The genome of a tortoise herpesvirus (testudinid herpesvirus 3) has a novel structure and contains a large region that is not required for replication *in vitro* or virulence *in vivo*

Running title: Testudinid herpesvirus 3 genome and pathogenesis

Frédéric Gandar,[1,2] Gavin S. Wilkie,[3] Derek Gatherer,[4] Karen Kerr,[3] Didier Marlier,[2] Marianne Diez,[5] Rachel E. Marschang,[6] Jan Mast,[7] Benjamin G. Dewals,[1] Andrew J. Davison,[3,#] Alain F.C. Vanderplasschen[1,#]

Immunology-Vaccinology, Department of Infectious and Parasitic Diseases, Fundamental and Applied Research for Animals & Health (FARAH), Faculty of Veterinary Medicine, University of Liège, Liège, Belgium[1]

Clinic for Birds, Rabbits and Rodents, Department of Clinical Sciences, FARAH, Faculty of Veterinary Medicine, University of Liège, Liège, Belgium[2]

MRC – University of Glasgow Centre for Virus Research, Glasgow, United Kingdom[3]

Division of Biomedical and Life Sciences, Faculty of Health and Medicine, Lancaster University, Lancaster, United Kingdom[4]

Nutrition of Companion Animals, Department of Animal Production, FARAH, Faculty of Veterinary Medicine, University of Liège, Liège, Belgium[5]

Laboratory for Clinical Diagnostics, Laboklin GmbH & Co. KG, Bad Kissingen, Germany[6]

Biocontrol Department, Research Unit Electron Microscopy, Veterinary and Agrochemical Research Centre, VAR-CODA-CERVA, Ukkel, Belgium[7]

# These authors contributed equally. Address correspondence to Alain F.C. Vanderplasschen (a.vdplasschen@ulg.ac.be) and Andrew J. Davison (andrew.davison@glasgow.ac.uk)

Abstract word count: 247 words

Text word count: 6568 words

ABSTRACT

Testudinid herpesvirus 3 (TeHV-3) is the causative agent of a lethal disease affecting several tortoise species. The threat that this virus poses to endangered animals is focusing efforts on characterizing its properties, in order to enable the development of prophylactic methods. We have sequenced the genomes of the two most studied TeHV-3 strains (1976 and 4295). TeHV-3 strain 1976 has a novel genome structure and is most closely related to a turtle herpesvirus, thus supporting its classification into genus *Scutavirus*, subfamily *Alphaherpesvirinae*, family *Herpesviridae*. The sequence of strain 1976 also revealed viral counterparts of cellular interleukin-10 and semaphorin, which have not been described previously in members of subfamily *Alphaherpesvirinae*. TeHV-3 strain 4295 is a mixture of three forms (m1, m2, and M), in which, in comparison to strain 1976, the genomes exhibit large, partially overlapping deletions of 12.5 to 22.4 kb. Viral subclones representing these forms were isolated by limiting dilution, and each replicated in cell culture comparably to strain 1976. With the goal of testing the potential of the three forms as attenuated vaccine candidates, strain 4295 was inoculated intranasally into Hermann's tortoises (*Testudo hermanni*). All inoculated subjects died, and PCR analyses demonstrated the ability of the m2 and M forms to spread and invade the brain. In contrast, the m1 form was detected in none of the organs tested, suggesting its potential as the basis of an attenuated vaccine candidate. Our findings represent a major step towards characterizing TeHV-3 and developing prophylactic methods against it.

45 **IMPORTANCE**

46

47 Testudinid herpesvirus 3 (TeHV-3) causes a lethal disease in tortoises, several species of
48 which are endangered. We have characterized the viral genome, and used this information to
49 take steps towards developing an attenuated vaccine. We have sequenced the genomes of two
50 strains (1976 and 4295), compared their growth *in vitro*, and investigated the pathogenesis of
51 strain 4295, which consists of three deletion mutants. The major findings are: (i) TeHV-3 has
52 a novel genome structure; (ii) its closest relative is a turtle herpesvirus; (iii) it contains
53 interleukin-10 and semaphorin genes, the first time these have been reported in an
54 alphaherpesvirus*;* (iv) a sizeable region of the genome is not required for viral replication *in*
55 *vitro* or virulence *in vivo*; and (v) one of the components of strain 4295, which has a deletion
56 of 22.4 kb, exhibits properties indicating that it may serve as the starting point for an
57 attenuated vaccine.

58

INTRODUCTION

The order *Herpesvirales* contains a large number of enveloped, double-stranded DNA viruses that share structural, genetic and biological properties, and is divided into three families infecting a wide range of hosts (1). One of these families, the *Herpesviridae*, contains viruses infecting mammals, birds, or reptiles, and is subdivided into three subfamilies, the *Alpha*-, *Beta*-, and *Gammaherpesvirinae* (2). Members of these subfamilies are referred to colloquially as alpha-, beta-, and gammaherpesviruses, respectively.

All herpesviruses of reptiles identified to date group among the alphaherpesviruses, in lineages distinct from herpesviruses of mammals or birds (3). Many of the hosts of these viruses belong to the order Testudines (also called Chelonii), and include pond turtles, marine turtles, and terrestrial tortoises. Among members of the order Testudines, herpesvirus infections have been described chiefly in the latter two groups. Among marine turtles, the genome of chelonid herpesvirus 5 (ChHV-5), which is thought to be the causative agent of fibropapillomatosis, has been cloned as a bacterial artificial chromosome from infected tissue and sequenced. This virus has been classified into the genus *Scutavirus* (4). The phylogenetic relationship between herpesviruses infecting marine turtles and those infecting tortoises is unclear (3).

Tortoises exist as at least 40 species belonging to the family Testudinidae. Herpesviruses have been isolated from healthy or sick individuals belonging to several of these species. Based on partial sequencing of the viral DNA polymerase gene, four genotypes have been identified, leading to the nomenclature testudinid herpesvirus 1 to 4 (TeHV-1 to TeHV-4) (3). Among these genotypes, TeHV-3 appears to be the most pathogenic, and has been shown to affect several tortoise species, with those from the genus *Testudo* (e.g., *Testudo hermanni*) being the most sensitive (5, 6). Young tortoises are more susceptible to TeHV-3 disease than adults, and can suffer from mortality rates of up to 100%. These pathological and

84 epidemiological features, and the fact that many of the susceptible host species are
85 endangered, contribute to ecological concerns over this virus. Clinical signs depend on several
86 factors, including host species, age, the season at which infection occurs, and the viral strain
87 involved (7-12). The main clinical signs are nasal discharge, rhinitis, conjunctivitis associated
88 with blepharospasm, and diphtheritic plaques in the oral cavity and esophagus (13). Weight
89 loss, cachexia, central nervous symptoms (such as circling and head tilt), and death are
90 observed in advanced stages of the disease. The virus has been isolated from several tissues,
91 suggesting a broad tropism (8, 11, 14, 15). With the goal of controlling the threat that TeHV-3
92 poses to tortoise populations, various inactivated vaccine candidates have been tested, but
93 none has proved efficacious (16, 17). Obvious alternatives to inactivated vaccines are
94 attenuated vaccines and subunit vaccines targeting key viral proteins. The knowledge required
95 for the development of such vaccines would include the genome sequence.

96 Here, we sequenced the genomes of representative TeHV-3 strains 1976 (18) and
97 4295 (9). Strain 1976 has a genome structure not reported previously among herpesviruses,
98 and is most closely related phylogenetically to turtle herpesviruses. This strain also contains
99 genes with cellular homologues that have not been described previously in
100 alphaherpesviruses. Strain 4295 consists of a mixture of three forms, of which the genomes
101 exhibit large, partially overlapping deletions in comparison with strain 1976. The effects of
102 these deletions on viral growth *in vitro* and virulence *in vivo* were investigated.

MATERIALS AND METHODS

Cells and viruses

Terrapene heart cells (TH-1, subline B1) (19) were cultured in medium (Dulbecco's modified Eagle's medium (DMEM; Sigma-Aldrich) containing 4.5 g/L glucose, 5 % fetal calf serum, and 1 % non-essential amino acids (Invitrogen)). Cells were cultured at 25°C in a humid atmosphere in the presence of 5 % CO_2. Two previously described TeHV-3 strains were used. Strain 1976 (passage 6) originated from the intestine of a Horsfield's tortoise (*T. horsfieldii*) that died from TeHV-3 infection (18). Strain 4295 (passage 14) was isolated from a pharyngeal swab performed on a clinically healthy Hermann's tortoise (*T. hermanni*) during an outbreak of herpesviral disease (9). The absence of extraneous contaminating agents in the TH-1 cells and TeHV-3 strains was confirmed by electron microscopic examination of mock-infected and infected TH-1 cells (Fig. 1). Clones of strain 4295 were produced by limiting dilution. Ten-fold serial dilutions of infected culture supernatant were inoculated onto TH-1 cells grown in 96-well plates. Clones were amplified from dilutions for which less than 10 % of the wells showed signs of infection. The purity and genotype of the clones were determined by PCR (see below; Fig. 7A and B).

Virion DNA production

Confluent TH-1 cells were infected with TeHV-3 at a multiplicity of infection (MOI) of 0.2 PFU/cell. To reduce contamination by cellular DNA, cell supernatant was harvested at the early stages of viral release, when approximately 10 % of cells were exhibiting cytopathic effect. Virions were semi-purified as described previously (20). Briefly, after removal of the cell debris by centrifugation (1000 *g*, 10 min, 4°C), viral particles were pelleted by ultracentrifugation through a 30 % sucrose cushion (100,000 *g*, 2 h, 4°C). DNA was purified from virions as described previously (21).

128 **DNA sequencing**

129 Virion DNA (1 μg) was sheared by sonication to an average size of 470 bp, and
130 sequencing libraries were prepared by using a KAPA library preparation kit (KAPA
131 Biosystems). The fragments were A-tailed and ligated to the NEBnext Illumina adaptor (New
132 England Biolabs), and NEBnext indexing primers were added by carrying out four cycles of
133 PCR in an ABI 7500 realtime cycler, using a KAPA HiFi Realtime library amplification kit.
134 The libraries were sequenced by using a MiSeq (Illumina) operating v2 chemistry, generating
135 data sets of 250 nucleotide (nt) paired-end reads. The reads were filtered for quality, adapter
136 sequences were removed by using Trim Galore v. 0.2.2
137 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore), and the reads were
138 assembled by using Velvet v. 1.2.07 (22) and AbySS v. 1.3.5 (23). Larger contigs were
139 generated from the assemblies by using Phrap v. 1.080812 (24, 25) and then IMAGE v. 2.31
140 (26). The data sets were assembled against the contigs by using BWA v. 0.6.2-r126 (27), and
141 the alignments were inspected by using Tablet 1.13.12.17 (28). Problematic regions, including
142 those representing deletions or duplications, and those containing relatively short reiterated
143 sequences (including mononucleotide tracts), were resolved by using custom Perl scripts to
144 count and extract individual reads from the data sets for further analysis (29). The sizes of the
145 more substantial reiterated sequences were estimated by PCR (Table 1), and, in most
146 instances, the products were sequenced by using Sanger technology.

147 Potential genome termini were identified in the BWA alignments from sets of reads
148 sharing a common end. They were confirmed for strains 1976 and 4295 (prior to subcloning)
149 by using published methodology (30), which involves ligating partially double-stranded DNA
150 adapters to blunt-ended or untreated viral DNA, followed by PCR using a combination of a
151 virus-specific primer and an adapter-specific primer (the former are listed in Table 1). The
152 inserts from 12 plasmids generated from each purified PCR product were sequenced by using

153 Sanger technology, and the genome termini were defined as being located at the positions
154 represented by the majority of clones. Analysis of both blunt-ended and untreated viral DNA
155 allowed unpaired nucleotides at the 3'-ends of the genome to be identified. The genome
156 termini of the three subclones (m1, m2, and M) of strain 4295 were not determined in this
157 way, but rather inferred from the data from strain 4295. The final genome sequences of strain
158 1976 and the three subclones (m1, m2, and M) of strain 4295 were constructed on the basis of
159 the locations of the termini, and the integrity of each was verified by aligning it against the
160 relevant data set using BWA, visualizing the alignment by using Tablet. The sequences were
161 deposited in NCBI GenBank (see below).

162 **Southern blot analysis**

163 Southern blot analysis of virion DNA, digested with EcoRI or KpnI restriction
164 endonucleases, was performed as described previously (31, 32). Probes were produced by
165 PCR amplification of strain 1976 DNA using specific primers (Table 1).

166 **Viral growth curves**

167 Triplicate cultures of TH-1 cells in 24-well plates were inoculated with TeHV-3 at an
168 MOI of 0.2 PFU/cell. After an incubation period of 4 h, the cells were washed with PBS and
169 overlaid with DMEM containing 4.5 g/liter glucose, 5 % FCS and 1 % non-essential amino
170 acids (Invitrogen). The supernatants of infected cultures were harvested at successive
171 intervals after infection and stored at -80°C. The amount of infectious virus was determined
172 by plaque assay on TH-1 cells as described previously (32, 33). The data, expressed as mean
173 titer and standard deviation (SD) of triplicate assays, were analyzed for significant differences
174 ($p<0.05$) using one-way ANOVA.

175

176 **Transmission electron microscopy**

177 TH-1 cells were infected with TeHV-3 at a MOI of 0.2 PFU/cell, and, at 6 days
178 post-infection, processed for electron microscopic examination as described elsewhere (32).

179

180 **Tortoises**

181 Five-year old Hermann's tortoises (*T. hermanni*) originating from a small colony bred
182 in captivity were kept individually in terrariums (width x depth x height: 0.9 x 0.45 x 0.6 m).
183 The environmental parameters were as follows. Relative humidity was maintained at 60-70
184 %. Lighting was controlled automatically on standard 12 h light and 12 h dark circadian
185 cycles, with a UVb light switched on during the 12 h light period. The temperature of an
186 infrared basking spot was regulated at 29°C during the light period and 24°C during the dark
187 period, respectively. A temperature gradient of approximately 6°C was present in the
188 terrariums, with the basking spot being the warmest place. Fresh water and vegetables were
189 provided daily. Clinical examinations of tortoises immediately prior to the experiments
190 revealed that they were healthy. Experiments were preceded by an acclimatization period of 2
191 weeks.

192 **Inoculation of tortoises with TeHV-3**

193 Tortoises were sedated with alfaxolone (Dechra Veterinary Products) injected intravenously
194 at a dose of 7 mg/kg. TeHV-3 was then inoculated by intranasal instillation of $1x10^5$ PFU
195 distributed equally between both nostrils (total volume of 50 µl). The animals were examined
196 twice daily until the end of the experiment. Animals expressing significant apathy,
197 neurological signs, or respiratory distress were euthanized in accord with the end point
198 defined by the local bioethics committee.

199

200 **Ethics statement**

201 The experiments, maintenance, and care of tortoises complied with the guidelines of
202 the European Convention for the Protection of Vertebrate Animals used for Experimental and
203 other Scientific Purposes (CETS no. 123). The animal study was approved by the local ethics
204 committee of the University of Liège, Belgium (laboratory accreditation no. LA 1610010,
205 protocol no. 1217). All efforts were made to minimize suffering and to respect the 3Rs rule.

206 **Quantification of viral gene copies in organs by qPCR**

207 DNA was isolated from 25 mg of organs stored at -80°C in RNAlater (Invitrogen) by
208 using a DNA mini kit (Qiagen). The viral genome was quantified by amplifying fragments of
209 the TeHV-3 UL13 and *T. hermanni* β-actin genes, using real-time SYBR green-based PCR.
210 The primers used are listed in Table 1. The qPCR reactions were performed using a CFX96
211 Touch real-time PCR detection system and iTaq universal probe supermix as detection
212 chemistry (Bio-Rad Laboratories). Master-mix for qPCR consisted of 1 x iTaq universal
213 probe supermix, 200 nM each primer, and 200 ng sample DNA in a final volume of 15 µl.
214 The UL13 amplification program included an initial denaturation step at 95°C for 3 min,
215 followed by 50 cycles with a denaturation step at 95°C for 30 s, an annealing step at 58.5°C
216 for 30 s, and an elongation step at 72°C for 30 s. The β-actin amplification program included
217 an initial denaturation step at 95°C for 3 min, followed by 50 cycles with a denaturation step
218 at 95°C for 30 s, an annealing step at 60°C for 30 s, and elongation step at 72°C for 30 s. At
219 the end of these amplification programs, the dissociation stage was performed (95°C for 10 s),
220 and the melting curve was determined by increasing the temperature from 60 to 95°C at the
221 rate of 0.1°C/s. All reactions were carried out in triplicate.

222 Data for validation of qPCR (efficiency (E), coefficient of determination (R^2), and
223 slope) were analyzed by using Bio-Rad CFX Manager 3.1 software (Bio-Rad Laboratories)
224 with the auto method (UL13: E = 94.7 %, R^2 = 0.994, slope = -3.456; β-actin : E = 95.2 %, R^2

225 = 0.982, slope = -3.443). The mean values of the number of TeHV-3 genome copies in
226 various organs were compared by using the Kruskal-Wallis test. A value of $p<0.05$ was
227 considered significant, and a value of $p<0.01$ was considered highly significant.

