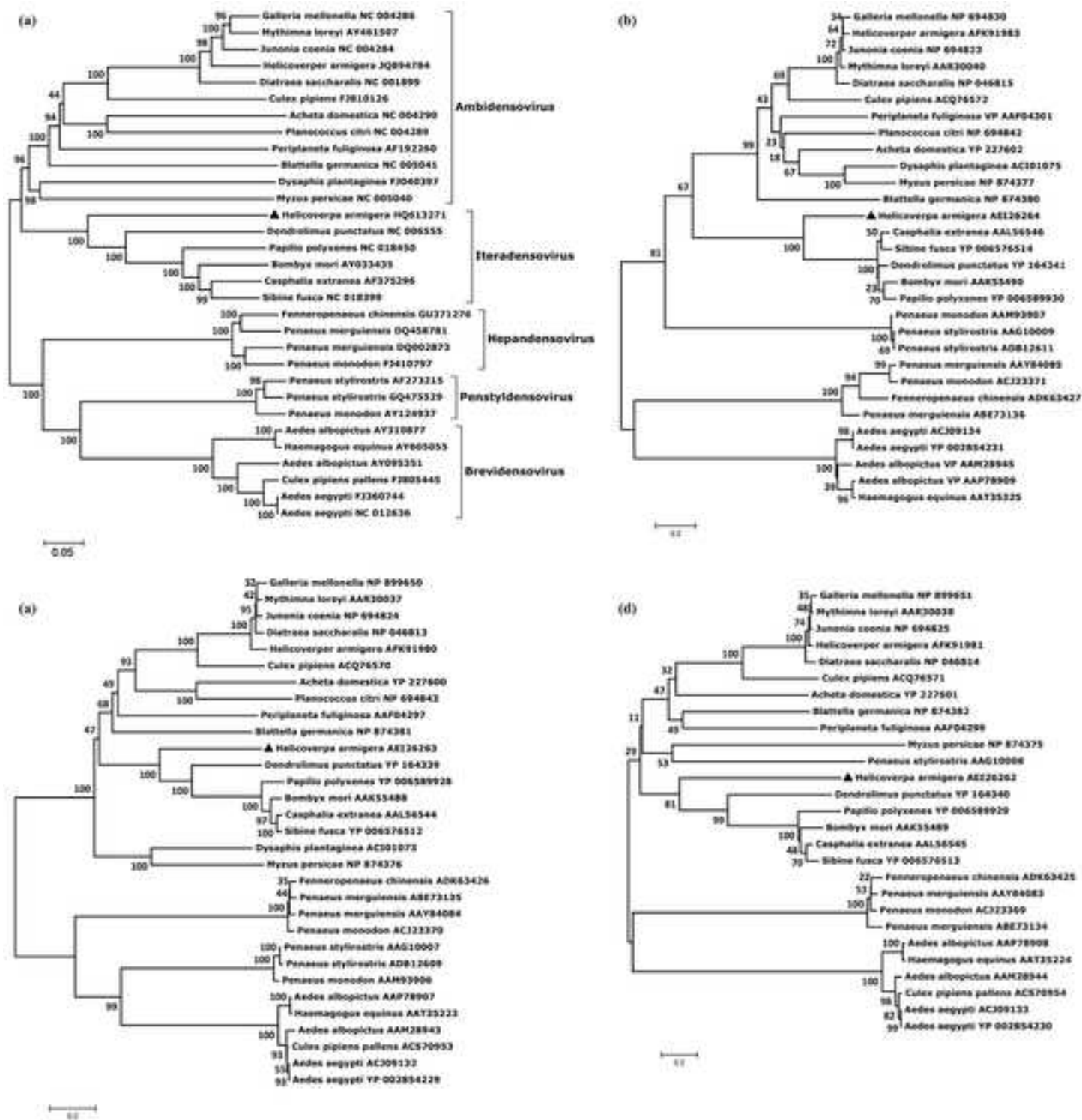
26 531 model, (c) the amino acid sequence of the NS1 ORF with LG+G+I model, and (d) the amino acid

