View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repository of the Academy's Library

1 # Novel parvoviruses in reptiles and genome sequence of a
2 lizard parvovirus shed light on *Dependoparvovirus* genus
3 evolution

4

5 **Running title:** Novel parvoviruses in reptiles

6

7 **Authors:** Judit J. Pénzes[1*], Hanh T. Pham[2], Mária Benkő[1], Peter Tijssen[2]

8

9 **Addresses**

10 [1]Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian
11 Academy of Sciences, 21 Hungária krt., Budapest, Hungary, H-1143

12 [2]INRS-Institut Armand-Frappier, Université du Québec, 531 Boulevard des Prairies, Laval,
13 QC H7V 1B7, Canada

14

15 ***Corresponding author**

16 E-mail: penzes.judit@agrar.mta.hu

17 Phone: +36 1 467-4084

18

19

20

21

22 Contents Category: Animal – Small DNA viruses

23 Key words: reptile, lizard, *Dependoparvovirus,* AAV: autonomous replication, evolution

24 Accession numbers:

25 Bearded dragon parvovirus – KP733794

26 Pygmy chameleon parvovirus – KP733796

27 Corn snake parvovirus – KP733795

28

29 **Abstract**

30 Here, we report the detection and partial genome characterization of two novel
31 reptilian parvoviruses derived from a short-tailed pygmy chameleon (*Rampholeon*
32 *brevicaudatus*) and a corn snake (*Pantherophis guttatus*) along with the complete genome
33 analysis of the first lizard parvovirus, obtained from four bearded dragons (*Pogona vitticeps*).
34 Both homology searches and phylogenetic tree reconstructions demonstrated that all are
35 members of the *Dependoparvovirus* genus. Even though most dependoparvoviruses replicate
36 efficiently only in co-infections with large DNA viruses, no such agents could be detected in
37 one of the bearded dragon samples, hence the possibility of autonomous replication was
38 explored. The alternative ORF encoding the full assembly-activating protein (AAP), typical
39 for the genus, could be obtained from reptilian parvoviruses for the first time, with a structure
40 that appears to be more ancient than that of avian and mammalian parvoviruses. All three
41 viruses were found to harbor short introns as previously observed for snake adeno-associated
42 virus (SAAV), shorter than that of any non-reptilian dependoparvovirus. According to the
43 phylogenetic calculations based on full non-structural protein (Rep) and AAP sequences, the
44 monophyletic cluster of reptilian parvoviruses seems to be the most basal out of all lineages of
45 genus *Dependoparvovirus*. The suspected ability for autonomous replication, results of
46 phylogenetic tree reconstruction, intron lengths and the structure of the AAP, suggested that a
47 single Squamata origin instead of the earlier assumed diapsid (common avian-reptilian) origin
48 is more likely for the genus *Dependoparvovirus* of the *Parvoviridae* family.

49

50

51 **Introduction**

52 Members of the *Parvoviridae* family are non-enveloped viruses of icosahedral symmetry with
53 a diameter of ~25 nm. Their linear, single-stranded DNA genome (of 4–6.3 kb) has a well-
54 conserved organization of two major ORFs (*rep* and *cap*) encoding the replication or non-
55 structural (Rep) and the capsid (VP1, VP2, VP3) proteins, respectively. The genome is
56 flanked by palindromic sequences that form a hairpin-like, partially double-stranded
57 secondary structure, essential for replication (Tijssen *et al.*, 2011). For dependoparvoviruses,
58 as well as many other parvoviruses, these telomeres form inverted terminal repeats (ITRs).

59 Parvoviruses (PVs) occur in numerous vertebrate and invertebrate hosts. Accordingly,
60 the family is divided into two subfamilies (*Parvovirinae* and *Densovirinae*), of which the
61 former infects vertebrates (Cotmore *et al.*, 2014; Tijssen *et al.*, 2011). Dependoparvoviruses
62 or so-called adeno-associated viruses (AAVs) are classified in the *Dependoparvovirus* genus,
63 which is known for the widest host spectrum out of the current eight genera of the
64 *Parvovirinae* subfamily. Although members of the genus *Aveparvovirus* infect galliform birds
65 (Zsák *et al.*, 2008), all members of the remaining six genera are restricted exclusively to
66 mammals (Cotmore *et al.*, 2014). Members of the genus *Dependoparvovirus* infect
67 representatives of all major amniotic groups, i.e. reptiles, birds, and mammals. However,
68 signs of PV infections in reptiles are rather scarce. To date only two successful isolations
69 were reported; one from a corn snake (*Pantherophis guttatus*) (Ahne & Scheinert, 1989), and
70 the other from a ball python (*Python regius*) (Farkas *et al.*, 2004; Ogawa *et al.*, 1992).
71 Furthermore, PV-like particles were observed in bearded dragons (*Pogona vitticeps*)
72 (Jacobson *et al.*, 1996) and in California mountain kingsnakes (*Lampropeltis zonata*
73 *multicincta*) (Wozniak *et al.*, 2000). As for molecular characterization, there is only one fully-
74 sequenced reptilian PV genome derived from the ball python isolate, named snake adeno-
75 associated virus (SAAV), representing the newly established species, *Squamate*

76 *dependoparvovirus 1* (Farkas *et al.*, 2004). Since then, the partial genome analysis of
77 serpentine adeno-associated virus 2, detected in an Indonesian pit viper (*Parias hageni*), has
78 been published (Farkas & Gál, 2008). Recently, the first, partial molecular data on a non-
79 serpentine reptilian PV, from a checkerboard worm lizard (*Trogonophis wiegmanni*), an
80 amphisbaenian, were reported (Pénzes & Benkő, 2014).