228 **Histological analyses**

229 Lung, spleen, brain (telencephalon), kidney, and liver from a mock-infected tortoise
230 and from infected tortoises were dissected immediately after euthanasia, fixed in 10 %
231 buffered formalin, and embedded in paraffin blocks. Sections of 5 μm were stained with
232 hematoxylin and eosin prior to microscopic analysis (34). For each sample, ten randomly
233 selected fields were examined in a blind test.

234 **Phylogenetic analyses**

235 Predicted amino acid sequences were obtained in this study or from GenBank. The
236 raw phylogenetic data derived from them are available at
237 http://dx.doi.org/10.17635/lancaster/researchdata/11. The sequences of herpesvirus DNA
238 polymerases were aligned by using Muscle (35) in MEGA (36). A Bayesian tree was then
239 constructed in BEAST (37), using the LG+Γ substitution model (38), a lognormal relaxed
240 clock model (39), and a Yule speciation tree model (40). The tree was run to convergence, at
241 which point all posterior probabilities on the nodes were 1. The amino acid sequence of
242 TeHV-3 gene TE7, which encodes a semaphorin (SEMA)-related protein, was aligned with
243 the sequences of SEMA-7s by using Muscle (35), before constructing a maximum likelihood
244 tree in MEGA (36), using the LG+Γ substitution model (2). The amino acid sequence of
245 TeHV-3 gene TE8, which encodes an interleukin-10 (IL-10)-related protein, was analyzed by
246 using Treeadder (41) with a previously published tree (42), in order to create a set of trees
247 with the TeHV-3 TE8 sequence added at every possible position. The likelihood of each tree

248 was assessed by using Tree-puzzle (43), and bootstrap significance was assessed by using
249 Consel (44).

250 **Structural analyses**

251 Solved structures from PDB were selected for the SEMA and IL-10 protein families.
252 Homology models for TeHV-3 SEMA and IL-10 were constructed in relation to these solved
253 structures by using MOE 2014.09 (Chemical Computing Group, Montreal). Briefly, an initial
254 partial geometry was copied from the template chains in the solved structures, using all
255 co-ordinates in which residue identity was conserved. Otherwise, only backbone coordinates
256 were used. Based on this initial geometry, Boltzmann-weighted randomized modelling (45)
257 was employed, with segment searching for regions that could not be mapped onto the initial
258 geometry (46). A total of 100 models were constructed. On completion of segment addition,
259 each model was energetically minimized in the AMBER99 force field (47). The highest
260 scoring intermediate model was then determined by the generalized Born/volume integral
261 (GB/VI) methodology (48). Stereochemical quality of homology models was assessed using
262 Ramachandran plots (49).

263

264 **Nucleotide sequences**

265 The genome sequences of TeHV-3 strain 1976 and the three subclones (m1, m2, and
266 M) of strain 4295 were deposited in NCBI GenBank under accession numbers KM924292,
267 KR363629, KR363628, and KM924293, respectively.

268 **RESULTS AND DISCUSSION**

269 The initial goal of this study was to determine the sequence of the TeHV-3 genome.
270 To achieve this, the two most studied isolates (strains 1976 and 4295) were sequenced.
271 Analysis of high-throughput sequence data for strain 1976 indicated the presence of a single
272 population, whereas that for strain 4295 implied the presence of a mixture of three closely
273 related populations, each being most simply interpreted as a deletion mutant derived from an
274 ancestral genome similar to that of strain 1976.

275 **Genome sequence of strain 1976**

276 The filtered data set for strain 1976 consisted of 3,787,248 reads, of which 1,488,829
277 (39 %) aligned with the finished sequence at an average coverage of 2305 reads/nt. The
278 genome is 160,358 bp in size, and consists of a central, long unique region (U_L; 107,928 bp),
279 extended at its right end by a short unique region (U_S; 20,375 bp) flanked by an inverted
280 repeat (IR_S and TR_S; 8536 bp), and at its left end by a third unique region (U_T; 12,595 bp) also
281 flanked by an inverted repeat (TR_T and IR_T; 1194 bp), yielding the overall configuration
282 TR_T-U_T-IR_T-U_L-IR_S-U_S-TR_S (Fig. 2). Each 3'-end of the genome consists of a single,
283 unpaired nucleotide complementary to the nucleotide at the other 3'-end. All alphaherpesvirus
284 genome structures share the U_L-IR_S-U_S-TR_S component, in some cases with U_L flanked by an
285 inverted repeat to produce TR_L-U_L-IR_L-IR_S-U_S-TR_S (2). However, the TR_T-U_T-IR_T
286 component is novel, yielding a genome structure that has not been reported previously in the
287 family *Herpesviridae*. The alphaherpesvirus genomes most reminiscent of TeHV-3 in this
288 regard are equid herpesviruses 1 and 4 (genus *Varicellovirus*), which contain a small inverted
289 repeat (87 and 86 bp, respectively) at the left end of U_L, separated from its counterpart by a
290 short sequence (944 and 667 bp, respectively) (50, 51). However, the sizes of TR_T/IR_T and U_T
291 in TeHV-3 are much larger than in these viruses, and each contains protein-coding sequences

292 (see below). The average nucleotide composition of the strain 1976 genome is 46 % G+C,
293 with TR_T/IR_T, U_T, U_L, TR_S/IR_S, and U_S being 65, 45, 44, 55, and 45 %, respectively. Further
294 details of the strain 1976 sequence, including an annotation of predicted protein-coding
295 content (see below), are available in the GenBank accession. Four partial sequences from
296 strain 1976 have been published previously, and all are identical to the corresponding sections
297 of the genome sequence. These include 8667 bp extending from within UL40 to within UL36
298 (GenBank accession AY338245), and much shorter sequences from UL39 (DQ343900), UL5
299 (DQ343892), and UL30 (DQ343881) (6).

300 In alphaherpesvirus genomes, unique regions flanked by inverted repeats are typically
301 present in either orientation in virion DNA, giving rise to the presence of more than one
302 genome isomer (2). Consequently, the orientations of U_T and U_S in the strain 1976 genome
303 were investigated by Southern blot analysis (Fig. 3). Fig. 3A illustrates the four possible
304 arrangements of U_T-U_L and U_L-U_S, and the restriction endonuclease fragments potentially
305 generated by digestion with EcoRI or KpnI. Hybridizations performed with appropriate
306 probes showed that U_T and U_S are present in either orientation, whereas U_L, which is not
307 flanked by an inverted repeat, is present in a single orientation (Fig. 3B). These results
308 indicate that TeHV-3 virion populations contain a mixture of four genome isomers differing
309 in the relative orientations of U_T and U_S.

310 **Genome sequence of strain 4295**

311 The filtered data set for strain 4295 consisted of 9,772,400 reads, of which 3,106,488
312 (32%) aligned with the finished sequence at an average coverage of 5195 reads/nt. Analysis of
313 the data indicated that the DNA contained a mixture of three related genome populations,
314 each most simply interpreted as being a deleted form of a genome similar to that of strain
315 1976. The extents of the deletions in relation to strain 1976 are marked by horizontal green
316 bars in Fig. 2. The three genome forms were named m1 and m2 (the minor forms, estimated

317 from counting reads representing the novel junction, representing approximately 23 and 11 %
318 of the population, respectively), and M (the major form representing approximately 66 % of
319 the population). This interpretation was confirmed by subcloning the m1, m2, and M forms by
320 limiting dilution and sequencing their genomes at an average coverage of 25, 46, and 22
321 reads/nt, respectively. Compared with the strain 1976 sequence, the m1, m2, and M forms
322 exhibit large, partially overlapping deletions of 22,424, 12,485, and 18,315 bp, respectively,
323 extending from U_T (in the orientation shown in Fig. 2), across IR_T, to U_L. No reads matching
324 the region in the strain 1976 genome that corresponds to the region absent from all three
325 forms (nt 4780-14989 in strain 1976) were detected in strain 4295 prior to subcloning,
326 indicating that none of putative parental genome remained. In addition, the data supported the
327 presence of two sizeable duplications in strain 4295 prior to subcloning and in the subclones,
328 one located near the right end of U_L (present in the m1 and m2 forms) and the other in IR_S and
329 TR_S (present in all three forms). The extents of these duplications relative to strain 1976 are
330 marked by horizontal orange bars in Fig. 2. Further details of the sequences of the strain 4295
331 subclones, including annotations of coding content, are available in the GenBank accessions.
332 No reads diagnostic of the deletions and duplications in strain 4295 (i.e., reads representing
333 the novel junctions) were detected in strain 1976. The genomes of strain 1976 and 4295 (the
334 M form) are closely related, differing by 193 substitutions (including IR_S but not TR_S, and not
335 counting insertions or deletions).

336 **General features of the TeHV-3 genome**

337 Standard bioinformatic approaches were taken to predict the locations of open reading
338 frames (ORFs) in the strain 1976 genome that encode functional proteins (for example, see
339 (30)). In general, ORFs were included that potentially encode proteins of 100 or more amino
340 acid residues, that do not extensively overlap ORFs predicted to be functional (i.e., ORFs
341 encoding proteins that are similar to proteins of known function or have features suggesting

342 function, such as hydrophobic domains), and that are located appropriately in relation to
343 potential mRNA polyadenylation signals. Particular attention was paid to ORFs that have
344 counterparts in other herpesviruses. Four ORFs potentially encoding proteins of fewer than
345 100 residues were also added because they are related to recognized herpesvirus proteins
346 (UL11) or have other distinguishable features (TE35, TE36, and TE39). Splicing was
347 predicted in five ORFs (TE12, TE13, ORF13, UL15, and TE25). The first ATG in each ORF
348 was assigned as the initiation codon, except in cases in which use of a subsequent ATG was
349 supported by alignments with related proteins or in which it provided a putative signal
350 peptide.

351 The analysis indicated that the strain 1976 genome encodes a total of 107 predicted
352 functional ORFs (Fig. 2 and Table 2), three of which are duplicated in the inverted repeats
353 (TE1 in TR_T/IR_T, and TE25 and RS1 in TR_S/IR_S). These 107 ORFs are conserved in strain
354 4295, except for those affected by the deletions, and TE35, which is frameshifted (and
355 therefore considered marginal in strain 1976). A total of 19 ORFs belong to six families of
356 paralogous genes: the TE3 family with nine members, and the TE15, TE22, TE27, UL55, and
357 US3 families, each with two members. The genome also contains three copies of a potential
358 origin of DNA replication (ori), identified as an A+T-rich region capable of forming a hairpin
359 structure and containing characteristic sequence motifs involved in binding the UL9 DNA
360 replication origin-binding helicase (52-54) (Fig. 2). Approximately 37 % of strain 1976
361 genomes appear to contain a 154 bp deletion that results in the absence of one of the copies of
362 ori from the region between UL1 and TE25. This deletion also appears to be present in
363 approximately 50 % of strain 4295 genomes prior to subcloning, but their distribution in this
364 strain and in subclones m1 and m2 was not determined with certainty because of the
365 ambiguity caused by duplications in this region.

366

367 **Relationships between TeHV-3 and other herpesviruses**

368 The strain 1976 ORFs are shown in three categories in Fig. 2, according to their
369 conservation in other herpesviruses. The first category (shaded red) consists of the 44 genes
370 that are thought to have been inherited from the ancestor of the alpha-, beta-, and
371 gammaherpesviruses. The second category (shaded blue) comprises the 21 additional ORFs
372 (not counting duplicates) that have orthologs in alphaherpesviruses of mammals or birds. The
373 third category (shaded orange) contains the remaining 42 ORFs (TE1-TE42), all of which
374 lack counterparts in other herpesviruses, except for TE31, which has an orthologue in the
375 turtle alphaherpesvirus ChHV-5 (4). This category of TeHV-3-specific gene includes the TE3,
376 TE15, TE22, and TE27 families, as well as genes TE7, TE8, and TE11, which are similar to
377 SEMAs, IL-10s, and C-type lectins, respectively.

378 The full list of TeHV-3 strain 1976 ORFs is provided in Table 2, including those
379 having counterparts in ChHV-5 (marked with asterisks in the first column). The order of
380 conserved ORFs in the TeHV-3 genome is the same as that in ChHV-5, except for the absence
381 or disruption of 14 ORFs in the latter (UL44-UL51, UL54, UL39, UL40, and UL13) and the
382 inversion of UL55 (of which TeHV-3 has two counterparts, UL55A and UL55B).
383 Comparison with other alphaherpesvirus lineages (in the mammalian alphaherpesvirus genera
384 *Simplexvirus* and *Varicellovirus*, and the avian alphaherpesvirus genera *Mardivirus* and
385 *Iltovirus*) (55) indicates that the ancestral alphaherpesvirus contained conserved ORFs
386 arranged in the same order as in TeHV-3 from UL43 rightwards. The ancestral state from
387 UL54 leftwards is more difficult to discern, as this region appears to have undergone
388 rearrangements in various lineages.

389 Phylogenetic analysis of short regions of the genome has indicated previously that
390 TeHV-3 is an alphaherpesvirus (6). This conclusion is supported from the complete genome
391 sequence by the presence in TeHV-3 of 21 ORFs that have orthologs in alphaherpesviruses

392 but not in beta- or gammaherpesviruses (Fig. 2), and by phylogenetic analysis of the DNA
393 polymerase (Fig. 4A). This analysis also indicates that the closest known relative of TeHV-3
394 is ChHV-5, the current sole member of the genus *Scutavirus* (1). However, despite their
395 relationship, the two viruses do not share the same genome structure, ChHV-5 apparently
396 adopting the simpler U_L-IR_S-U_S-TR_S arrangement (4). Given the relatively large phylogenetic
397 distance between TeHV-3 and ChHV-5, it is a matter of judgment whether TeHV-3 should be
398 classified as a new species into the same genus as ChHV-5 or into a new genus. Since these
399 are the only relevant viruses that have been examined in sufficient detail, and both infect
400 Testudines, we recommend the former as the safer option. Regardless of the eventual
401 taxonomical outcome, the results establish a robust phylogenetic relationship between tortoise
402 and turtle herpesviruses.

403 **Phylogenic analysis and homology modelling of the TeHV-3 TE7 and TE8 proteins**

404 The finding that the TeHV-3 TE7 and TE8 genes encode SEMA and IL-10 homologs,
405 respectively, is particularly interesting for the following reasons. First, SEMA homologs have
406 been reported in the family *Herpesviridae* only in the *Macavirus* genus of subfamily
407 *Gammaherpesvirinae* (56). Second, although numerous IL-10 homologs have been described
408 in the subfamilies *Betaherpesvirinae* and *Gammaherpesvirinae* (42), they have not been
409 reported in the subfamily *Alphaherpesvirinae*. We performed phylogenetic analyses to
410 determine whether the origins of the TeHV-3 SEMA and IL-10 homologs could be traced. We
411 also performed homology modelling analyses, in order to assess whether these viral genes
412 encode functional homologs of the cellular genes.

413 A maximum likelihood phylogenetic tree placed the TeHV-3 SEMA (the TE7 protein)
414 as the nearest neighbor to a cluster of poxvirus SEMAs, with low bootstrap confidence (Fig.
415 4B). No viral SEMA is close enough to any cellular counterpart to justify the deduction of an
416 evolutionary scenario involving a recent gene transfer from host to virus. In contrast, as

417 previously described (42), some viral IL-10s (in HHV-4, EHV-2, and ORFV) cluster
418 phylogenetically with cellular IL-10s, implying relatively recent gene transfers and a degree
419 of functional conservation (Fig. 4C). However, other viral IL-10s (including that of TeHV-3)
420 are more divergent from host IL-10s, implying earlier gene transfer events with subsequent
421 genetic drift and possible functional divergence. As a consequence, no inference concerning
422 the evolutionary history of the TeHV-3 IL-10 (the TE8 protein) can be made, other than the
423 speculation that it originated from a host in a relatively ancient transfer event, as appears to be
424 the case for many other viral IL-10s (42).