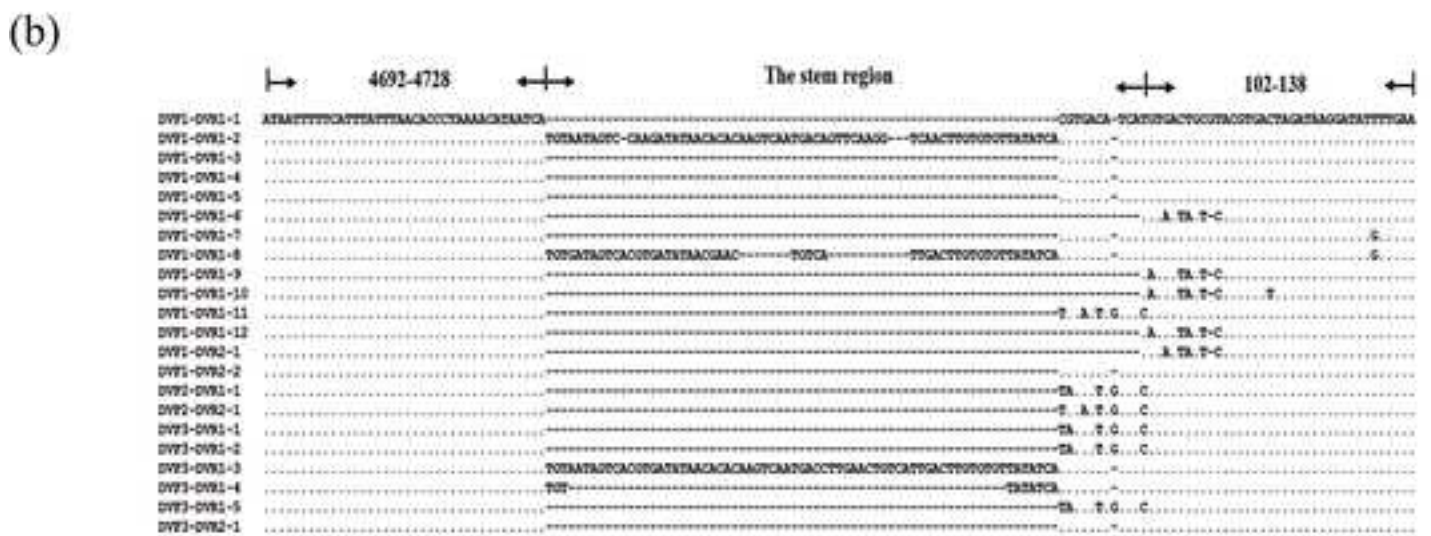
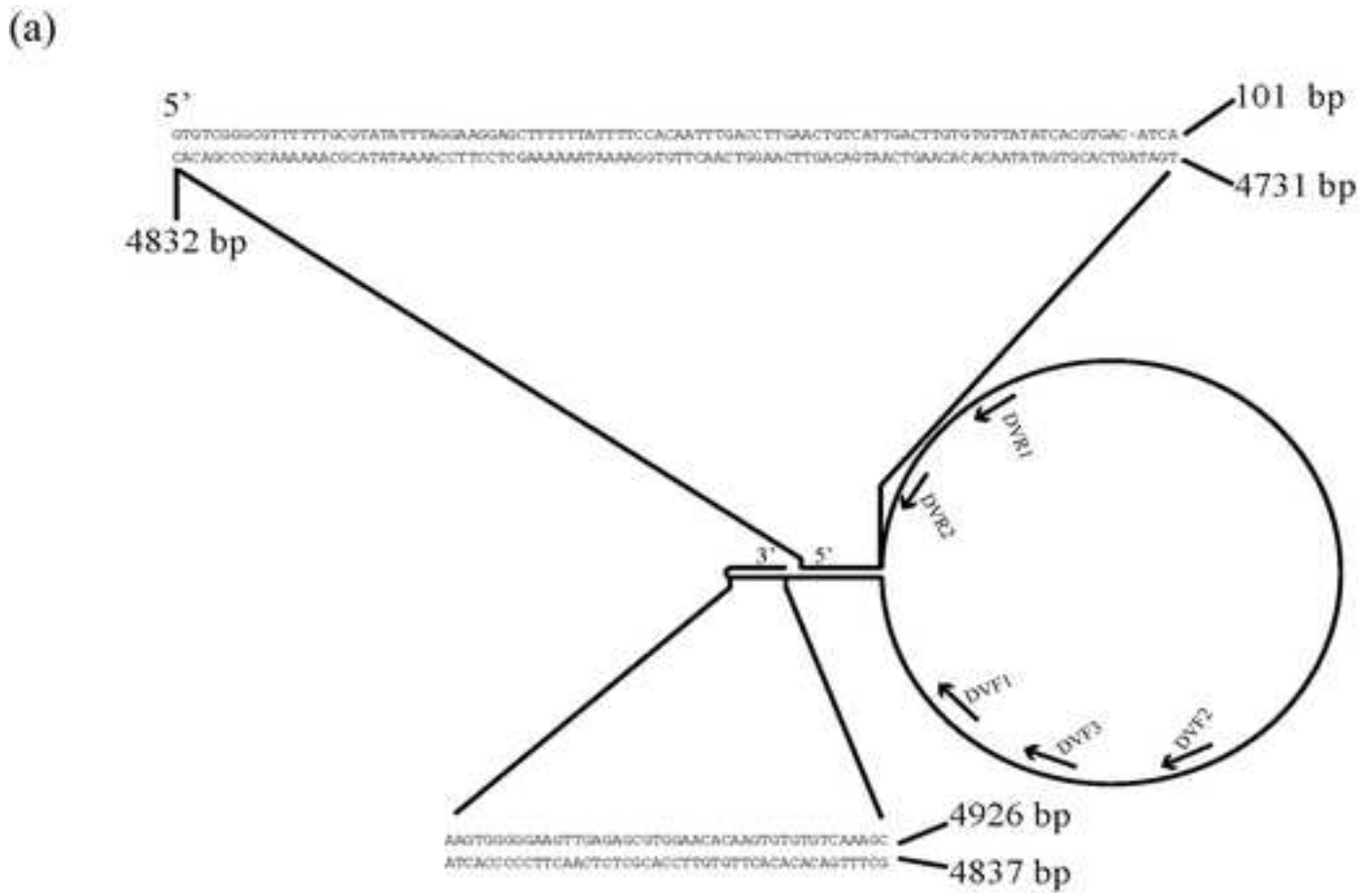
27 532 sequence of the NS2 ORF with JTT+G model. “▲” represents the sequence of HaDV2. Bootstrap