81 The name, *Dependoparvovirus*, reflects a common feature of its members, i.e. their
82 dependence on some helper viruses (usually adeno- or herpesviruses) for efficient replication.
83 However, the viruses causing the so called Derzsy's disease in geese and Muscovy ducks are
84 capable of autonomous replication even though they are members of the *Dependoparvovirus*
85 genus (Brown *et al.*, 1995; Le Gall-Recule & Jestin, 1994; Zádori *et al.*, 1995). Based on the
86 ability of autonomous replication of anseriform PVs and the basal phylogenetic position of the
87 SAAV, a diapsid (common reptile-bird) origin of the genus has been proposed (Farkas *et al.*,
88 2004; Zádori *et al.*, 1995). Nonetheless, PVs in reptiles have been found with concomitant
89 adeno- or herpesvirus infection in all cases reported previously (Ahne & Scheinert, 1989;
90 Farkas & Gál, 2008; Heldstab & Bestetti, 1984; Jacobson *et al.*, 1996; Kim *et al.*, 2002;
91 Wozniak *et al.*, 2000). The only exception to date is the amphisbaenian PV, with which no
92 simultaneous infection by large DNA viruses could be observed (Pénzes & Benkő, 2014).

93 In this study, our aim was to assess the prevalence and diversity besides extending the
94 number of the currently known, scarce reptilian PVs. Our goal was furthermore to analyze the
95 complete genome organization of more reptilian PVs, including lizard ones. Moreover, we
96 expected these results to shed more light on the evolution of the *Dependoparvovirus* genus.

97

98

99 **Results**

100 **PCR screening**

101 Samples of captivity-kept, deceased reptiles were screened by a consensus PCR method
102 (Pénzes & Benkő, 2014) throughout the years from 2009 to 2013. Six positive results were
103 obtained out of the 162 independent samples. This would imply an average infection rate of
104 3.7%. Although samples from a large taxonomic scale were screened, all the positive cases
105 were obtained exclusively from members of the Squamata order. A novel, previously
106 unrecognized PV was detected by PCR in four out of nine bearded dragons (*Pogona vitticeps*)
107 and was designated bearded dragon parvovirus (BDPV). Three out of the four samples turned
108 out to be positive also for adenoviruses (AdV) during the parallel examinations (Pénzes &
109 Doszpoly, 2011). The adenovirus-negative sample originated from an adult female bearded
110 dragon, with malfunctioning ovaries and aberrant yolk formation. The sample proved to be
111 negative for any large DNA viruses that encode a DNA-dependent DNA polymerase gene in
112 their genome as well. The other positive individuals were only 2-3 months old, displaying
113 neurological signs as metabolic disorders such as bone malformation, dysecdysis and
114 anorexia. We detected another novel lizard PV in one out of 23 adenovirus-infected short-
115 tailed pygmy chameleons (*Rampholeon brevicaudatus*) and referred to it as pygmy chameleon
116 parvovirus (PCPV). As for snakes, in co-infection with snake AdV-1, a novel snake PV was
117 obtained from one out of four corn snake samples (*Pantherophis guttatus*) and was referred to
118 as corn snake parvovirus (CSPV). All novel reptilian PVs displayed the highest similarity
119 with homologous *cap* fragments of members of the *Dependoparvovirus* genus according to
120 BLAST homology searches.

121

122 **Complete and partial genome characterization of novel reptilian parvoviruses**

123 The amplification of the short fragment from *rep* was successful in case of all three novel
124 PVs; hence a longer fragment encompassed by the short PCR fragments of the *cap* and the
125 *rep* could be amplified. Its length corresponded to 1487 nt in PCPV (GenBank accession:
126 KP733796), and 1821 nt in CSPV (GenBank accession: KP733795). As for BDPV, the
127 complete genome sequence could be obtained, including the ITRs (GenBank accession:
128 KP733794). The comparison of the homologous, approx. 1500-nt-long fragment in all three
129 viruses is presented in Fig. 1. In all cases, the stop codon of the Rep and the start codons of all
130 VP proteins occur adjacent to a putative promoter that is homologous with the adeno-
131 associated virus 2 (AAV2) P40, being the most downstream promoter out of the three that are
132 typical for the *Dependoparvovirus* genus (Kotin & Smith, 2001; Tijssen *et al.*, 2011). The
133 length of the non-coding region, situated between the two major ORFs, is 17 nt in both PCPV
134 and BDPV, and 18 in CSPV. By analogy with AAV2 (Qiu *et al.*, 2006), we could predict two
135 introns in all genomes, which are spliced from a presumed common donor site and from two,
136 distinct acceptor sites. In all cases, the complete sequence of an alternative ORF, of the
137 putative assembly-activating protein (AAP) (Naumer *et al.*, 2012; Sonntag *et al.*, 2011) could
138 be identified. It harbors an alternative start codon (CTG), similarly to primate AAVs. The
139 comparison of the novel reptilian AAP aa sequences with that of other dependoparvoviruses
140 revealed the essential core region to be the most conserved. The hydrophobic N terminal
141 region, the other essential motif for capsid assembly (Naumer *et al.*, 2012), displayed
142 remarkable variety both in length and in the number of hydrophobic aa clusters throughout the
143 whole genus. The proline-rich region is basically absent in the squamate AAPs. Only three
144 threonine/serine (T/S)-rich regions could be identified. The alignment is presented in Fig. 2.

145 The sequence between nt 2230-2370 is highly conserved in all three viruses, since it
146 contains a phospholipase A2 motif (PLA2) (Zádori *et al.*, 2001).