425 We investigated the possibility of functional divergence by using homology modelling
426 of the TE7 and TE8 proteins in comparison with cellular SEMAs and IL-10s, respectively
427 (Fig. 5 and 6). The templates were selected for superposition on the basis of the lowest root
428 mean square deviation (RMSD) values exhibited by the top BLAST hits. This process
429 indicated that mouse SEMA-3A and human IL-10 were the best templates (Table A1). It
430 should not be inferred from these choices that we believe the TE7 protein to be an ortholog of
431 mouse SEMA-3A or the TE8 protein an ortholog of human IL-10. Indeed, the top BLAST hit
432 of the TE7 protein among the cellular SEMAs was SEMA-7A; hence the use of this group of
433 proteins in Fig. 4B. On the basis of the templates selected, modelling clearly demonstrates
434 that the TE7 protein (Fig. 5B) and the TE8 protein (Fig. 6B) are both capable of assuming
435 backbone structures highly similar to their cellular counterparts, despite having a high degree
436 of sequence divergence. Both TE7 and TE8 homology models contain some residues that are
437 sub-optimal in terms of their stereochemistry (Fig. 5C and 6C). However, these are mostly at
438 turns in the backbone or loop regions (Fig. 5B and 6B). Major secondary structural regions,
439 by contrast, are well formed in terms of their stereochemistry. We conclude that functional
440 divergence is unlikely to have occurred to an extent such that these two viral proteins no
441 longer operate in ways analogous to their cellular equivalents, and therefore that the TE7 and

442 TE8 proteins are likely to have maintained SEMA and IL-10 functions. This conclusion is
443 further supported by key residues of TE7 and TE8 proteins. The three-dimensional structure
444 of SEMA-7A bound to its receptor PLEXIN-C1 has been resolved, and key residues in the
445 SEMA sequence for the interaction with its receptor have been identified (57). Interestingly,
446 most of these residues present in the blade 3, 4c-4d loop, and extrusion helix 2 regions were
447 conserved in TE7. Among the residues conserved in TeHV-3, only one of them (Ile238) had a
448 sub-optimal stereochemistry (Fig. 5B). Together, these data suggested that the TeHV-3 TE7
449 protein might also signal through PLEXIN-C1 and function similarly to other viral SEMAs
450 (56). As for the TE8 protein, its sequence exhibits most of the residues found in the two
451 family signature motifs characteristic of all cellular Il-10s: L-[FILMV]-X3-[ILV]-X3-
452 [FILMV]-X5-C-X5-[ILMV]-[ILMV]-X(3)-L-X2-[IV]-[FILMV] and KA-X2-E-X-D-[ILV]-
453 [FLY]-[FILMV]-X2-[ILMV]-[EKQZ] (residues that are not conserved in the TE8 protein are
454 underlined).

455 **Requirement *in vitro* of the genome regions deleted in the three forms of strain 4295**

456 The m1, m2, and M forms of strain 4295 exhibit partially overlapping deletions in
457 their genomes (Fig. 2). The observation that all three forms lack the region corresponding to
458 coordinates 4780-14989 in strain 1976 indicates that this 10,210 bp sequence is not essential
459 for viral replication *in vitro*. As the three forms resulted from a co-culture, it is possible that
460 each of them requires complementation by the others in order to provide the functions that
461 have been lost. To test this hypothesis, strain 4295 was subcloned by limiting dilution, and the
462 subclones were analyzed by PCR (Fig. 7A, B). This experiment demonstrated that the three
463 forms are capable of growing *in vitro*, despite relatively large deletions. To determine whether
464 the deletions have quantitative effects on viral growth, the three subclones were compared
465 using a multi-step growth assay. They replicated comparably with each other and strain 1976,
466 thereby demonstrating that none of the deleted genes affected viral replication in the assay

467 used (Fig. 7C). Finally, the morphogenic properties of strains 1976 and 4295 were compared
468 by electron microscopic examination of infected TH-1 cells (Fig. 1). In a blind test, it was not
469 possible to differentiate the two strains, both of which exhibited all the features typical of
470 TeHV-3 and herpesviruses in general.

471 **Pathogenesis of strain 4295**

472 The tortoises that are most sensitive to TeHV-3 infection belong to the genus *Testudo*
473 (58). All species in this genus are protected by the Convention on International Trade in
474 Endangered Species of wild fauna and flora (CITES). Consequently, their use in *in vivo*
475 experiments is highly restricted and, moreover, carries the mandatory condition that the
476 scientific objective of any such experiments must contribute to species conservation. As
477 indicated above, no safe and efficacious vaccine is yet available against TeHV-3. The
478 observation that strain 4295 consists of a mixture of three deletion mutants prompted us to test
479 the potential of this mixture as an attenuated vaccine. With this goal in mind, three tortoises
480 were inoculated with strain 4295 (the same passage that the one used for sequencing and
481 which was shown to be a mixture of the m1, m2 and M forms) by intranasal instillation, and
482 one sentinel tortoise was mock-infected (Fig. 8). Our intention was to observe the animals for
483 2 months, in order to assess the safety of the inoculated material, and then to challenge them
484 with strain 1976 and evaluate the immune protection conferred. However, all three infected
485 tortoises became apathetic at about 20 days post-inoculation (PI), and then reduced feeding
486 progressively, to become anorexic between 22 and 29 days PI. Significant nasal discharge was
487 observed by 30 to 37 days PI, with mild blepharoedema. Diphtheritic plaques were observed
488 in the buccal cavity of one tortoise (4295/41 D). The animals were euthanized at 38 and 41
489 days PI due to extreme weakness (defined as one of the endpoints by the local ethics
490 committee). The mock-infected tortoise did not show any symptoms throughout the course of

491 the experiment. It was euthanized at 75 days PI in order to serve as a negative control for
492 further analyses (see below).

493 At necropsy, few macroscopic lesions were observed. A single tortoise (4295/41 D)
494 exhibited diphtheritic membranes on the buccal and esophageal mucosa. No other lesions
495 were observed in the infected tortoises or the mock-infected tortoise. Histopathological
496 analysis of various organs (lung, spleen, brain, kidney, and liver) of the infected tortoises
497 revealed mild lesions consistent with previous reports (8, 13) (Fig. 9). These lesions consisted
498 of interstitial heterophilic infiltration of the lung, congestion and heterophilic infiltration of
499 the red pulp of the spleen, and glial cell infiltration of the brain (telencephalon). The kidney
500 and the liver did not show significant histopathological modifications. Eosinophilic
501 intranuclear inclusion bodies were not detected in any of the samples examined.

502 To investigate tissue tropism, viral load was estimated in various organs (lung, spleen,
503 brain, kidney, and liver) using a qPCR method developed for the purpose. Consistent with
504 earlier reports (8, 13, 16) , the virus was detected in all organs tested (Fig. 10A). However, in
505 all three infected tortoises, the highest viral load was detected in the brain ($p<0.05$), slightly
506 greater than the viral load in the spleen. The other organs revealed lower and comparable viral
507 loads ($p<0.01$). At three week intervals, peripheral mononuclear blood cells were collected
508 during the entire course of the experiment and subjected to qPCR analyses. TeHV-3 was not
509 detected in any of the samples (data not shown).

510 Finally, we used the PCR assays described above to determine whether the viral load
511 detected in the various organs in Fig. 10A represented all three forms of 4295 (Fig. 10B). The
512 m1 form was not detected in any sample. The m2 form was detected with the highest
513 frequency, and was identified in all organs shown to be positive for TeHV-3 by PCR of a
514 gene present in all forms (UL13). Like the m2 form, the M form was detected in the brain of
515 all infected subjects. However, its presence in the other organs was reduced compared to the

516 m2 form. The M form was detected in only a fraction of the lungs and the spleens positive for
517 the m2 form, and, with exception of one sample, it was not detected in the kidney and liver
518 samples that were positive for the m2 form.

519 As the PCR data presented above were derived from tortoises that had been
520 co-infected with a mixture of the three forms present in strain 4295, they should be interpreted
521 cautiously in terms of the fitness of individual forms. Indeed, even if it might be viewed as
522 unlikely, *in vivo* interactions between the forms, such as helper effects resulting from
523 superinfection of the same host cells or secretion of soluble factors, cannot be excluded.
524 Consequently, the conclusions described below will need to be confirmed by experimental
525 infection of animals by the individual genotypes. The ability of the m2 form to spread
526 throughout the body suggests that the region of the TeHV-3 genome encompassing genes TE3
527 to TE11 (corresponding to the deletion in this form) is not essential for virulence and
528 neuroinvasiveness. Similarly, the ability of the M form to invade the brain suggests that the
529 region encoding genes TE4 to TE16 is not essential for neuroinvasiveness. However,
530 compared to the m2 form, the restricted tropism of the M form for the other organs suggests
531 that one or more genes in the region containing genes TE12 to TE16 may contribute to viral
532 spread *in vivo*.

533 Setting aside any possible interactions that may have occurred between the three forms
534 *in vivo*, the results of the analyses presented in Fig. 10B suggest that the region encompassing
535 TE4 to TE11 (the deletion common to all three forms), as well as not being required for viral
536 growth *in vitro*, is not essential for virulence of TeHV-3 *in vivo*. Importantly, the inability of
537 the m1 form to spread *in vivo* (despite co-infection with the two other genotypes)
538 demonstrates that the region encompassing genes UL55B to TE19 contains one or more genes
539 that are essential (alone or in combination) for viral spread *in vivo*. As the m1 form was
540 shown to grow *in vitro* as efficiently as the two other forms (m2 and M) and strain 1976, the

541 results of the present study encourage the testing of the m1 subclone of strain 4295 as an
542 attenuated vaccine candidate. While this study was under review, a version of the strain 1976
543 genome sequence became available (59). In comparison with our sequence, this contains a
544 9521 nt deletion (nt 5469-14989) starting in TE5 and ending near the left end of U_L, and thus
545 lacks seven genes, including TE7 and TE8. The similarity in location between this deletion
546 and those in strain 4295 indicates that the TeHV-3 genome has a propensity for losing
547 information in this region. Based on the pathogenic properties of strain 4295, we predict that
548 the deletion in strain 1976 will not affect viral spread *in vivo*. However, the sequence also has
549 frameshifts in TE19 that, if not due to error, may affect virulence, since this gene is absent
550 from the m1 form of strain 4295.

551 Our study reports the complete genome sequences of TeHV-3 strains 1976 and 4295,
552 the latter comprising a mixture of three deletion mutants that were sequenced as the mixture
553 and as subclones. The sequence of strain 1976 revealed a novel genome structure in the
554 family *Herpesviridae*. Genetic analysis highlighted the presence of genes related to cellular
555 genes (SEMA and IL-10) that have not been reported previously in alphaherpesviruses.
556 Phylogenetic analysis showed that TeHV-3 is most closely related to turtle herpesviruses, and
557 suggested the classification of this virus in the genus *Scutavirus*. Importantly, *in vitro* and *in*
558 *vivo* analyses demonstrated that the TeHV-3 genome contains large regions that are essential
559 neither for viral replication *in vitro* nor for virulence *in vivo*. They also indicated that spread
560 of the m1 form of strain 4295 is attenuated *in vivo*, thus indicating that this form would be a
561 good starting point for the development of vaccine candidates. In conclusion, the present
562 study represents a major step towards the characterization of an important viral pathogen of
563 tortoises and the development of effective prophylactic measures against TeHV-3 disease.

564 **ACKNOWLEDGEMENTS**

565 This work was supported by the University of Liège, the Belgian Science Policy
566 (Belspo) (BELVIR IAP7/45), the Fonds National Belge de la Recherche Scientifique (FNRS),
567 and the UK Medical Research Council (grant number MC_UU_12014/3). B.D. is a research
568 fellow of the FNRS. We thank Wai Kwong Lee and Andrew Carswell (BHF Glasgow
569 Cardiovascular Research Centre, University of Glasgow) for providing Sanger DNA
570 sequencing services.

REFERENCES

1. **Pellett PE, Davison AJ, Eberle R, Ehlers B, Hayward GS, Lacoste V, Minson AC, Nicholas J, Roizman B, Studdert MJ, Wang F.** 2012. Order - Herpesvirales, p 99-107. *In* King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), Virus Taxonomy doi:http://dx.doi.org/10.1016/B978-0-12-384684-6.00005-7. Elsevier, San Diego.
2. **Pellett PE, Roizman B.** 2013. Herpesviridae, p 1802-1822. *In* Knipe DM, Howley PM (ed), Fields Virology, 6th ed, vol 2. Lippincott Williams & Wilkins, Philadelphia, PA.
3. **Bicknese EJ, Childress AL, Wellehan JF, Jr.** 2010. A novel herpesvirus of the proposed genus Chelonivirus from an asymptomatic bowsprit tortoise (*Chersina angulata*). J Zoo Wildl Med **41:**353-358.
4. **Ackermann M, Koriabine M, Hartmann-Fritsch F, de Jong PJ, Lewis TD, Schetle N, Work TM, Dagenais J, Balazs GH, Leong JA.** 2012. The genome of Chelonid herpesvirus 5 harbors atypical genes. PLoS One **7:**e46623.
5. **Origgi FC.** 2006. Herpesvirus in tortoises, p 814-821. *In* Mader DR (ed), Reptile Medicine and Surgery, 2nd ed. Saunders, Saint-Louis, MO.
6. **Marschang RE, Gleiser CB, Papp T, Pfitzner AJ, Bohm R, Roth BN.** 2006. Comparison of 11 herpesvirus isolates from tortoises using partial sequences from three conserved genes. Vet Microbiol **117:**258-266.
7. **Origgi FC, Klein PA, Mathes K, Blahak S, Marschang RE, Tucker SJ, Jacobson ER.** 2001. Enzyme-linked immunosorbent assay for detecting herpesvirus exposure in Mediterranean tortoises (spur-thighed tortoise [Testudo graeca] and Hermann's tortoise [Testudo hermanni]). J Clinical Microbiol **39:**3156-3163.

595 8. **Origgi FC, Romero CH, Bloom DC, Klein PA, Gaskin JM, Tucker SJ, Jacobson**
596 **ER.** 2004. Experimental transmission of a herpesvirus in Greek tortoises (Testudo
597 graeca). Vet Pathol **41:**50-61.
598 9. **Marschang RE, Gravendyck M, Kaleta EF.** 1997. Herpesviruses in tortoises:
599 investigations into virus isolation and the treatment of viral stomatitis in Testudo
600 hermanni and T. graeca. J Vet Med Series B **44:**385-394.
601 10. **Muro J, Ramis A, Pastor J, Velarde R, Tarres J, Lavin S.** 1998. Chronic rhinitis
602 associated with herpesviral infection in captive spur-thighed tortoises from Spain. J
603 Wildl Dis **34:**487-495.
604 11. **Martel A, Blahak S, Vissenaekens H, Pasmans F.** 2009. Reintroduction of clinically
605 healthy tortoises: the herpesvirus Trojan horse. J Wildl Dis **45:**218-220.
606 12. **Zapata AG, Varas A, Torroba M.** 1992. Seasonal variations in the immune system
607 of lower vertebrates. Immunol Today **13:**142-147.
608 13. **Teifke JP, Lohr CV, Marschang RE, Osterrieder N, Posthaus H.** 2000. Detection
609 of chelonid herpesvirus DNA by nonradioactive in situ hybridization in tissues from
610 tortoises suffering from stomatitis-rhinitis complex in Europe and North America. Vet
611 Pathol **37:**377-385.
612 14. **Une Y, Uemura K, Nakano Y, Kamiie J, Ishibashi T, Nomura Y.** 1999.
613 Herpesvirus infection in tortoises (Malacochersus tornieri and Testudo horsfieldii).
614 Vet Pathol **36:**624-627.
615 15. **Hervas J, Sanchez-Cordon PJ, de Chacon Lara F, Carrasco L, Gomez-**
616 **Villamandos JC.** 2002. Hepatitis associated with herpes viral infection in the tortoise
617 (Testudo horsfieldii). J Vet Med B Infect Dis Vet Public Health **49:**111-114.

618 16. **Marschang RE, Milde K, Bellavista M.** 2001. Virus isolation and vaccination of
619 Mediterranean tortoises against a chelonid herpesvirus in a chronically infected
620 population in Italy. Dtsch Tierarztl Wochenschr **108:**376-379.
621 17. **Blahak S, Steiner M.** 2010. Epidemiological investigations into the persistence of
622 antibodies against tortoise herpesvirus strains in the species *T. graeca*, *T. hermanni*
623 and *T. horsfieldii*, abstr 1st Intenational Conference on Reptile and Amphibian
624 Medicine, Munich, March 4th, 2010. Verlag, Munich, Germany.
625 18. **Marschang RE, Frost JW, Gravendyck M, Kaleta EF.** 2001. Comparison of 16
626 chelonid herpesviruses by virus neutralization tests and restriction endonuclease
627 digestion of viral DNA. J Vet Med B Infect Dis Vet Public Health **48:**393-399.
628 19. **Clark HF, Karzon DT.** 1967. Terrapene heart (TH-1), a continuous cell line from the
629 heart of the box turtle Terrapene carolina. Exp Cell Res **48:**263-268.
630 20. **Lete C, Palmeira L, Leroy B, Mast J, Machiels B, Wattiez R, Vanderplasschen A,**
631 **Gillet L.** 2012. Proteomic characterization of bovine herpesvirus 4 extracellular
632 virions. J Virol **86:**11567-11580.
633 21. **Markine-Goriaynoff N, Georgin JP, Goltz M, Zimmermann W, Broll H,**
634 **Wamwayi HM, Pastoret PP, Sharp PM, Vanderplasschen A.** 2003. The core 2
635 beta-1,6-N-acetylglucosaminyltransferase-mucin encoded by bovine herpesvirus 4 was
636 acquired from an ancestor of the African buffalo. J Virol **77:**1784-1792.
637 22. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly
638 using de Bruijn graphs. Genome Res **18:**821-829.
639 23. **Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I.** 2009. ABySS: a
640 parallel assembler for short read sequence data. Genome Res **19:**1117-1123.
641 24. **Ewing B, Hillier L, Wendl MC, Green P.** 1998. Base-calling of automated
642 sequencer traces using phred. I. Accuracy assessment. Genome Res **8:**175-185.