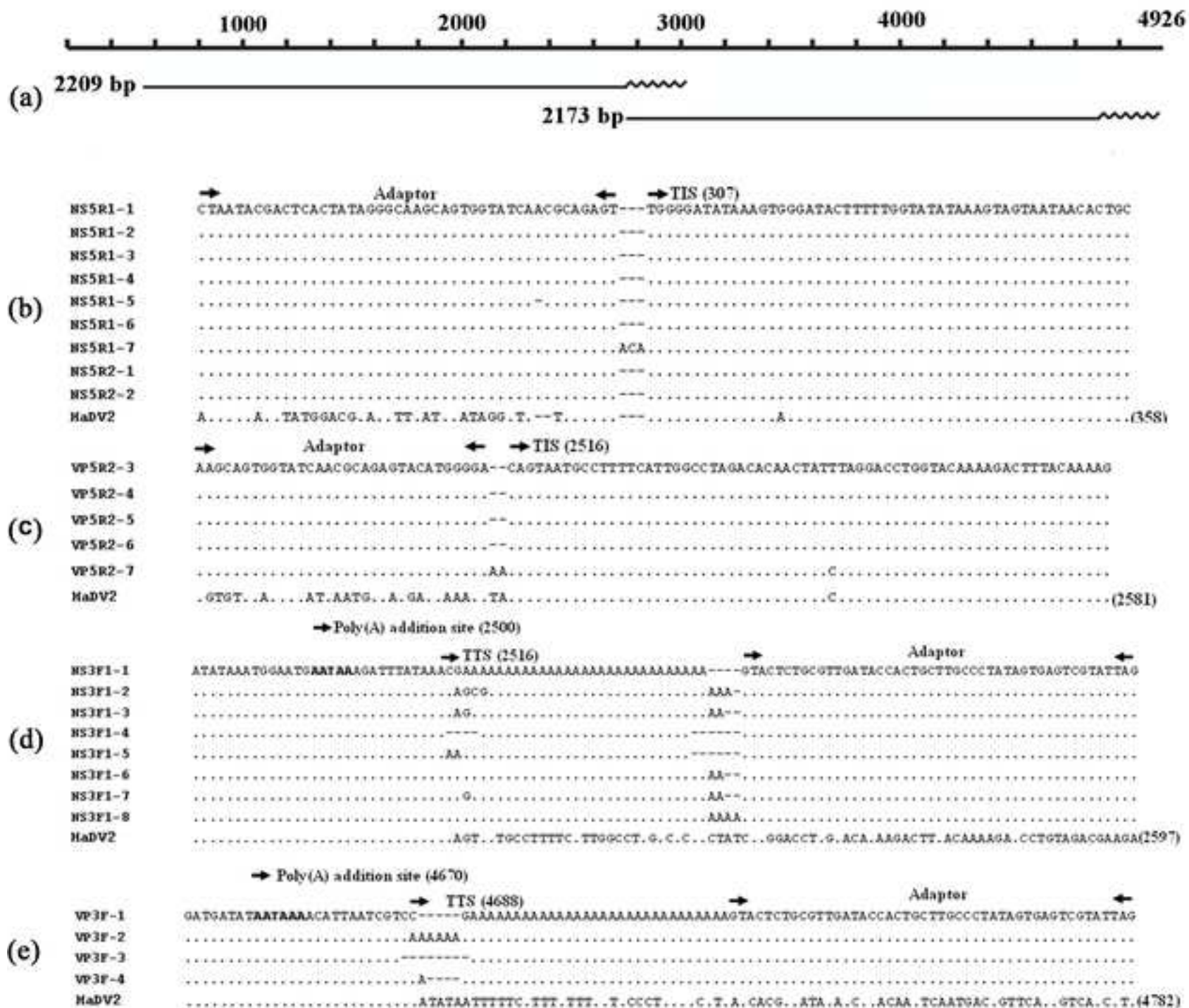
28 533 values (1000 pseudoreplicates) > 50% are indicated on the nodes. **Fig. S4** Northern blot analysis of