147 The complete genome of the BDPV consisted of 4590 nt with ITRs of 257 nt each.
148 The genome organization is presented in Fig 3(A). Within the ITRs, a regular, T-shaped
149 hairpin structure could be predicted. Such secondary structure within the ITRs is typical of
150 members of genus *Dependoparvovirus*, except goose PV (GPV) and duck PV (DPV) (Zádori
151 *et al.*, 1995). The predicted hairpin consisted of 130 nt as shown in Fig. 3(B). The genome
152 core flanked by the ITRs consisted of 4076 nt only, being the shortest out of all
153 dependoparvoviruses. The length of the full genome, including the ITRs, qualified as the
154 second shortest after that of SAAV. The genome contained two major ORFs corresponding to
155 the *rep* and *cap* genes of all PVs known to date (Qiu *et al.*, 2006), with lengths of 1518 and
156 2178 nt, respectively. The splice donor- and acceptor sites of the Rep-protein-coding genes of
157 the other two reptilian PVs had similar positions (Fig. 1). Three putative promoters,
158 corresponding to the P5, P19 and P40 of AAV2, were identified in the genome at positions of
159 nts 323, 803, and 1852 (TATA boxes are shown in Fig. 1 and Fig. 3(A)). We identified two
160 predicted poly(A) sites in the BDPV genome, out of which the one with a higher score
161 follows the *cap* gene at position 4278. Another, supposedly cryptic poly(A) site was,
162 however, observed within the intron in the middle of the genome analogous to AAV5 at the
163 position of 2076 (Qiu *et al.*, 2006).
164 As for the proteins, the BDPV genome is predicted to encode at least two Rep
165 proteins. The presence of the two presumed introns and the predicted inner start codon,
166 following the putative promoter homologous to P19, however, suggests that there might be
167 four Reps expressed, homologous to those of AAV2, respectively. In the protein sequence of
168 the Rep1, supposedly homologous to the AAV2 Rep78, the putative replication initiative
169 motif I and II (Ilyna & Koonin, 1992) as well as the putative tripartite helicase superfamily III
170 motifs could be identified (Tijssen & Bergoin, 1995; Smith et al., 1999).The PLA2 motif
171 (Zádori *et al.*, 2001) was present in the N-terminal region of the putative VP1 protein

172 sequence in all three viruses. The putative AAP binding site was identified close to the C-
173 terminal of the VP proteins (Naumer *et al.*, 2012). It surrounds M689 at the aa sequence of
174 VEMLWEV. The putative Rep protein sequence shared most identity with its homologue in
175 SAAV (58%) and not less than 34% with all other members of the *Dependoparvovirus* genus.
176 The VP protein sequence also displayed the highest identity with that of SAAV (70%) while
177 this value was not lower than 56% in case of other dependoparvoviruses.
178
179 **Phylogenetic analysis**
180 Phylogenetic tree reconstructions were performed in case of all three proteins. As for the VP
181 protein, only the partial, 226-aa-long fragments obtained during the PCR screening were used.
182 The phylogenetic trees according to the Rep and AAP proteins were based on the entire
183 deduced aa sequence. All three novel reptilian PVs could be included in the
184 *Dependoparvovirus* genus according to the short fragment of the VP protein, presented in Fig.
185 4(A). The complete Rep protein sequence of the BDPV, clustered with SAAV Rep protein,
186 forming the most basal group within dependoparvoviruses as shown in Fig. 4(B). The AAP-
187 based tree presented in Fig. 4(C) indicated that all reptilian PVs formed a monophyletic
188 cluster, yet this did not apply for all diapsid PVs.
189
190 **Discussion**
191 **Parvovirus infection in lizards implies the ability of autonomous replication of reptilian**
192 **dependoparvoviruses**
193 A complete clone and sequence were obtained of the bearded dragon lizard PV and an
194 incomplete sequence of the short-tailed pygmy chameleon PV, besides detecting another
195 novel snake PV that is only the third one to date. Our primers (Pénzes & Benkő, 2014) proved
196 to be suitable for detecting reptilian PVs. The infection rate of 3.7% revealed during the

197 screening was low, especially if compared to that of mammalian dependoparvoviruses.
198 According to the few studies carried out so far it has been reported to be 19.9% in primates
199 (Gao *et al.*, 2003) and 22.4% in bats (Li *et al.*, 2010). It is not certain whether the low
200 infection rate of reptilian PVs mirrors real values, or if our PCR system failed to amplify viral
201 DNA in some cases. However, the low infection rate of pygmy chameleons supports the low
202 infection rate in reptiles in general; only one sample was found to be positive among a total of
203 26 samples, including 23 AdV-positive ones. Nevertheless, this is the first evidence for PV
204 infection of any member of the Chamaeleonidae family. The infection rate among bearded
205 dragons was much higher (~ 44%). Although PV-like particles in bearded dragons have been
206 reported before, these are the first molecular data corroborating their relationship to
207 *Parvoviridae*. The same applies to the CSPV, the hitherto only third snake PV from which
208 genomic sequence data became available. It is interesting that both snake PVs, SAAV and
209 CSPV, were derived from SnAdV-1 positive cases (Farkas *et al.*, 2004; Ogawa *et al.*, 1992).
210 The fact that BDPV could be found in a sample of an animal with no simultaneous
211 infection by any potential helper virus is in concordance with previous findings. No potential
212 helper virus could be demonstrated in the worm lizard containing the amphisbaenian PV
213 either (Pénzes & Benkő, 2014). These results suggested that reptilian dependoparvoviruses,
214 just like anseriform dependoparvoviruses, might also be capable of autonomous replication.
215 However, there is no experimental evidence to support this assumption, hence it remains only
216 a plausible hypothesis for now.
217
218 **Complete genome characterization of the first lizard parvovirus and partial**
219 **characterization of other reptilian parvoviruses**
220 The genome of the BDPV, reported here, is the first complete lizard PV genome, and only the
221 second complete reptilian PV besides SAAV (Farkas *et al.*, 2004). Despite their short

222 genomes, their ITRs are the second longest within the *Dependoparvovirus* genus. The
223 terminal 122 nt of the 154-nt-long ITRs of SAAV correspond to the palindrome hairpin
224 structure (Farkas *et al.*, 2004). The ITRs of BDPV were 103 nt longer, even though the
225 number of nt involved in the hairpin was rather similar, i.e. 130 nt. The suspected Rep-
226 binding site precedes the terminal resolution site by 18 nt (17 in SAAV). We are currently not
227 aware of the function(s) of the extremely long single-stranded region of BDPV ITRs.
228 Nevertheless, recent studies have suggested that particular sequences of AAV ITRs play a
229 crucial role in increasing site-specific integration into host cell genomes (Galli & Cervelli,
230 2014). A putative similar role of the long BDPV ITRs may be possible as well.