643 25. **Ewing B, Green P.** 1998. Base-calling of automated sequencer traces using phred. II.
644 Error probabilities. Genome Res **8:**186-194.
645 26. **Tsai IJ, Otto TD, Berriman M.** 2010. Improving draft assemblies by iterative
646 mapping and assembly of short reads to eliminate gaps. Genome Biol **11:**R41.
647 27. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler
648 transform. Bioinformatics **26:**589-595.
649 28. **Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD,**
650 **Marshall D.** 2013. Using Tablet for visual exploration of second-generation
651 sequencing data. Brief Bioinform **14:**193-202.
652 29. **Gatherer D, Seirafian S, Cunningham C, Holton M, Dargan DJ, Baluchova K,**
653 **Hector RD, Galbraith J, Herzyk P, Wilkinson GW, Davison AJ.** 2011. High-
654 resolution human cytomegalovirus transcriptome. Proc Natl Acad Sci U S A
655 **108:**19755-19760.
656 30. **Wilkie GS, Davison AJ, Watson M, Kerr K, Sanderson S, Bouts T, Steinbach F,**
657 **Dastjerdi A.** 2013. Complete genome sequences of elephant endotheliotropic
658 herpesviruses 1A and 1B determined directly from fatal cases. J Virol **87:**6700-6712.
659 31. **Costes B, Fournier G, Michel B, Delforge C, Raj VS, Dewals B, Gillet L, Drion P,**
660 **Body A, Schynts F, Lieffrig F, Vanderplasschen A.** 2008. Cloning of the koi
661 herpesvirus genome as an infectious bacterial artificial chromosome demonstrates that
662 disruption of the thymidine kinase locus induces partial attenuation in Cyprinus carpio
663 koi. J Virol **82:**4955-4964.
664 32. **Costes B, Raj VS, Michel B, Fournier G, Thirion M, Gillet L, Mast J, Lieffrig F,**
665 **Bremont M, Vanderplasschen A.** 2009. The major portal of entry of koi herpesvirus
666 in Cyprinus carpio is the skin. J Virol **83:**2819-2830.

667 33. **Vanderplasschen A, Bublot M, Dubuisson J, Pastoret P-P, Thiry E.** 1993.
668 Attachment of the Gammaherpesvirus Bovine Herpesvirus 4 Is Mediated by the
669 Interaction of gp8 Glycoprotein with Heparinlike Moieties on the Cell Surface.
670 Virology **196:**232-240.
671 34. **Dewals B, Boudry C, Gillet L, Markine-Goriaynoff N, de Leval L, Haig DM,**
672 **Vanderplasschen A.** 2006. Cloning of the genome of Alcelaphine herpesvirus 1 as an
673 infectious and pathogenic bacterial artificial chromosome. J Gen Virol **87:**509-517.
674 35. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and
675 high throughput. Nucleic Acids Res **32:**1792-1797.
676 36. **Kumar S, Nei M, Dudley J, Tamura K.** 2008. MEGA: a biologist-centric software
677 for evolutionary analysis of DNA and protein sequences. Brief Bioinform **9:**299-306.
678 37. **Drummond AJ, Suchard MA, Xie D, Rambaut A.** 2012. Bayesian phylogenetics
679 with BEAUti and the BEAST 1.7. Mol Biol Evol **29:**1969-1973.
680 38. **Le SQ, Gascuel O.** 2008. An improved general amino acid replacement matrix. Mol
681 Biol Evol **25:**1307-1320.
682 39. **Drummond AJ, Ho SY, Phillips MJ, Rambaut A.** 2006. Relaxed phylogenetics and
683 dating with confidence. PLoS Biol **4:**e88.
684 40. **Gernhard T.** 2008. The conditioned reconstructed process. J Theor Biol **253:**769-778.
685 41. **Gatherer D.** 2007. Letter: TreeAdder: A Tool to Assist the Optimal Positioning of a
686 New Leaf into an Existing Phylogenetic Tree. Open Bioinfo J **1:**1-2.
687 42. **Ouyang P, Rakus K, van Beurden SJ, Westphal AH, Davison AJ, Gatherer D,**
688 **Vanderplasschen AF.** 2014. IL-10 encoded by viruses: a remarkable example of
689 independent acquisition of a cellular gene by viruses and its subsequent evolution in
690 the viral genome. J Gen Virol **95:**245-262.

691 43. **Schmidt HA, Strimmer K, Vingron M, von Haeseler A.** 2002. TREE-PUZZLE:
692 maximum likelihood phylogenetic analysis using quartets and parallel computing.
693 Bioinformatics **18:**502-504.
694 44. **Shimodaira H, Hasegawa M.** 2001. CONSEL: for assessing the confidence of
695 phylogenetic tree selection. Bioinformatics **17:**1246-1247.
696 45. **Levitt M.** 1992. Accurate modeling of protein conformation by automatic segment
697 matching. J Mol Biol **226:**507-533.
698 46. **Fechteler T, Dengler U, Schomburg D.** 1995. Prediction of protein three-
699 dimensional structures in insertion and deletion regions: a procedure for searching data
700 bases of representative protein fragments using geometric scoring criteria. J Mol Biol
701 **253:**114-131.
702 47. **Wang J, Cieplak P, Kollman PA.** 2000. How well does a restrained electrostatic
703 potential (RESP) model perform in calculating conformational energies of organic and
704 biological molecules? J Comput Chem **21:**1049-1074.
705 48. **Labute P.** 2008. The generalized Born/volume integral implicit solvent model:
706 estimation of the free energy of hydration using London dispersion instead of atomic
707 surface area. J Comput Chem **29:**1693-1698.
708 49. **Ramachandran GN, Ramakrishnan C, Sasisekharan V.** 1963. Stereochemistry of
709 polypeptide chain configurations. J Mol Biol **7:**95-99.
710 50. **Telford EA, Watson MS, McBride K, Davison AJ.** 1992. The DNA sequence of
711 equine herpesvirus-1. Virology **189:**304-316.
712 51. **Telford EA, Watson MS, Perry J, Cullinane AA, Davison AJ.** 1998. The DNA
713 sequence of equine herpesvirus-4. J Gen Virol **79 (Pt 5):**1197-1203.

714 52. **Stow ND, Davison AJ.** 1986. Identification of a varicella-zoster virus origin of DNA
715 replication and its activation by herpes simplex virus type 1 gene products. J Gen
716 Virol **67 (Pt 8):**1613-1623.
717 53. **Koff A, Tegtmeyer P.** 1988. Characterization of major recognition sequences for a
718 herpes simplex virus type 1 origin-binding protein. J Virol **62:**4096-4103.
719 54. **Weir HM, Stow ND.** 1990. Two binding sites for the herpes simplex virus type 1
720 UL9 protein are required for efficient activity of the oriS replication origin. J Gen
721 Virol **71 (Pt 6):**1379-1385.
722 55. **Davison AJ.** 2010. Herpesvirus systematics. Vet Microbiol **143:**52-69.
723 56. **Myster F, Palmeira L, Sorel O, Bouillenne F, DePauw E, Schwartz-Cornil I,**
724 **Vanderplasschen A, Dewals BG.** 2015. Viral semaphorin inhibits dendritic cell
725 phagocytosis and migration but is not essential for gammaherpesvirus-induced
726 lymphoproliferation in malignant catarrhal fever. J Virol **89:**3630-3647.
727 57. **Liu H, Juo ZS, Shim AH, Focia PJ, Chen X, Garcia KC, He X.** 2010. Structural
728 basis of semaphorin-plexin recognition and viral mimicry from Sema7A and A39R
729 complexes with PlexinC1. Cell **142:**749-761.
730 58. **Pasmans F, Blahak S, Martel A, Pantchev N.** 2008. Introducing reptiles into a
731 captive collection: the role of the veterinarian. Vet J **175:**53-68.
732 59. **Origgi FC, Tecilla M, Pilo P, Aloisio F, Otten P, Aguilar-Bultet L, Sattler U,**
733 **Roccabianca P, Romero CH, Bloom DC, Jacobson ER.** 2015. A Genomic
734 Approach to Unravel Host-Pathogen Interaction in Chelonians: The Example of
735 Testudinid Herpesvirus 3. PLoS One **10:**e0134897.
736

737 **FIGURE LEGENDS**

738 **FIG 1** Morphology of TeHV-3. TH-1 cells were infected with strains 1976 and 4295. Six
739 days post-infection, cells were processed for electron microscopy examination. In a blind test,
740 it was not possible to differentiate the two strains. The images in this figure represent TH-1
741 cells infected with strain 1976. (A) General view of the various compartments of an infected
742 cell. (B) Nucleoplasm. (C) Nucleus and cytoplasm. (D) Cytoplasm and extracellular space.

743 **FIG 2** Map of the TeHV-3 strain 1976 genome. The unique regions (U_T, U_L, and U_S) are
744 shaded white, and the inverted repeats (TR_T, IR_T, IR_S, and TR_S) are shaded light yellow.
745 Predicted functional ORFs are depicted by colored arrows, with nomenclature below. Red
746 shading indicates ORFs inherited from the ancestor of the alpha-, beta-, and
747 gammaherpesviruses. Blue shading indicates ORFs that have orthologs in mammalian or
748 avian alphaherpesviruses. Light orange shading indicates the remaining ORFs. Introns are
749 shown as narrow white bars. Reiterations are shown by grey shading within or between ORFs,
750 and three copies of ori by vertical, red bars. The deletions and the duplications present in the
751 m1, m2, and M forms of TeHV-3 strain 4295 are marked above the genome by horizontal
752 green and orange bars, respectively.

753 **FIG 3** Relative orientations of unique regions in the TeHV-3 strain 1976 genome. (A) A
754 schematic representation of the genome is shown at the top, the orientations of U_T, U_L, and U_S
755 corresponding to those in Fig. 2. Below this, the four possible combinations of the
756 orientations of U_T-U_L and U_L-U_S are presented, with the majority of U_L omitted. For each
757 combination, the sizes (kb) of restriction endonuclease fragments at or near the genome ends
758 or the U_T-U_L and U_L-U_S junctions are shown (EcoRI above the genome and KpnI below).
759 White bars (P1 to P6) represent the positions of the probes used for hybridization. Asterisks

760 highlight restriction endonuclease fragments for which a positive band was observed in
761 Southern blot analyses. (B) The panels on the right show a Southern blot analysis of TeHV-3
762 strain 1976 DNA digested with EcoRI or KpnI and hybridized to probes P1 to P6. Black
763 arrowheads indicate all bands compatible with the predicted fragments. The panel on the left
764 shows an ethidium bromide-stained profile of marker fragments (MS) and strain 1976 DNA
765 digested with EcoRI or KpnI. The 0.65 and 0.72 kb fragments were detected, but are not
766 visible on these images, which have been restricted to fragments above 2 kb.

767 **FIG 4** Phylogenetic analyses. In each panel, the scale indicates substitutions per site.
768 Abbreviations for herpesvirus names: El, elephantid; H, human; Mu, murid; E, equid; F, felid;
769 Bo, bovid; S, suid; Ce, cercopithecine; Pt, pteropodid; Pn, panine; Mc, macacine; Pa, papiine;
770 Sa, saimiriine; Ga, gallid; Me, meleagrid; Co, columbid; An, anatid; Ac, acciptrid; Ps,
771 psittacid; Te, testudinid; Ch, chelonid; Ov, ovine; Al, alcelaphine; Cy, cyprinid; and HV,
772 herpesvirus (followed by a hyphenated number). Other abbreviations: CNPV, canarypox
773 virus; FPV, fowlpox virus; PEPV, penguin poxvirus; PNPV, pigeon poxvirus; SGIV,
774 Singapore grouper iridovirus; VARV, variola virus; HSPV, horse poxvirus; VACV, vaccinia
775 virus; ECTV, ectromelia virus; CPXV, cowpox virus; RPXV, rabbit poxvirus; ORFV, ORF
776 virus; and SWPV, swinepox virus. (A) Bayesian tree for herpesvirus DNA polymerase
777 proteins. All nodes have posterior probabilities of 1. Viruses that have not yet been classified
778 are marked by asterisks (1). (B) Maximum likelihood tree for vertebrate SEMA-7A proteins
779 and their viral homologs. Node bootstrap values greater than 70 are marked. (C) Maximum
780 likelihood tree for IL-10 homologues. Node bootstrap values greater than 70 are marked.

781 **FIG 5** Homology model of the TeHV-3 SEMA (TE7) protein. (A) Mouse SEMA-3A PDB
782 1Q47. (B) Homology model of TeHV-3 TE7 constructed from 1Q47 with sub-optimal
783 residues shown. (C) Ramachandran plot of homology model used to identify sub-optimal

784 residues. (D) Mouse SEMA-3A PDB 1Q47 (blue) superposed with the homology model of
785 TeHV-3 TE7 (red).

786 **FIG 6** Homology model of the TeHV-3 IL-10 (TE8) protein. (A) Human IL-10 PDB 1ILK.
787 (B) Homology model of TeHV-3 IL-10 constructed from 1ILK with sub-optimal residues
788 shown. (C) Ramachandran plot of homology model used to identify sub-optimal residues. (D)
789 Human IL-10 PDB 1ILK (blue) superposed with the homology model of TeHV-3 IL-10 (red).

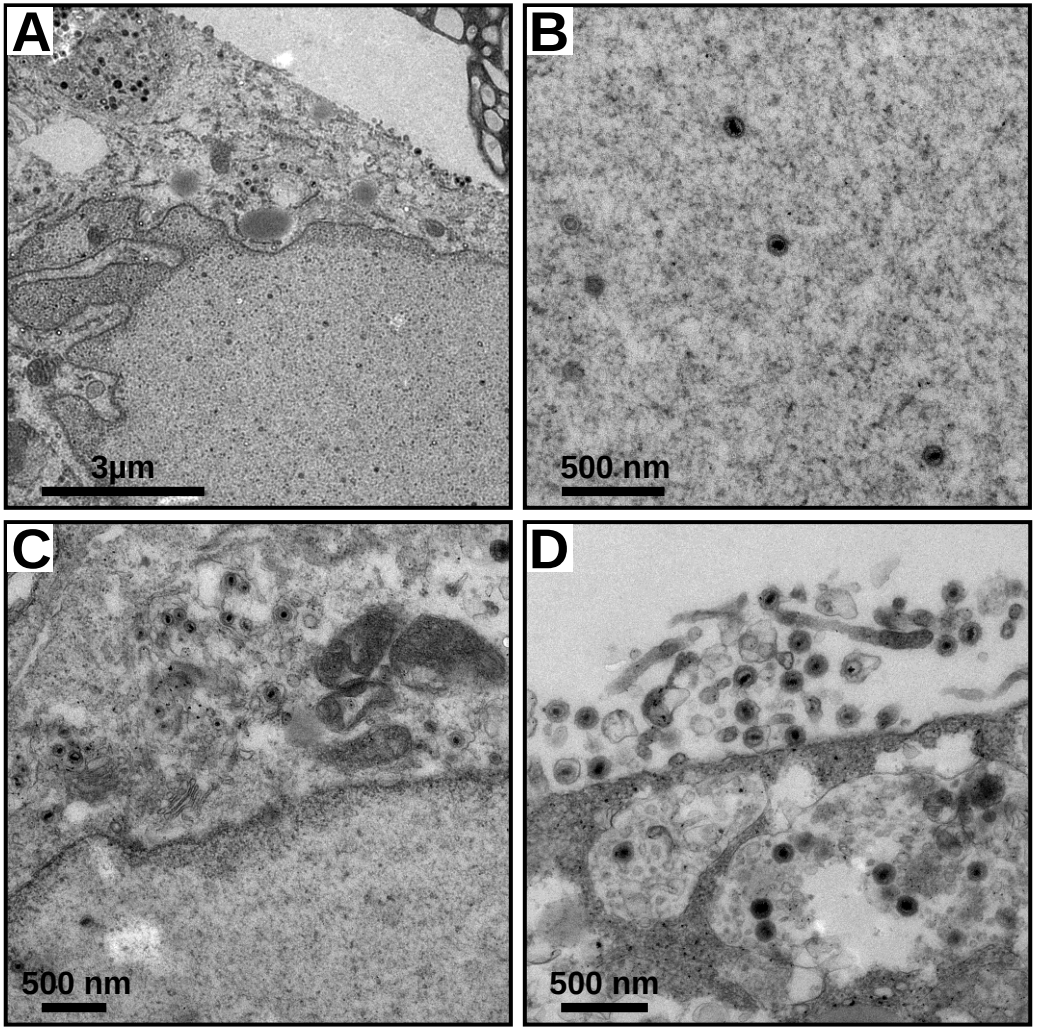
790 **FIG 7** Effects of the deletions in the m1, m2 and M forms of strain 4295 on viral growth *in*
791 *vitro*. (A) Schematic representation of the regions in the strain 1976 genome corresponding to
792 the deletions in the three forms of strain 4295, with the coordinates of the deletions indicated
793 above the strain 1976 genome. Arrows represent primers (Table 1) designed for PCR
794 amplification of the regions containing the deletions. Amplicon sizes are indicated in bold
795 below the genomes of the strain 4295 subclones. PCR amplification performed with the
796 UL13f and UL13r primers led to a product of 106 bp (B) Characterization by PCR of the
797 strain 4295 subclones representing the three genome forms. Strain 4295 prior to subcloning
798 (containing the three forms) and strain 1976 were used as controls. The positions of markers
799 (bp) are marked by arrowheads. (C) Multi-step growth curves of the strain 4295 subclones
800 and strain 1976.