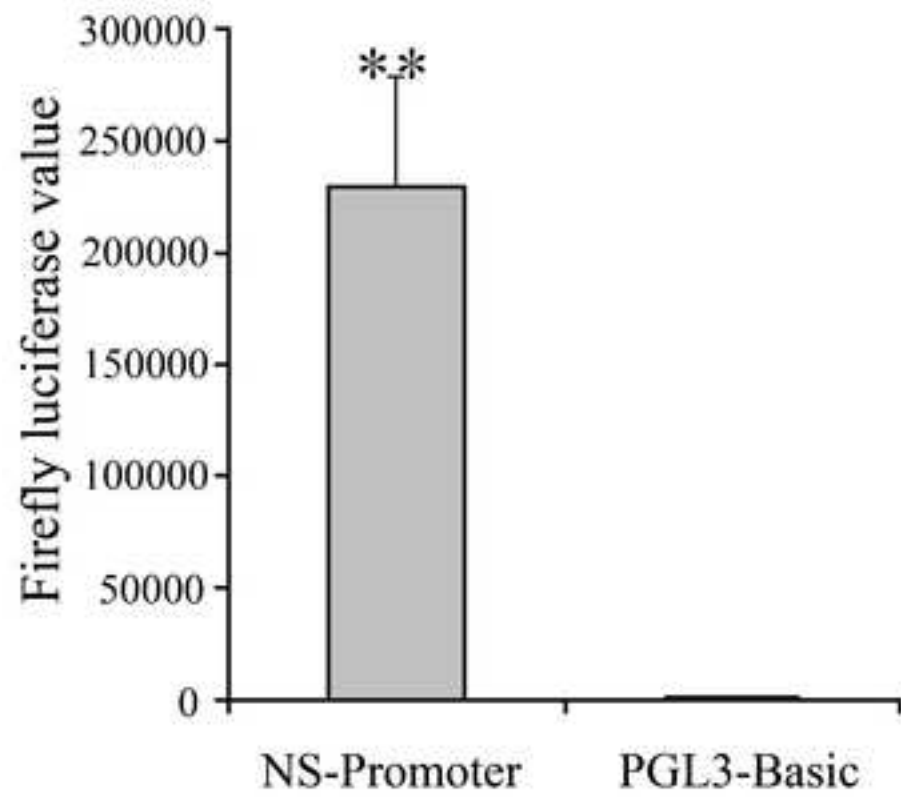
29 534 the HaDV2 transcripts showed two bands of 2.2 kb with the NS and the VP probe, respectively. **Fig.**

30 535 **S5** Dose-responses of anti-NS1, anti-NS2 and anti-VP antibodies using ELISA.

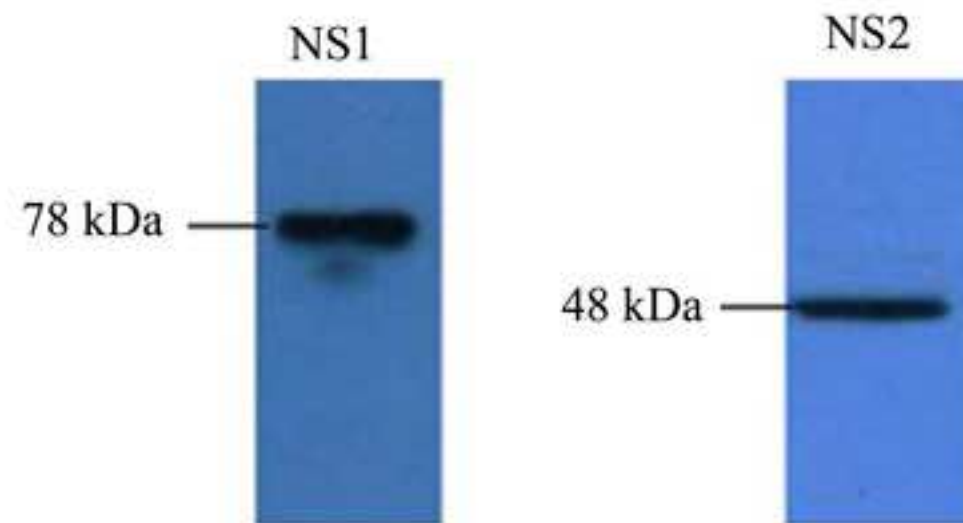








(a)



(b)