231 The organization of the BDPV genome is typical for members of the
232 *Dependoparvovirus* genus, including the three promoters, the length of the non-coding region
233 flanked by the two ORFs (15 to18 bases) (Li *et al.*, 2010), and the presence of alternative
234 splicing involving one donor and two acceptor sites (Cotmore *et al.*, 2014; Qiu & Pintel,
235 2008; Qiu *et al.*, 2006; Tijssen *et al.*, 2011). The length of the first intron in mammalian
236 dependoparvoviruses is around 300 nt (Chiorini *et al.*, 1999; Ruffing *et al.*, 1994) while 205
237 to 215 nt in those of avian origin (Estevez & Villegas, 2004; Zádori *et al.*, 1995) and only 160
238 in SAAV (Farkas *et al.*, 2004), similarly to the three novel reptilian dependoparvoviruses
239 (Table 1).

240 The polyadenylation strategy of dependoparvoviruses varies in different species.
241 Although we identified a putative inner poly(A) site within the intron of these novel PVs, it
242 remains unknown if polyadenylation takes place at this position or exclusively at the predicted
243 principal, downstream poly(A) site.

244 Alternative ORFs have been identified multiple times in genomes of PVs with
245 different evolutional backgrounds (Allander *et al.*, 2001; Day & Zsak, 2010; Tse *et al.*, 2011;
246 Zádori *et al.*, 2005). In the *Dependoparvovirus* genus, the recently discovered ORF of the

247 AAP is completely contained within the *cap* gene (Sonntag *et al.*, 2011). This applies to the
248 three novel reptilian PVs as well (Fig. 1). A scaffolding role is ascribed to this protein and is
249 essential for capsid assembly. The various regions in the protein sequence corresponding to
250 this role has been well-characterized (Naumer *et al.*, 2012). However, the function of the
251 proline-rich and T/S-rich regions is unknown, even though they are rather conserved and
252 occur repetitively (Naumer *et al.*, 2012). In case of squamate hosts, the proline-rich region is
253 basically absent, reduced to a single PE motif, while it is present in the AAAV and is slightly
254 reduced in anseriform PVs. The number of the T/S-rich regions also varies among viruses of
255 different host origins; there are five in mammalian AAVs, four in avian PVs, whereas only
256 three in those of squamate origin.

257 Even before the discovery of AAP, the C-terminal 29 aa of the VP protein had been
258 demonstrated to be essential for capsid assembly (Ruffing *et al.*, 1994; Wu *et al.*, 2000), and it
259 is now known to correspond to the AAP binding site (Naumer *et al.*, 2012). The seven-aa-
260 long motif, in proximity to the C terminus of the BDPV VPs, is identical with those of
261 anseriform PVs and SAAV.

262

263 **Reptilian parvoviruses cluster as a monophyletic group within genus *Dependoparvovirus***

264 Phylogeny reconstructions based on the short, 226-aa-long fragment of the VP protein
265 proved to be suitable for PV-classification at genus level, as all genera clustered as
266 monophyletic groups on the midpoint-rooted phylogenetic tree (Fig. 4(A)). This also
267 corresponded to the previous prediction (Pénzes & Benkő, 2014). Furthermore, the three
268 novel reptilian PVs unite in one monophyletic group with amphisbaenian PV and SAAV,
269 being the most descended within genus *Dependoparvovirus*.

270 The unrooted phylogenetic tree, based on the whole Rep sequence and presented with
271 a midpoint-root, contradicted this; the clade formed by reptilian PVs (i.e. BDPV and SAAV)

272 is a basal cluster of the same genus as shown in Fig. 4(B). Currently this contradiction is
273 difficult to resolve, however, the less significant role of the reptilian adaptive immune system
274 should also be considered in this issue (Zimmerman *et al.*, 2010). This might suggest the
275 selection pressure to be somewhat lower on antigenic viral proteins compared to that of
276 mammalian ones. As this might lead to the more flexible evolution of reptilian parvoviral
277 capsid proteins, it could also manifest in homoplasy of the mammalian ones. The *rep* gene,
278 which encodes a non-structural protein is probably exempt from this selective pressure and
279 therefore presumably is more suitable for examining evolutional relationships below the
280 genus level, when the examination is based on complete aa sequences.
281 According to AAP sequences, reptilian PVs display monophyly with anseriform
282 dependoparvoviruses, yet avian AAV (AAAV) clusters with mammalian AAVs. It is worth
283 mentioning that on a Rep-based tree, published earlier in an in silico analysis on endogenous
284 viral elements (Katzourakis & Gifford, 2010), AAAV also appeared split from the anseriform
285 dependoparvoviruses by an integrated sequence found in a mammalian (dolphin) genome.
286 As the protein is specific for the *Dependoparvovirus* genus, the phylogenetic calculations do
287 not provide any information on which clade is more basal. Phylogenetic calculations based on
288 the AAP aa sequence, however, raise further concerns, namely its complete overlap with the
289 *cap*, coding capsid proteins responsible for antigenic traits. When resolving this contradiction,
290 it is important to point out that the *cap* is more conserved within the *Dependoparvovirus*
291 genus (Cotmore et al., 2014). In case of hepadnaviruses, a similar conservation has been
292 detected within overlapping ORFs, explaining their suitability for phylogeny reconstructions
293 (Mizokami *et al*., 1997). In this case, it is likely that the AAP sequence is under a stronger
294 functional selection pressure than capsid proteins are, hence it might serve as a better subject
295 for eluding phylogenetic relationships than capsid proteins in general.
296