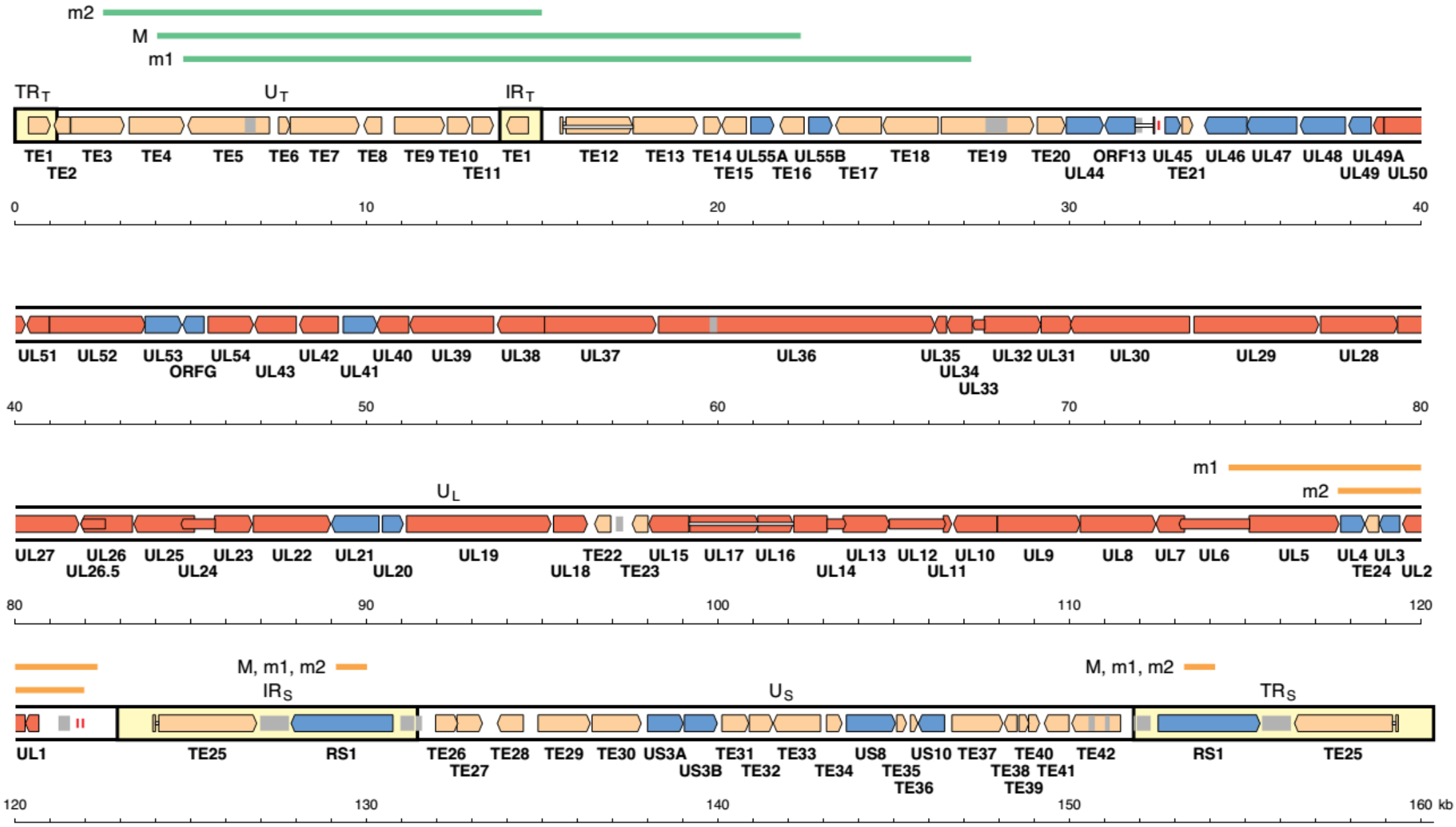
801 **FIG 8** Pathogenesis of strain 4295. On day 0, tortoises (n=4) aged 5.4 years (mean weight ±
802 SD: 219.5 g ± 53.1 g) were infected (n=3) with strain 4295 (consisting of a mixture of three
803 deletion mutants) or mock-infected (n=1). Percentage survival is expressed according to days
804 PI. The tortoises were named according to the following scheme: the inoculation performed
805 (4295 or mock)/the time PI at which death occurred (when more than one tortoise died on the
806 same day, they were further ranked by the addition of a lower case letter), and an upper case

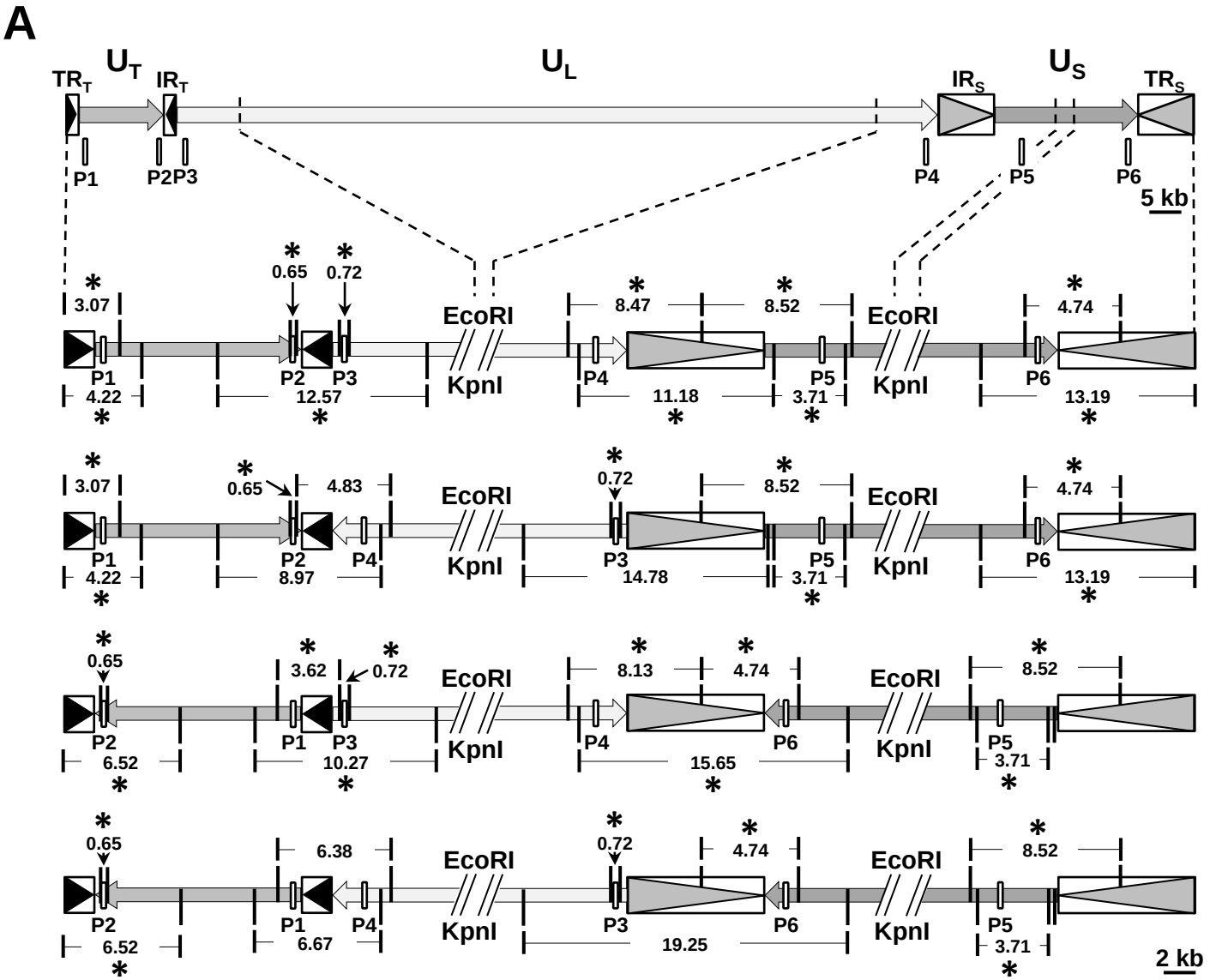
807 letter to describe the clinical state before euthanasia (D, diseased; H, healthy).

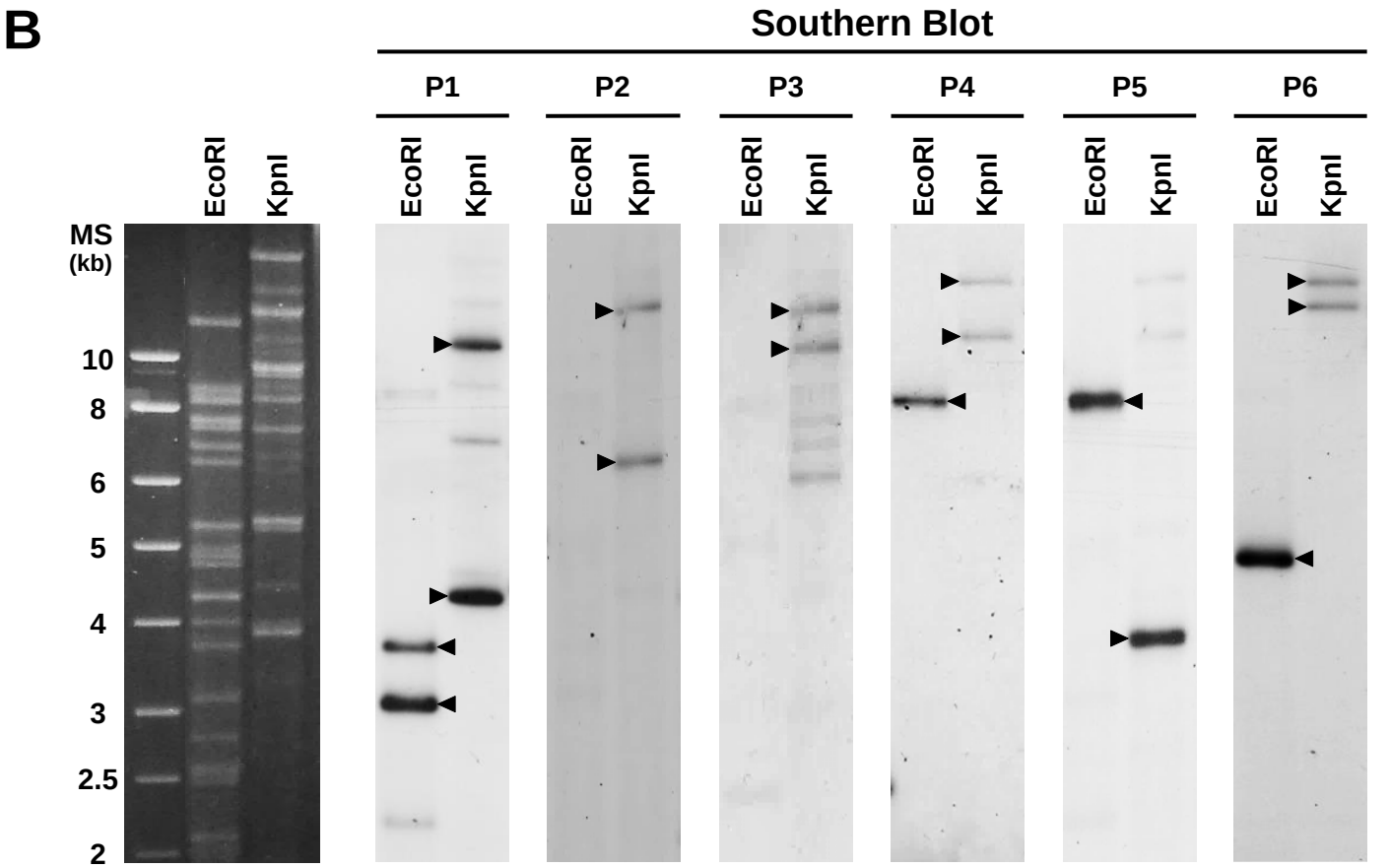
808 **FIG 9** Histopathological characterization of the lesions induced by strain 4295 (consisting of
809 a mixture of three deletion mutants). The indicated organs were collected from all tortoises at
810 the end of the experiment described in Fig. 8, and were processed for histological
811 examination. The images were collected from tortoises Mock/75 H and 4295/41 D, the latter
812 having been selected as representative of the infected group. H, heterophil; F, faveolae; GC,
813 glial cell; C, capillary; PT, proximal tubule; DT, distal tubule; G, glomerulus; S, sinusoid; and
814 M, melanomacrophage. Bars = 20 μm.

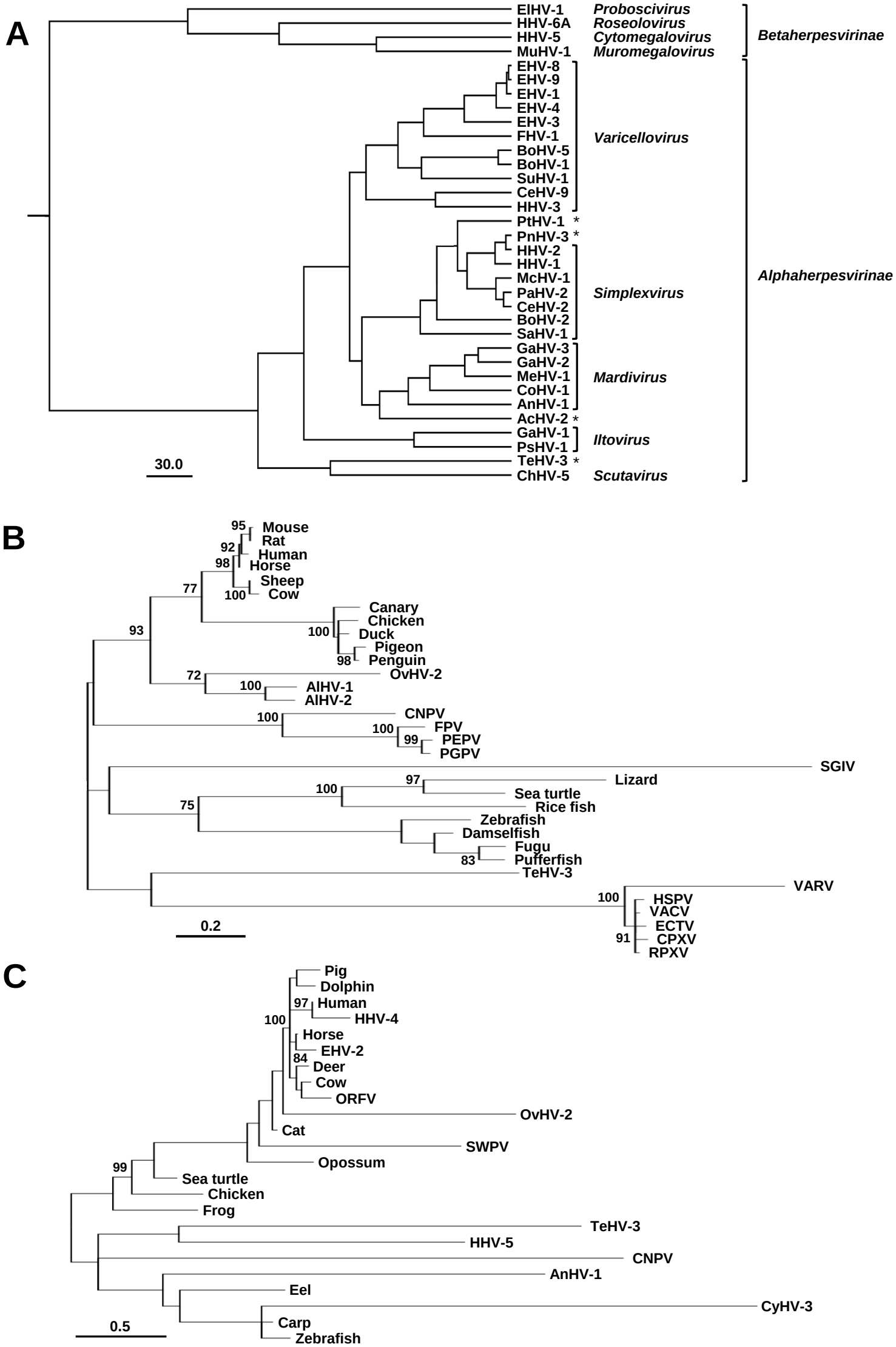
815 **FIG 10** Tissue tropism of strain 4295. DNA was extracted from the indicated organs of all
816 tortoises at the end of the experiment described in Fig. 8. (A) Analysis of viral gene copy
817 number by qPCR. Individual values represent the mean of triplicate measurements ± SD. (B)
818 Analysis of the presence of the three forms present in strain 4295, using primers specific for
819 each form (m1, m2, and M) and all three forms (UL13). See Fig. 7B and Table 1 for details of
820 the primers. Strain 4295 (containing the three forms) was used as positive (CT) control.
821

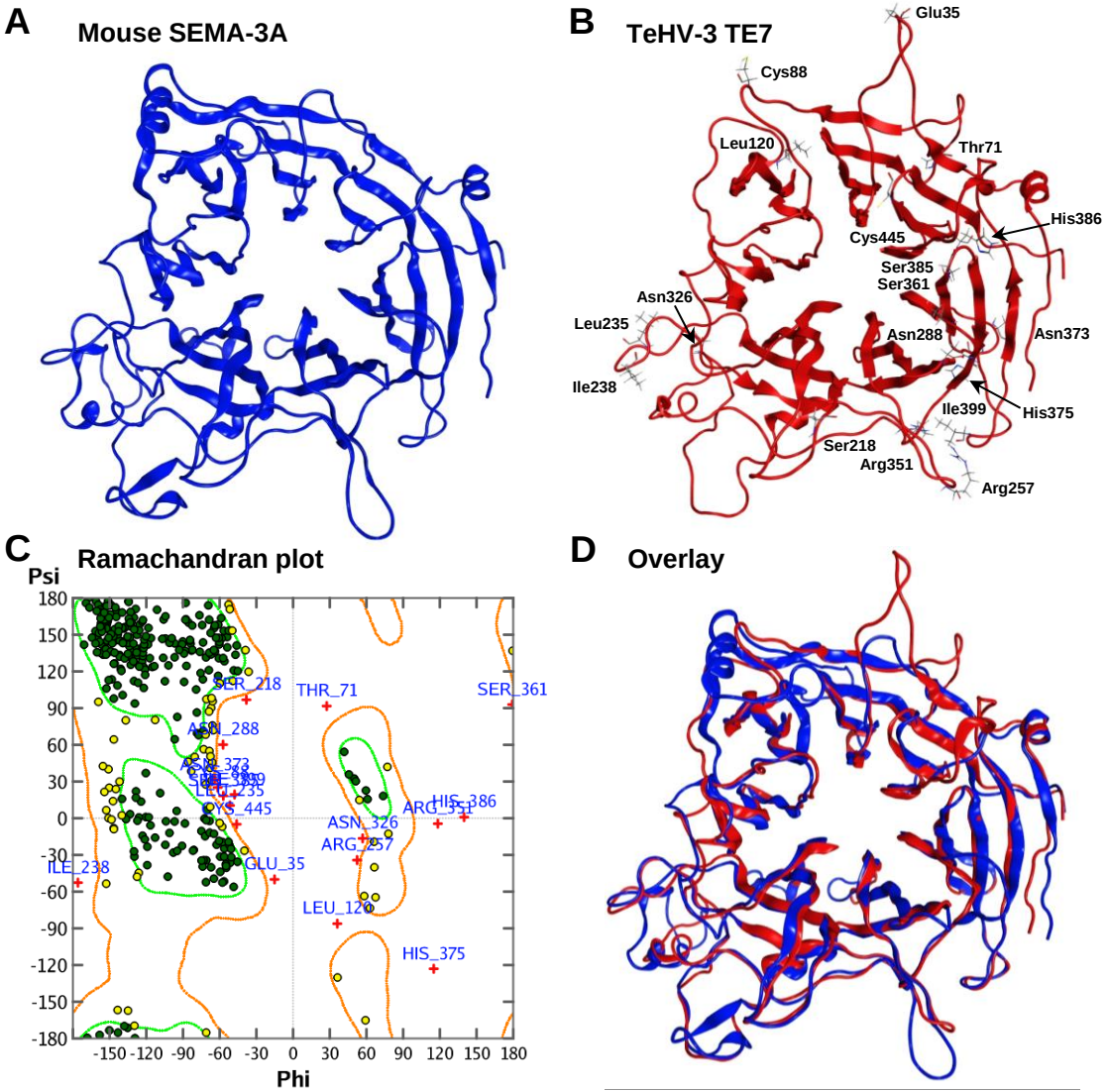

A
3µm
B
500 nm
C
500 nm
D
500 nm

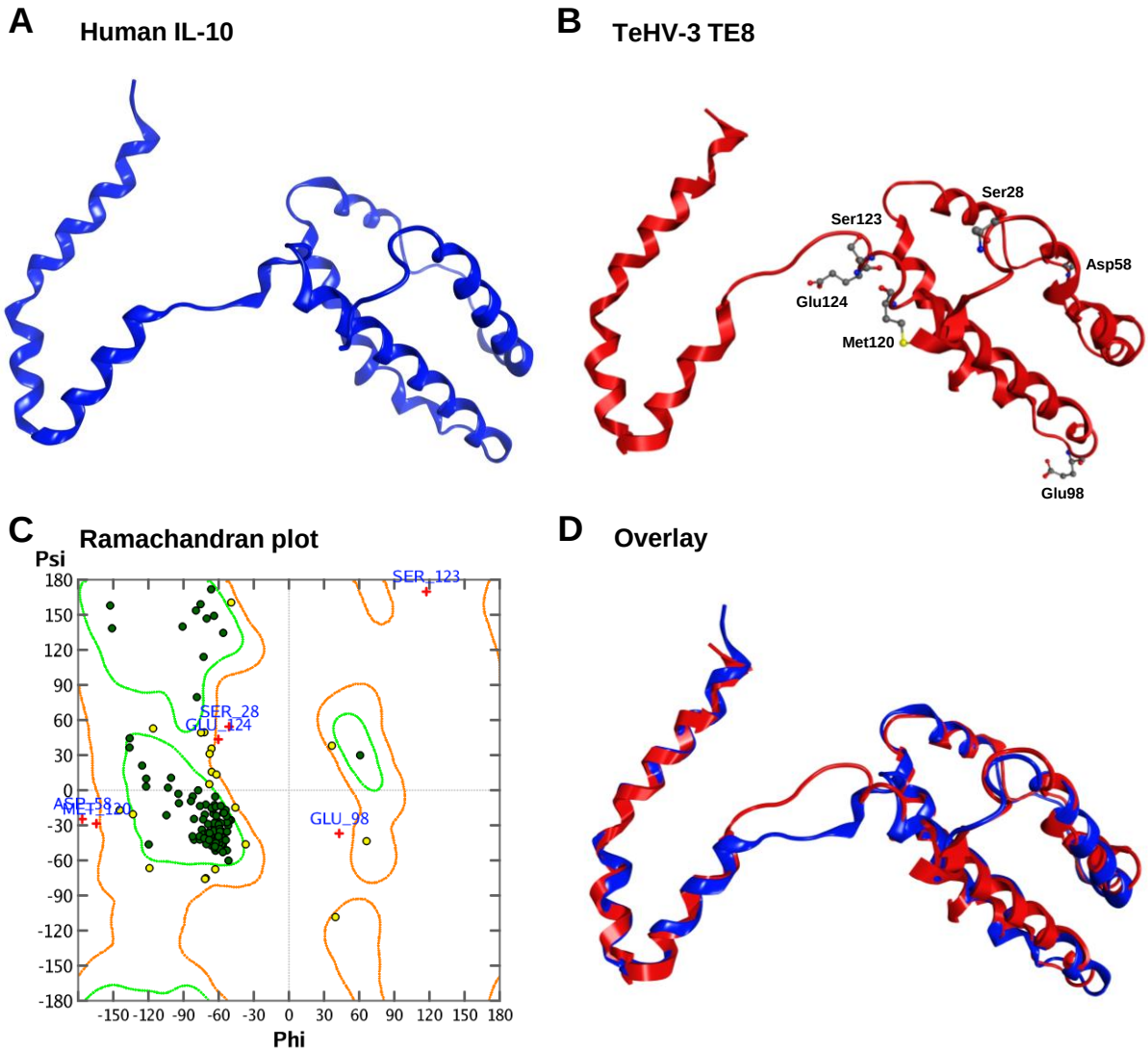

m2
M
m1
TR$_T$
U$_T$
IR$_T$
TE1 TE2 TE3 TE4 TE5 TE6 TE7 TE8 TE9 TE10 TE11 TE1 TE12 TE13 TE14 TE15 UL55A TE16 UL55B TE17 TE18 TE19 TE20 UL44 ORF13 UL45 TE21 UL46 UL47 UL48 UL49A UL49 UL50
0 10 20 30 40
UL51 UL52 UL53 ORFG UL54 UL43 UL42 UL41 UL40 UL39 UL38 UL37 UL36 UL35 UL34 UL33 UL32 UL31 UL30 UL29 UL28
40 50 60 70 80
m1
m2
U$_L$
UL27 UL26 UL26.5 UL25 UL24 UL23 UL22 UL21 UL20 UL19 UL18 TE22 TE23 UL15 UL17 UL16 UL14 UL13 UL12 UL11 UL10 UL9 UL8 UL7 UL6 UL5 UL4 UL3 TE24 UL2
80 90 100 110 120
M, m1, m2
M, m1, m2
IR$_S$
U$_S$
TR$_S$
UL1 TE25 RS1 TE26 TE27 TE28 TE29 TE30 US3A US3B TE31 TE32 TE33 TE34 US8 TE35 TE36 US10 TE37 TE38 TE39 TE40 TE41 TE42 RS1 TE25
120 130 140 150 160 kb

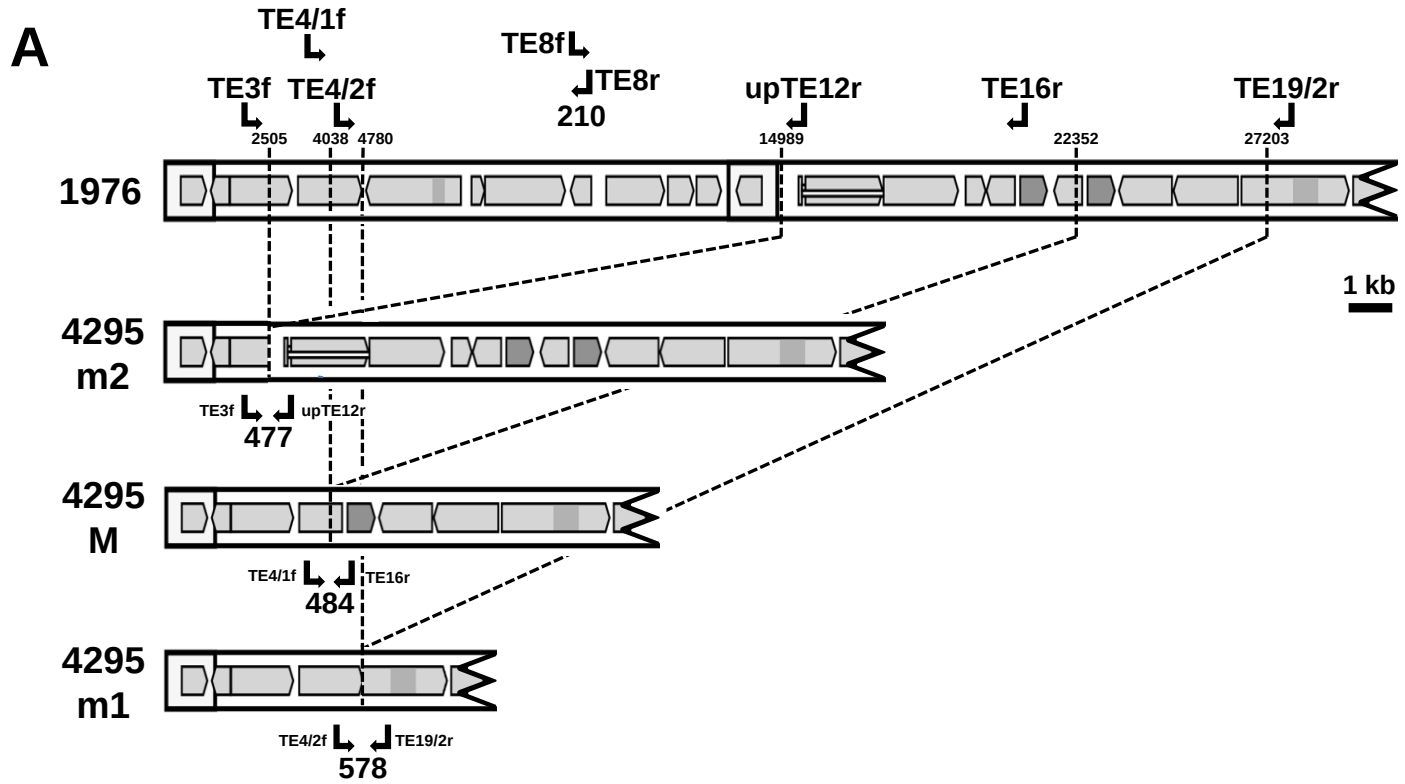

A
U_T
U_L
U_S
TR_T
IR_T
IR_S
TR_S
P1
P2P3
P4
P5
P6
5 kb
EcoRI
KpnI
3.07
0.65
0.72
8.47
8.52
4.74
4.22
12.57
11.18
3.71
13.19
4.83
8.97
14.78
3.62
6.52
10.27
8.13
15.65
6.38
6.67
19.25
2 kb