297 **Conclusions**

298 Here we report the first examinations on the prevalence and diversity of dependoparvoviruses
299 in captive reptiles. Despite a low prevalence, as the result of this study, the number of known
300 reptilian PVs was doubled. BDPV is the first lizard, and the second reptilian PV for which the
301 complete genome sequence has been obtained. As the complete protein sequence of both
302 major ORFs is known, the taxonomic classification of this virus could be determined. BDPV
303 fulfills the criteria to be designated as a new species within genus *Dependoparvovirus*
304 (Cotmore *et al.*, 2014); hence the name *Squamate dependoparvovirus 2* is suggested.

305 It has been shown that ancestral members of the *Dependoparvovirus* genus were
306 integrated into their vertebrate host's genome over the past 50 million years of evolution
307 (Belyi *et al.*, 2010). It has been observed that certain artifactual circumstances enable AAV2
308 to replicate autonomously, similarly to autonomous PVs of other genera (Yacobson et al.,
309 1987). This suggests the theory that dependoparvoviruses may have descended from
310 autonomous ancestors. As the only members with autonomous replication capability were
311 derived from anseriform birds, i.e. GPV and DPV, a diapsid origin of the *Dependoparvovirus*
312 genus was presumed (Zádori *et al.*, 1995). The basal position of the SAAV on Rep-based
313 phylogenetic trees was in concordance with this assumption (Farkas *et al.*, 2004), even though
314 no data were available on its mode of replication. The fact that both amphisbaenian PV
315 (Pénzes & Benkő, 2014) and BDPV were detected without the presence of a potential helper
316 virus, might provide further support to this theory. Nevertheless, helper-independent
317 replication of these viruses is required to be demonstrated in virus free cell cultures first, in
318 order to draw strong conclusions.

319 As early as 2004, it was observed that the length of the introns in the SAAV genome
320 was shorter than in both avian and mammalian dependoparvoviruses (Farkas *et al.*, 2004).
321 The three novel reptilian PVs also possess short introns (Table 1). If we suppose the intron

322 length to expand continuously throughout the evolution of dependoparvoviruses, then a
323 Squamata origin of the genus turns out to be more likely. This is further supported by the
324 phylogeny based on the Rep sequence. A similar expanding tendency in case of the AAP
325 conserved motifs is observed, i.e. there are five in total T/S-rich regions in mammalian AAVs,
326 only four in the avian and three in reptilian dependoparvoviruses. There is, however, the
327 exception of the AAAV AAP which has the same organization as the AAPs of mammalian
328 viruses. Even the phylogeny reconstruction does not display any monophyly of this avian PV
329 with anseriform PVs, but with mammalian AAVs instead. This clustering cannot be observed,
330 nevertheless, with the Rep-based phylogenetic tree, yet the monophyly of the most basal clade
331 of squamate PVs is still valid. Earlier results on dependoparvovirus Rep-based phylogeny,
332 however, indicated AAAV to be split from the anseriform dependoparvoviruses by an
333 endogenous viral element originating from a cetacean (Katzourakis &Gifford, 2010).
334 In summary, our current results suggest that evolution of dependoparvoviruses support
335 rather an exclusive Squamata origin than a common avian-reptilian one. Further research on
336 novel non-squamate reptilian dependoparvoviruses, provided if such viruses exist, would be
337 required to elucidate this enigma.
338

339 **Methods**

340 **Samples and DNA-extraction**

341 162 independent samples originating from various hosts including members of all reptilian
342 orders except for *Rhyncochephalia* were screened. The most represented order was Squamata,
343 from which 110 samples were examined. The samples were collected from deceased
344 individuals obtained from local pet stores or private pet owners.
345 For nucleic acid extraction, small (15−25 mg) pieces from the internal organs (lungs,
346 liver, intestines, gonads and kidney) were transferred to 2-ml microcentrifuge safe-lock tubes

347 and homogenized by using a TissueLyser LT bead mill (Qiagen®, Hilden, Germany). DNA
348 was purified using the DNeasy Blood and Tissue Kit (Qiagen®, Hilden, Germany) according
349 to the protocol recommended for animal tissues with an overnight incubation at 55°C.
350

351 **PCR primers and conditions**

352 To check the presence of parvoviral DNA, a consensus primer pair aiming at a conserved
353 region of the *cap* gene was used (forward: GGYGCCGAKGGAGTGGGYAATKCCTC,
354 reverse: TCAAARTTRTTBCCBGTYCTYAGCAT) (Pénzes & Benkő, 2014). As for the
355 PCR program an initial denaturation step at 92°C for 5 minutes was followed by 45 cycles of
356 denaturation at 92°C for 30 seconds, annealing at 46°C for 60 seconds and elongation at 72°C
357 for another 60 seconds. Final elongation was performed at 72°C for 5 minutes, expected to
358 result an approximately 600 base pair(bp)-long fragment amplified from the VP gene. For
359 AdV detection, a very sensitive consensus nested PCR that targets a highly conserved region
360 of the adenoviral DNA-dependent DNA polymerase gene (Wellehan *et al.*, 2004) was applied.
361 Amplification of the herpesviral DNA polymerase gene fragment was carried out by another
362 nested consensus PCR (VanDevanter *et al.*, 1996). Furthermore an expanded-range PCR
363 designed for the DNA polymerase gene of other large DNA viruses was performed (Hanson
364 *et al.*, 2006).
365 In case of PV positivity, a short fragment from the *rep* gene was targeted for amplification. To
366 this end two different sets of consensus primer pairs were designed, with the sequences of 5'-
367 GTDAAYTGGACYAAYGMRAAC-3' and 5'-AACATNCKBTCYTSYARNGG-3' in case
368 of set A and 5'-TGTGTCARGTMTWTGATGGKAA-3' and 5'-
369 CAATTCAGGRTAACATTCNRWACA-3' for set B. The size of the expected PCR products
370 was approx. 250 or 400 bp, respectively. A PCR program consisting of an initial denaturation
371 at 94°C for 5 min, 45 cycles of 94°C for 1 min, 41°C for 1 min, and 72°C for 1 min was

372 utilized while finishing with a final synthesis step at 72°C for 3 min. In order to sequence the
373 full genome, primers facing outwards from the already known sequences were designed, and
374 then submitted to PCR with a reaction mix including only these primers. The PCR products
375 were cloned and sequenced. PCR reactions were in 50 µl volume with AmpliTaq Gold® DNA
376 Polymerase (Life Technologies Corporation®, Carlsbad, CA, USA) or, in case of fragments
377 larger than 1000 bp, with Phusion® High-Fidelity DNA Polymerase (ThermoFisher
378 Scientific®, Waltham, MA, USA) according to the manufacturer's recommendations.
379
380 **Obtaining the genome end structures**
381 Because of their secondary structure, the ITRs could not be obtained via the single-primer
382 PCR. To solve this problem, a phosphorylated adaptor 5'-
383 ATCCACAACAACTCTCCTCCTC-3' was attached using T4 RNA ligase (New England
384 Biolabs®, Ipswich, MA, USA) to both genome ends. Additional primers were designed
385 specifically to the nearest ClaI restriction sites at both genome ends. A special PCR was
386 applied in 25 µl with OneTaq® Hot Start DNA Polymerase (New England Biolabs®, Ipswich,
387 MA, USA) including GC enhancer and 3 µl of 2 mM EDTA, and using the adaptor reverse
388 primer. The PCR fragments obtained were cloned and sequenced. In case of incomplete ITRs,
389 specific primers were designed for both the flip and flop structures in order to acquire the
390 remaining nucleotides.
391
392 **Sequencing, molecular cloning and sequence analysis**
393 Sequencing reactions were performed with the BigDye® Terminator v3.1 Cycle Sequencing
394 Kit (Life Technologies Corporation®, Carlsbad, CA, USA), and sent for analysis by a
395 commercial service on an ABI PRISM 3100 Genetic Analyzer (Life Technologies
396 Corporation®, Carlsbad, CA, USA). PCR fragments that were amplified with only one primer

397 were molecularly cloned with the CloneJET® PCR Cloning Kit™ (Thermo Scientific®
398 Waltham, MA, USA), and sequenced with primers specific for the plasmid. Genome end
399 fragments were cloned to pGEM-T easy vectors (Promega®, Fitchburg, WI, USA), into SURE
400 2 electrocompetent cells. For identification and comparison of the nt sequences, the Blastx
401 tool was used at the NCBI website. Sequence editing and assembly was performed with the
402 Staden Sequence Analysis Package (Staden *et al.*, 2000) with occasional manual corrections.

403

404 **Phylogeny reconstructions**

405 Multiple alignments were constructed based on aa sequences of the complete Rep and AAP as
406 well as on the 224-aa-long fragment of the VP protein, using ClustalX v2.1 (Larkin *et al.*,
407 2007). The alignment was then submitted to model selection carried out by ProtTest v2.4
408 (Abascal *et al.*, 2005). Guide tree was calculated via PHYLIP v3.6 using the Protdist
409 application to obtain the distance matrix (JTT substitution model) out of which the tree was
410 constructed by Fitch (Fitch-Margoliash model with global rearrangements). The maximum
411 likelihood phylogenetic trees were calculated by the PhyML 3.0 web server based on best
412 models according to the Akaike information criterion (Guindon *et al.*, 2010). Bootstrap
413 analysis was also performed in 100 repeats.

414

415 **Acknowledgements**

416 The authors gratefully acknowledge the financial support provided by the Hungarian
417 Scientific Research Fund (OTKA grant K100163) and the Natural Sciences and Engineering
418 Research Council of Canada (NSERC). Thanks are due to Giulia Dowgier and Vito Collela
419 (Aldo Moro University, Bari) for their help in the PCR screening during an Erasmus training
420 programme in Budapest.

421

422 **References**

423 **Abascal, F., Zardoya, R. & Posada, D. (2005).** ProtTest: selection of best-fit models of
424 protein evolution. *Bioinformatics* **21**, 2104-2105.
425 **Ahne, W. & Scheinert, P. (1989).** Reptilian viruses: isolation of parvovirus-like particles
426 from corn snake Elapha guttata (Colubridae). *Zentralbl Veterinarmed B* **36**, 409-412.
427 **Allander, T., Emerson, S. U., Engle, R. E., Purcell, R. H. & Bukh, J. (2001).** A virus
428 discovery method incorporating DNase treatment and its application to the
429 identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A* **98**, 11609-
430 11614.
431 **Belyi, V. A., Levine, A. J. & Skalka, A. M. (2010).** Sequences from ancestral single-
432 stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are
433 more than 40 to 50 million years old. *J Virol* **84**, 12458-12462.
434 **Brown, K. E., Green, S. W. & Young, N. S. (1995).** Goose parvovirus--an autonomous
435 member of the dependovirus genus? *Virology* **210**, 283-291.
436 **Chiorini, J. A., Kim, F., Yang, L. & Kotin, R. M. (1999).** Cloning and characterization of
437 adeno-associated virus type 5. *J Virol* **73**, 1309-1319.
438 **Cotmore, S. F., Agbandje-McKenna, M., Chiorini, J. A., Mukha, D. V., Pintel, D. J.,
439 Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Gatherer, D. &
440 Davison, A. J. (2014).** The family Parvoviridae. *Arch Virol* **159**, 1239-1247.
441 **Day, J. M. & Zsak, L. (2010).** Determination and analysis of the full-length chicken
442 parvovirus genome. *Virology* **399**, 59-64.
443 **Estevez, C. & Villegas, P. (2004).** Sequence analysis, viral rescue from infectious clones and
444 generation of recombinant virions of the avian adeno-associated virus. *Virus Res* **105**,
445 195-208.