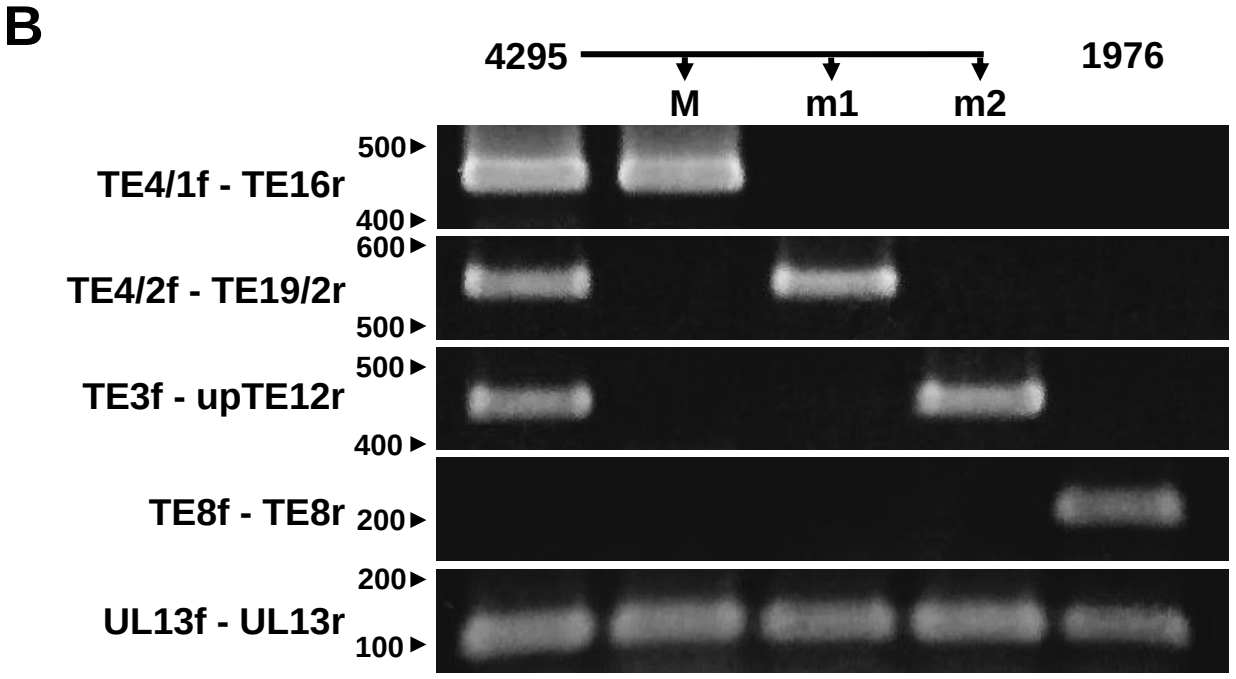

B
Southern Blot
P1
P2
P3
P4
P5
P6
EcoRI
KpnI
MS
(kb)
10
8
6
5
4
3
2.5
2

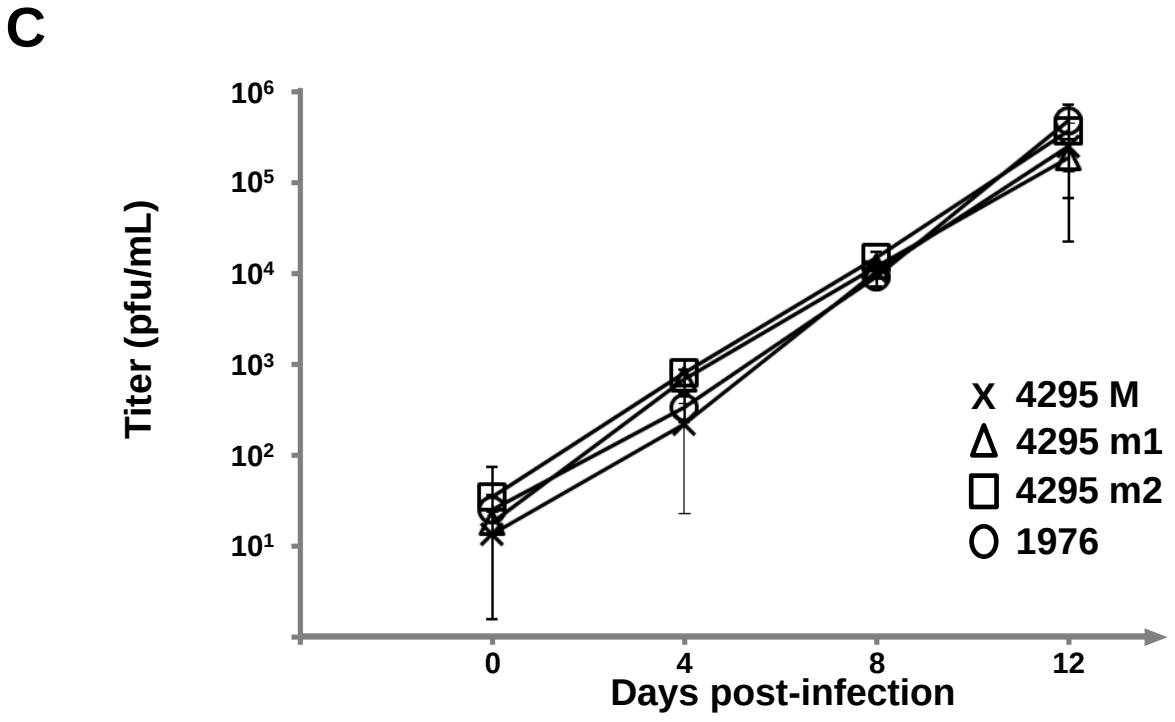

A
ElHV-1
HHV-6A
HHV-5
MuHV-1
Proboscivirus
Roseolovirus
Cytomegalovirus
Muromegalovirus
Betaherpesvirinae
EHV-8
EHV-9
EHV-1
EHV-4
EHV-3
FHV-1
BoHV-5
BoHV-1
SuHV-1
CeHV-9
HHV-3
Varicellovirus
PtHV-1 *
PnHV-3 *
HHV-2
HHV-1
McHV-1
PaHV-2
CeHV-2
BoHV-2
SaHV-1
Simplexvirus
GaHV-3
GaHV-2
MeHV-1
CoHV-1
AnHV-1
Mardivirus
AcHV-2 *
GaHV-1
PsHV-1
Iltovirus
TeHV-3 *
ChHV-5
Scutavirus
Alphaherpesvirinae
30.0
B
95
Mouse
Rat
92
Human
98
Horse
Sheep
77
100
Cow
Canary
Chicken
93
100
Duck
Pigeon
98
Penguin
72
OvHV-2
100
AlHV-1
AlHV-2
100
CNPV
100
FPV
99
PEPV
PGPV
SGIV
97
Lizard
100
Sea turtle
75
Rice fish
Zebrafish
Damselfish
Fugu
83
Pufferfish
TeHV-3
VARV
100
HSPV
VACV
ECTV
91
CPXV
RPXV
0.2
C
Pig
Dolphin
97
Human
100
HHV-4
Horse
EHV-2
84
Deer
Cow
ORFV
OvHV-2
Cat
SWPV
Opossum
99
Sea turtle
Chicken
Frog
TeHV-3
HHV-5
CNPV
AnHV-1
Eel
CyHV-3
Carp
Zebrafish
0.5

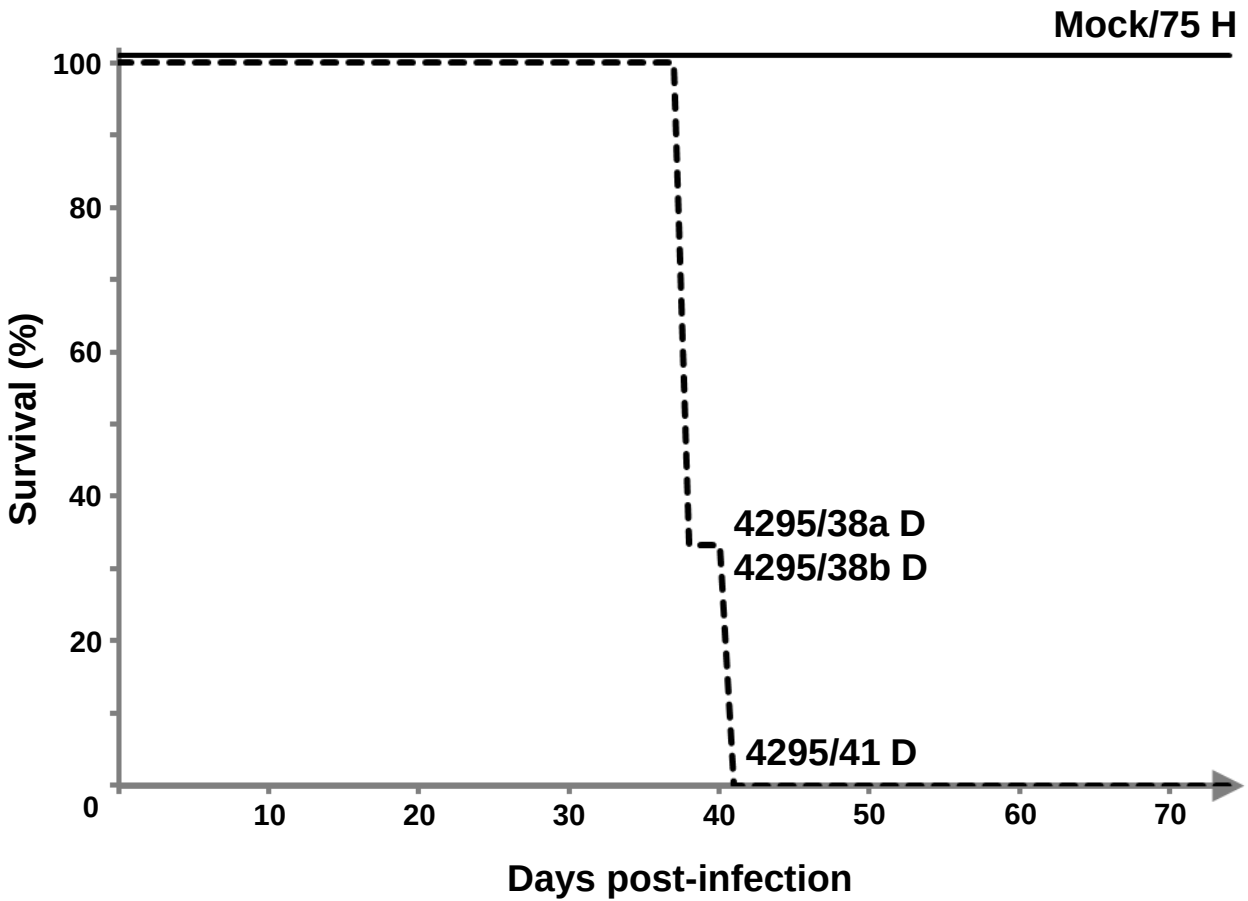

A
Mouse SEMA-3A
B
TeHV-3 TE7
Glu35
Cys88
Leu120
Thr71
His386
Cys445
Ser385
Ser361
Asn326
Leu235
Asn288
Asn373
Ile238
Ile399
His375
Ser218
Arg351
Arg257
C
Ramachandran plot
Psi
Phi
SER_218
THR_71
SER_361
ASN_288
LEU_235
CYS_445
ASN_326
ARG_257
ILE_238
GLU_35
LEU_120
HIS_375
D
Overlay

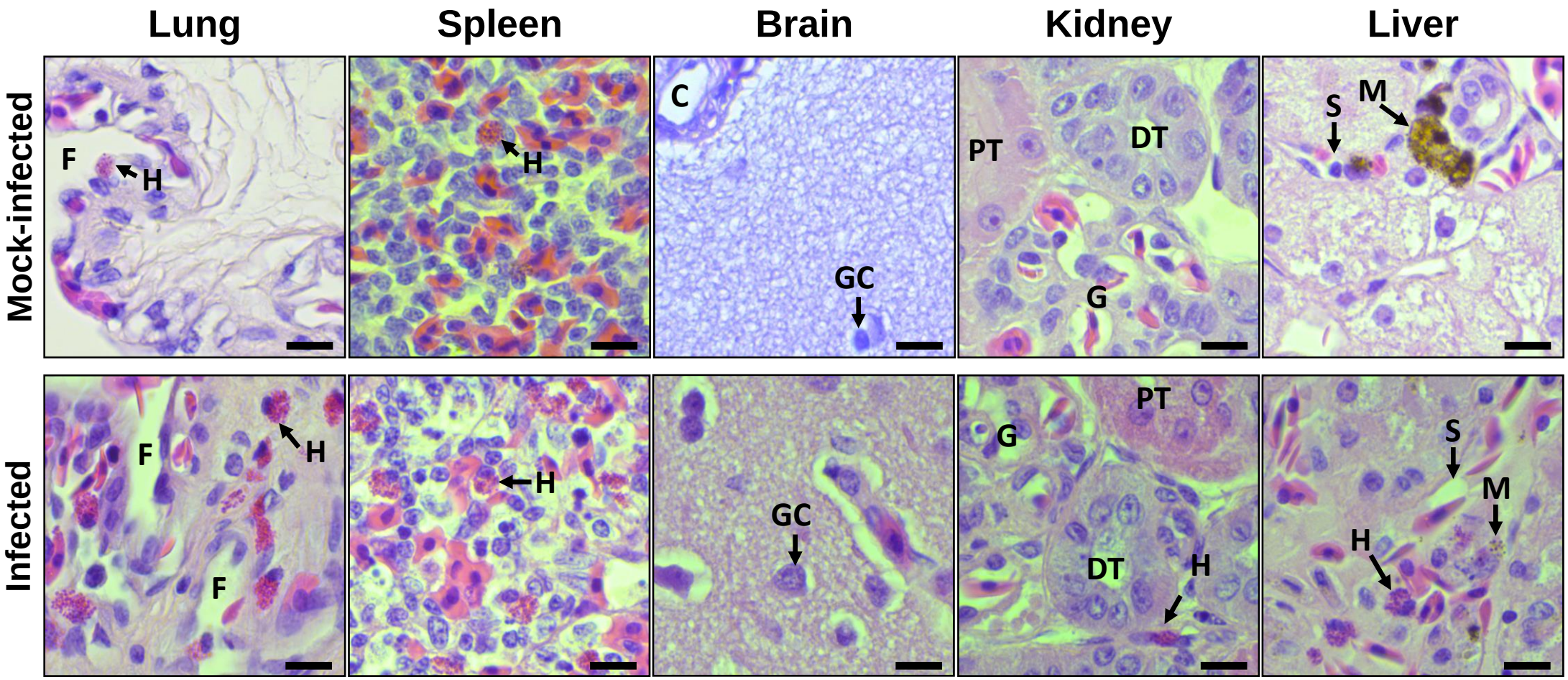

A
Human IL-10
B
TeHV-3 TE8
Ser28
Ser123
Asp58
Glu124
Met120
Glu98
C
Ramachandran plot
Psi
Phi
SER_123
SER_28
GLU_124
ASP_58
MET_120
GLU_98
D
Overlay


A
TE4/1f
TE8f
TE3f
TE4/2f
TE8r
210
upTE12r
TE16r
TE19/2r
2505
4038
4780
14989
22352
27203
1976
1 kb
4295
m2
TE3f
upTE12r
477
4295
M
TE4/1f
TE16r
484
4295
m1
TE4/2f
TE19/2r
578


B
4295
1976
M
m1
m2
TE4/1f - TE16r
500
400
TE4/2f - TE19/2r
600
500
TE3f - upTE12r
500
400
TE8f - TE8r
200
UL13f - UL13r
200
100


C
Titer (pfu/mL)
10^6
10^5
10^4
10^3
10^2
10^1
0
4
8
12
Days post-infection
4295 M
4295 m1
4295 m2
1976


Mock/75 H
100
80
60
40
20
0
Survival (%)
4295/38a D
4295/38b D
4295/41 D
10
20
30
40
50
60
70
Days post-infection


Lung
Spleen
Brain
Kidney
Liver
Mock-infected
Infected
F
H
C
GC
PT
DT
G
S
M

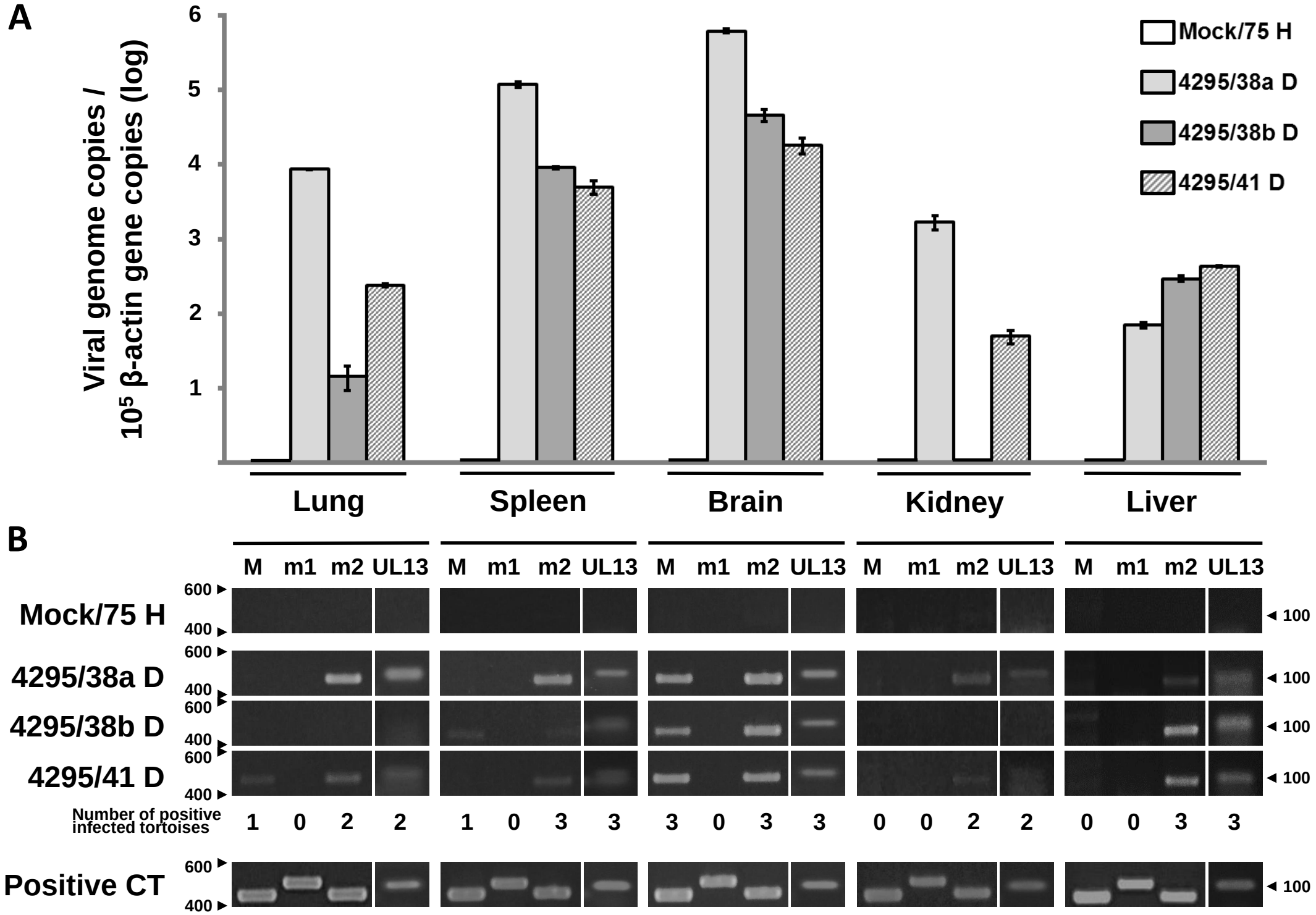

A
Viral genome copies / 10^5 β-actin gene copies (log)
6
5
4
3
2
1
Mock/75 H
4295/38a D
4295/38b D
4295/41 D
Lung
Spleen
Brain
Kidney
Liver
B
M m1 m2 UL13
Mock/75 H
4295/38a D
4295/38b D
4295/41 D
600
400
100
Number of positive infected tortoises
1 0 2 2 1 0 3 3 3 0 3 3 0 0 2 2 0 0 3 3
Positive CT

TABLE 1 Oligonucleotide primers.