446 **Farkas, S.. & Gál, J. (2008).** First Hungarian report of inclusion body hepatitis associated
447 with adenovirus and secondary parvovirus infection in an Indonesian pit viper (Parias
448 (Trimeresurus) hageni). *Magy Allatorvosok* **130**, 775-761. [In Hungarian]
449 **Farkas, S. L., Zádori, Z., Benkő, M., Essbauer, S., Harrach, B. & Tijssen, P. (2004).** A
450 parvovirus isolated from royal python (Python regius) is a member of the genus
451 Dependovirus. *J Gen Virol* **85**, 555-561.
452 **Galli, A. & Cervelli, T. (2014).** Inverted terminal repeats of adeno-associated virus decrease
453 random integration of a gene targeting fragment in Saccharomyces cerevisiae. *BMC*
454 *Mol Biol* **15**, 5.
455 **Gao, G., Alvira, M. R., Somanathan, S., Lu, Y., Vandenberghe, L. H., Rux, J. J.,**
456 **Calcedo, R., Sanmiguel, J., Abbas, Z. & Wilson, J. M. (2003).** Adeno-associated
457 viruses undergo substantial evolution in primates during natural infections. *Proc Natl*
458 *Acad Sci U S A* **100**, 6081-6086.
459 **Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O.**
460 **(2010).** New algorithms and methods to estimate maximum-likelihood phylogenies:
461 assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.
462 **Hanson, L. A., Rudis, M. R., Vasquez-Lee, M. & Montgomery, R. D. (2006).** A broadly
463 applicable method to characterize large DNA viruses and adenoviruses based on the
464 DNA polymerase gene. *Virol J* **3**, 28.
465 **Heldstab, A. & Bestetti, G. (1984).** Virus associated gastrointestinal disease in snakes. *J Zoo*
466 *Anim Med* **15**, 118-128.
467 **Jacobson, E. R., Kopit, W., Kennedy, F. A. & Funk, R. S. (1996).** Coinfection of a bearded
468 dragon, Pogona vitticeps, with adenovirus- and dependovirus-like viruses. *Vet Pathol*
469 **33**, 343-346.

470 **Ilyina, T. V. & Koonin, E.V. (1992).** Conserved sequence motifs in the initiator proteins for
471 rolling circle DNA replication encoded by diverse replicons from eubacteria,
472 eucaryotes and archaebacteria. *Nucleic Acids Res* **20,** 3279-3285.
473 **Katzourakis, A. & Gifford, R. J. (2010).** Endogenous viral elements in animal genomes.
474 *PLoS Genet* **6**, e1001191.
475 **Kim, D. Y., Mitchell, M. A., Bauer, R. W., Poston, R. & Cho, D. Y. (2002).** An outbreak of
476 adenoviral infection in inland bearded dragons (Pogona vitticeps) coinfected with
477 dependovirus and coccidial protozoa (Isospora sp.). *J Vet Diagn Invest* **14**, 332-334.
478 **Kotin, R.-M. & Smith, R.-H. (2001).** Dependovirus. In *The Springer Index of Viruses* pp.
479 703-707. Edited by C. A. Tidona, G. Darai & C. Büchen-Osmond. Berlin, Heidelberg:
480 Springer Verlag.
481 **Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A.,**
482 **McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J.**
483 **D., Gibson, T. J. & Higgins, D. G. (2007).** Clustal W and Clustal X version 2.0.
484 *Bioinformatics* **23**, 2947-2948.
485 **Le Gall-Recule, G. & Jestin, V. (1994).** Biochemical and genomic characterization of
486 muscovy duck parvovirus. *Arch Virol* **139**, 121-131.
487 **Li, Y., Ge, X., Hon, C. C., Zhang, H., Zhou, P., Zhang, Y., Wu, Y., Wang, L. F. & Shi, Z.**
488 **(2010).** Prevalence and genetic diversity of adeno-associated viruses in bats from
489 China. *J Gen Virol* **91**, 2601-2609.
490 **Mizokami, M., Orito, E., Ohba, K., Ikeo, K., Lau, J.Y.& Gojobori, T. (1997).** Constrained
491 evolution with respect to gene overlap of hepatitis B virus. *J Mol Evol* **44**, 83-90.
492 **Naumer, M., Sonntag, F., Schmidt, K., Nieto, K., Panke, C., Davey, N. E., Popa-Wagner,**
493 **R. & Kleinschmidt, J. A. (2012).** Properties of the adeno-associated virus assembly-
494 activating protein. *J Virol* **86**, 13038-13048.