Description	Primer name	Sequence (5'-3')	Coordinates[a]
For synthesis of probes for Southern blot analysis (probe name)			
P1	TE2 F	ATACAGTCCGTGGGATCCAG	1303 - 1322
	TE2 R	CACGTGAGGCACATAGGAGA	1516 - 1535
P2	TE11 F	CAGAGGCTGAAGGAAACTGG	13285 - 13304
	TE11 R	TCCTCCCGCTATAGGAAACC	13477 - 13496
P3	TE12 F	AAGCCTGGTGGTACGATGAC	15747 - 15766
	TE12 R	GCAATCTCCGATAAGCTCCA	15960 - 15979
P4	UL1 F	TTTCCCGTACCTCTGTCTCG	120587 - 120606
	UL1 R	ATGAGATGTTGCTGCGACTG	120960 - 120979
P5	TE28 F	ATAGCGGCCGACAATGTAAC	134019 - 134038
	TE28 R	TGGCCCGAAGTATTTTTACG	134176 - 134195
P6	TE42 F	GTCTCGGTCCATGGCTATTC	151441 - 151460
	TE42 R	TCAGGGAGTAGTGGGTGGAG	151666 - 151685
For qPCR analysis (gene amplified)			
TeHV-3 UL13	UL13 F	CGCATCCGTCAGGAATCTAT	103686 - 103705
	UL13 R	GGTCCCTCGTCCAACTAACA	103772 - 103791
***T. hermanni* β-actin gene[b]**	Beta-actin F	TCTGGTCGTACCACAGGTATTG	Not applicable
	Beta-actin R	AGATCCAGACGGAGGATGG	Not applicable
For analyzing specific features in the strain 1976 and strain 4295 genomes (targeted feature)[c]			
Reiteration at 6540-6842	TE5 F	AATACATGATACCAATCCCAGTTG	6473 - 6496
	TE5 R	CCAGAGGGGACACCGCAGATGACA	6902 - 6925
Reiteration at 27611-28227	TE19/1 F	ACTCCTGGCCACAGAACCAGTTGG	27559 - 27582
	TE19/1 R	TAAAAACATACCAGGAGGTTCCCA	28251 - 28274
Reiteration at 31883-32067	ORF13 F	ATCCAATACATATTTCACGG	31818 - 31837
	ORF13 R	CTGGTTCGAAACCTACGTATCGAG	32122 - 32145
Reiteration at 59762-59971	UL36 F	GTTGTTTACCCAGTTCCTGTACCA	59684 - 59707
	UL36 R	CATCTGGAAATAGTGTTGGCTATC	60040 - 60063
Reiteration at 97095-97300	TE22-TE23 F	CAAACGGCCATATCTCTTAG	97032 - 97051
	TE22-TE23 R	TAGAGCTCTGATCAATGTGTATAC	97387 - 97410
Reiteration at 121235-121565	UL1-IRS/1 F	TTTACAAAGCCGGGTGGAGCCTGG	121132 - 121155
	UL1-IRS/1 R	AGTCCCGGCATCGGCCGTGGTGGA	121629 - 121652
Duplication at 121745-121964 (ori)	UL1-IRS/2 F	TACAAGTACCTGATCGGGCT	121682 - 121701
	UL1-IRS/2 R	CCGAGATTTGGTACGGCTAGGACC	122396 - 122419
Reiteration at 126971-127795 and 151910-152308	TE25-RS1 F	ACGATTACGATGAGAGCACTGACA	126811 - 126834
	TE25-RS1 R	AGAACTGACCGGATTGGTGAACGA	127965 - 127988
Reiteration at 130962-131360 and 155475-156299	RS1-US F	GACCAGGCAGGTGCTTCATCCGTA	130866 - 130889
	RS1-US R	CTCCTAGCATTCCCATTGGC	131378 – 131397 [d]
Reiteration at 131398-131576	RS1-TE26 F	GTCATGTCAACCAGCCAATG	131365 - 131384
	RS1-TE26 R	TTCATGAGGGTCACACTGAC	131607 - 131626
Reiteration at 150542-150717	TE42/1 F	TTCTGATATCCTGGGGACAT	150461 - 150480
	TE42/1 R	ATAGTAGTGTTGATTCGTCC	150754 - 150773
Reiteration at 151008-151141	TE42/2 F	AGAGGTCGCTGCTCTCTTAACTGA	150907 - 150930
	TE42/2 R	TAGAGGTAGAGGCTACGCCAGGTC	151370 - 151393
Duplication in RS1 (strain 4295)[d]	RS1 M F	ACGACATCTGCTATCTCAGTGCAC	129099 - 129122
	RS1 M R	TATACCAGGCTTTCGATGCGCTGG	130014 - 130038
Left genome terminus	LEFT TERMINUS R	GCGAATCCGCAGGCAATCGCAACA	205 - 228
Left genome terminus[e]	LEFT TERMINUS F	GCGATCCAAATACGCGTAGCGATT	122704 - 122727
Right genome terminus	RIGHT TERMINUS F	ACCCTCCGAAGAGCAGACCATGAG	160211 – 160234
For genotyping strain 4295 forms (targeted form or gene)			
M	TE4/1 F	GATGGGTATGGAACGTCACC	3790 - 3809
	TE16 R	CGGCCATGGTTAGAAAAAGA	22567 - 22586
m1	TE4/2 F	CAATCATCTGAGCGTTGGAA	4534 - 4553
	TE19/2 R	ATTCGTCCGTCACAGTAGGG	27514 - 27533
m2	TE3 F	AAAGTCCGCTCCTCTCATCA	2281 - 2300
	up TE12 R	GCCGCCTAATAGGTTCTTTG	15221 - 15240
TE8	TE8 F	CGAGCAGCCTAATTCAGACC	10046 - 10065
	TE8 R	AACGCCTTTCTGAACGAAGA	10236 - 10255

[a] Coordinates are listed in relation to the strain 1976 genome.

[b] Amplified fragment (primers omitted): 5'-TCT GGT CGT ACC ACA GGT ATT GTG ATG GAC TCT GGT GAT GGT GTC ACC CAC ACT GTG CCC ATC TAT GAA GGT TAT GCC CTC CCC CAC GCC ATC CTC CGT CTG GAT CT-3'

[c] Principal primers only. Additional primers were used to generate confirmatory PCR products or sequence them. An adapter-specific primer was used in combination with virus-specific primers to locate the genome termini.

[d] Present in strain 4295 only. Coordinates correspond to the strain 1976 genome.

[e] U_L is inverted in a proportion of strain 4295 genomes.

TABLE 2 Features of predicted functional protein-coding regions in the TeHV-3 strain 1976 genome.

Gene name[a]	Gene family	Protein size (residues)	Protein name[b]	Protein features
TE1		207	Protein TE1	
TE2		152	Protein TE2	
TE3	TE3	505	Protein TE3	Contains signal peptide
TE4	TE3	520	Protein TE4	Contains signal peptide
TE5	TE3	774	Protein TE5	Contains signal peptide
TE6		106	Protein TE6	Contains potential transmembrane domain
TE7		653	Semaphorin	Type 1 membrane protein
TE8		168	Interleukin-10	Contains signal peptide
TE9	TE3	473	Membrane protein TE9	Type 1 membrane protein
TE10		210	Protein TE10	
TE11		200	Protein TE11	Contains potential transmembrane domain; similar to C-type lectins
TE1		207	Protein TE1	
TE12	TE3	658	Protein TE12	Contains signal peptide
TE13	TE3	636	Protein TE13	Contains signal peptide
TE14		162	Protein TE14	
TE15	TE15	231	Protein TE15	
UL55A	UL55	218	Nuclear protein UL55A	
TE16	TE15	229	Protein TE16	
UL55B*	UL55	221	Nuclear protein UL55B	
TE17		432	Protein TE17	
TE18		525	Protein TE18	
TE19		876	Membrane protein TE19	Type 1 membrane protein; contains immunoglobulin domains
TE20		262	Membrane protein TE20	Type 1 membrane protein; contains immunoglobulin domain
UL44		350	Envelope glycoprotein C	Type 1 membrane protein; binds cell surface heparan sulphate; binds complement C3b to block neutralization; involved in cell attachment
ORF13		302	Thymidylate synthase	Involved in nucleotide metabolism
UL45		146	Membrane protein UL45	Type 2 membrane protein; tegument-associated; possibly involved in membrane fusion
TE21		100	Protein TE21	Contains potential transmembrane domain
UL46		403	Tegument protein VP11/12	Modulates transactivating tegument protein VP16; possibly involved in gene regulation
UL47		473	Tegument protein VP13/14	Modulates transactivating tegument protein VP16; RNA-binding protein; possibly involved in gene regulation
UL48		428	Transactivating tegument protein VP16	Transactivates immediate early genes; involved in gene regulation; involved in virion morphogenesis
UL49		212	Tegument protein VP22	Involved in virion morphogenesis; possibly involved in RNA transport to uninfected cells
UL49A		100	Envelope glycoprotein N	Type 1 membrane protein; complexed with envelope glycoprotein M; involved in virion morphogenesis; involved in membrane fusion
UL50		447	Deoxyuridine triphosphatase	Involved in nucleotide metabolism
UL51		216	Tegument protein UL51	Involved in virion morphogenesis
UL52*		906	Helicase-primase primase subunit	Involved in DNA replication
UL53*		346	Envelope glycoprotein K	Type 3 membrane protein; 4 transmembrane domains; involved in virion morphogenesis; involved in membrane fusion
ORFG*		201	Protein IG	Contains a possible zinc-binding domain
UL54		953	Multifunctional expression regulator	RNA-binding protein; shuttles between nucleus and cytoplasm; inhibits pre-mRNA splicing; exports virus mRNA from nucleus; exerts most effects post-transcriptionally; involved in gene regulation; involved in RNA metabolism and transport
UL43*		396	Envelope protein UL43	Type 3 membrane protein; possibly involved in membrane fusion
UL42*		368	DNA polymerase processivity subunit	dsDNA-binding protein; involved in DNA replication
UL41*		319	Tegument host shutoff protein	mRNA-specific RNase; involved in cellular mRNA degradation
UL40		304	Ribonucleotide reductase subunit 2	Involved in nucleotide metabolism
UL39		796	Ribonucleotide reductase subunit 1	Involved in nucleotide metabolism
UL38*		447	Capsid triplex subunit 1	Complexed 1:2 with capsid triplex subunit 2 to connect capsid hexons and pentons; involved in capsid morphogenesis
UL37*		1056	Tegument protein UL37	Complexed with large tegument protein; involved in virion morphogenesis
UL36*		2619	Large tegument protein	Complexed with tegument protein UL37; ubiquitin-specific protease (N-terminal region); involved in capsid transport
UL35*		114	Small capsid protein	Located externally on capsid hexons; involved in capsid morphogenesis; possibly involved in capsid transport
UL34*		242	Nuclear egress membrane protein	Type 2 membrane protein; interacts with nuclear egress lamina protein; involved in nuclear egress
UL33*		115	DNA packaging protein UL33	Interacts with DNA packaging terminase subunit 2; involved in DNA encapsidation
UL32*		537	DNA packaging protein UL32	Involved in DNA encapsidation; possibly involved in capsid transport

UL31*		295	Nuclear egress lamina protein	Interacts with nuclear egress membrane protein; involved in nuclear egress
UL30*		1132	DNA polymerase catalytic subunit	Involved in DNA replication
UL29*		1183	Single-stranded DNA-binding protein	Contains a zinc-finger; involved in DNA replication; possibly involved in gene regulation
UL28*		730	DNA packaging terminase subunit 2	Involved in DNA encapsidation
UL27*		827	Envelope glycoprotein B	Type 1 membrane protein; possible membrane fusogen; binds cell surface heparan sulphate; involved in cell entry; involved in cell-to-cell spread
UL26.5*		233	Capsid scaffold protein	Clipped near C terminus; involved in capsid morphogenesis
UL26*		490	Capsid maturation protease	Serine protease (N-terminal region); minor scaffold protein (remainder of protein, clipped near C terminus); involved in capsid morphogenesis
UL25*		575	DNA packaging tegument protein UL25	Located on capsid near vertices; possibly stabilizes the capsid and retains the genome; involved in DNA encapsidation
UL24*		324	Nuclear protein UL24	
UL23*		355	Thymidine kinase	Involved in nucleotide metabolism
UL22*		733	Envelope glycoprotein H	Type 1 membrane protein; possible membrane fusogen; complexed with envelope glycoprotein L; involved in cell entry; involved in cell-to-cell spread
UL21*		450	Tegument protein UL21	Interacts with microtubules; involved in virion morphogenesis
UL20*		196	Envelope protein UL20	Type 3 membrane protein; 4 transmembrane domains; involved in virion morphogenesis; involved in membrane fusion
UL19*		1369	Major capsid protein	6 copies form hexons, 5 copies form pentons; involved in capsid morphogenesis
UL18*		320	Capsid triplex subunit 2	Complexed 2:1 with capsid triplex subunit 1 to connect capsid hexons and pentons; involved in capsid morphogenesis
TE22	TE22	153	Protein TE22	
TE23	TE22	150	Protein TE23	
UL15*		697	DNA packaging terminase subunit 1	Contains an ATPase domain; involved in DNA encapsidation
UL17*		652	DNA packaging tegument protein UL17	Capsid-associated; involved in DNA encapsidation; involved in capsid transport
UL16*		335	Tegument protein UL16	Possibly involved in virion morphogenesis
UL14*		177	Tegument protein UL14	Involved in virion morphogenesis
UL13		437	Tegument serine/threonine protein kinase	Involved in protein phosphorylation
UL12*		533	Deoxyribonuclease	Involved in DNA processing
UL11*		79	Myristylated tegument protein	Envelope-associated; involved in virion morphogenesis
UL10*		412	Envelope glycoprotein M	Type 3 membrane protein; 8 transmembrane domains; complexed with envelope glycoprotein N; involved in virion morphogenesis; involved in membrane fusion
UL9*		784	DNA replication origin-binding helicase	Involved in DNA replication
UL8*		716	Helicase-primase subunit	Involved in DNA replication
UL7*		271	Tegument protein UL7	Involved in virion morphogenesis
UL6*		669	Capsid portal protein	Dodecamer located at one capsid vertex in place of a penton; involved in DNA encapsidation
UL5*		851	Helicase-primase helicase subunit	Involved in DNA replication
UL4*		227	Nuclear protein UL4	Colocalizes with regulatory protein ICP22 and nuclear protein UL3 in small, dense nuclear bodies
TE24		131	Protein TE24	
UL3*		193	Nuclear protein UL3	Colocalizes with regulatory protein ICP22 and nuclear protein UL4 in small, dense nuclear bodies
UL2*		270	Uracil-DNA glycosylase	Involved in DNA repair
UL1*		127	Envelope glycoprotein L	Contains signal peptide; complexed with envelope glycoprotein H; involved in cell entry; involved in cell-to-cell spread
TE25	TE3	953	Protein TE25	Contains signal peptide
RS1*		965	Transcriptional regulator ICP4	Involved in gene regulation
TE26		210	Protein TE26	
TE27	TE27	241	Protein TE27	Contains potential transmembrane domain
TE28		246	Protein TE28	Contains potential transmembrane domain
TE29	TE3	494	Protein TE29	Contains signal peptide
TE30	TE3	465	Protein TE30	Contains signal peptide
US3A*	US3	338	Serine/threonine protein kinase US3A	
US3B*	US3	312	Serine/threonine protein kinase US3B	
TE31*		253	Membrane protein TE31	Type 1 membrane protein
TE32		222	Membrane protein TE32	Type 1 membrane protein; contains Ig domain
TE33		448	Protein TE33	Contains signal peptide
TE34		147	Protein TE34	
US8*		465	Envelope glycoprotein E	Type 1 membrane protein; involved in cell-to-cell spread
TE35		92	Protein TE35	Contains signal peptide
TE36		69	Protein TE36	Contains potential transmembrane domain
US10*		248	Virion protein US10	

TE37		481	Protein TE37	Contains signal peptide
TE38		120	Protein TE38	
TE39		94	Protein TE39	Contains potential transmembrane domains
TE40		103	Protein TE40	
TE41	TE27	233	Protein TE41	
TE42		462	Protein TE42	
RS1*		965	Transcriptional regulator ICP4	Involved in gene regulation
TE25	TE3	953	Protein TE25	Contains signal peptide

[a] Genes are listed as they are ordered in the genome. Duplicates (TE1, RS1, and TE25) are included. Asterisks mark genes that have orthologs in ChHV-5. Information on conservation in other herpesviruses is available in Fig. 2.

[b] The names of conserved proteins, and functional annotations, are derived from NCBI reference sequences (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10292).