495 **Ogawa, M., Ahne, W. & Essbauer, S. (1992).** Reptilian viruses: adenovirus-like agent
496 isolated from royal python (Python regius). *Zentralbl Veterinarmed B* **39**, 732-736.
497 **Papp, T., Fledelius, B., Schmidt, V., Kajan, G. L. & Marschang, R. E. (2009).** PCR-
498 sequence characterization of new adenoviruses found in reptiles and the first
499 successful isolation of a lizard adenovirus. *Vet Microbiol* **134**, 233-240.
500 **Pénzes, J.J. & Benkő, M. (2014).** Novel parvovirus from the worm lizard Trogonophis
501 wiegmanni - First virus ever detected in amphisbaenian hosts. *Acta Vet Hung* **62**, 284-
502 292.
503 **Pénzes, J. & Doszpoly, A. (2011).** Detection of adenoviral infection in bearded dragons
504 (Pogona vitticeps) in Hungary. *Magy Allatorvosok* **133**, 432-437. [In Hungarian]
505 **Qiu, J. & Pintel, D. (2008).** Processing of adeno-associated virus RNA. *Front Biosci* **13**,
506 3101-3115.
507 **Qiu, J., Yoto, Y., Tullis, G. & Pintel, D. J. (2006).** Parvovirus RNA processing strategies. In
508 *Parvoviruses*, pp. 253-273. Edited by J. R. Kerr, S. Cotmore, M. E. Bloom, R. M.
509 Linden & C. R. Parrish. London, UK: Hodder Arnold Publishers.
510 **Ruffing, M., Heid, H. & Kleinschmidt, J. A. (1994).** Mutations in the carboxy terminus of
511 adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD
512 integrin-binding motif. *J Gen Virol* **75 (Pt 12)**, 3385-3392.
513 **Smith, D.H.,Ward, P. & Linden, R.M. (1999).** Comparative Characterization of Rep
514 Proteins from the Helper-Dependent Adeno-Associated Virus Type 2 and the
515 Autonomous Goose Parvovirus. *J Virol* **73** 2930-2937.
516 **Sonntag, F., Kother, K., Schmidt, K., Weghofer, M., Raupp, C., Nieto, K., Kuck, A.,**
517 **Gerlach, B., Bottcher, B., Muller, O. J., Lux, K., Horer, M. & Kleinschmidt, J. A.**
518 **(2011).** The assembly-activating protein promotes capsid assembly of different adeno-
519 associated virus serotypes. *J Virol* **85**, 12686-12697.

520 **Staden, R., Beal, K. F. & Bonfield, J. K. (2000).** The Staden package, 1998. *Methods Mol*
521 *Biol* **132**, 115-130.
522 **Tijssen, P., Agbandje-McKenna, M., Almendral, J. M., Bergoin, M., Flegel, T. W.,**
523 **Hedman, K., Kleinschmidt, J. A., Li, Y., Pintel, D. J. & Tattersall, P. (2011).**
524 Parvoviridae. In *Virus taxonomy: classification and nomenclature of viruses: Ninth*
525 *report of the International Committee on Taxonomy of Viruses*, pp. 375-395. Edited by
526 A. M. Q. King, M. J. Adams, E. Carstens & E. J. Lefkowitz. San Diego: Elsevier.
527 Tijssen, P., Bergoin, M. (1995). Densonucleosis viruses constitute an increasingly diversified
528 subfamily among the parvoviruses. *Semin Virol* **6**, 347-355.
529 **Tse, H., Tsoi, H. W., Teng, J. L., Chen, X. C., Liu, H., Zhou, B., Zheng, B. J., Woo, P. C.,**
530 **Lau, S. K. & Yuen, K. Y. (2011).** Discovery and genomic characterization of a novel
531 ovine partetravirus and a new genotype of bovine partetravirus. *PLoS One* **6**, e25619.
532 **VanDevanter, D. R., Warrener, P., Bennett, L., Schultz, E. R., Coulter, S., Garber, R. L.**
533 **& Rose, T. M. (1996).** Detection and analysis of diverse herpesviral species by
534 consensus primer PCR. *J Clin Microbiol* **34**, 1666-1671.
535 **Wellehan, J. F., Johnson, A. J., Harrach, B., Benko, M., Pessier, A. P., Johnson, C. M.,**
536 **Garner, M. M., Childress, A. & Jacobson, E. R. (2004).** Detection and analysis of
537 six lizard adenoviruses by consensus primer PCR provides further evidence of a
538 reptilian origin for the atadenoviruses. *J Virol* **78**, 13366-13369.
539 **Wozniak, E.-J., DeNardo, D.-F., Brewer, A., Wong, V. & Tarara, R.-P. (2000).**
540 Identification of adenovirus and dependovirus-like agents in an outbreak of fatal
541 gastroenteritis in captive born California mountain kingsnakes, Lampropeltis zonata
542 multicincta. *J Herpet Med Surg* **10**, 4-7.
543 **Wu, P., Xiao, W., Conlon, T., Hughes, J., Agbandje-McKenna, M., Ferkol, T., Flotte, T.**
544 **& Muzyczka, N. (2000).** Mutational analysis of the adeno-associated virus type 2

545 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. *J Virol*
546 **74**, 8635-8647.

547 **Yakobson, B., Koch, T. & Winocour, E. (1987).** Replication of adenoassociated virus in
548 synchronized cells without the addition of a helper virus. *J Virol* 61, 972-981.

549 **Zádori, Z., Stefancsik, R., Rauch, T. & Kisary, J. (1995).** Analysis of the complete
550 nucleotide sequences of goose and Muscovy duck parvoviruses indicates common
551 ancestral origin with adeno-associated virus 2. *Virology* **212**, 562-573.

552 **Zádori, Z., Szelei, J., Lacoste, M. C., Li, Y., Gariepy, S., Raymond, P., Allaire, M., Nabi,**
553 **I. R. & Tijssen, P. (2001).** A viral phospholipase A2 is required for parvovirus
554 infectivity. *Dev Cell* **1**, 291-302.

555 **Zádori, Z., Szelei, J. & Tijssen, P. (2005).** SAT: a late NS protein of porcine parvovirus. *J*
556 *Virol* **79**, 13129-13138.

557 **Zimmerman, L. M., Vogel, L. A. & Bowden, R. M. (2010).** Understanding the vertebrate
558 immune system: insights from the reptilian perspective. *J Exp Biol* **213**, 661-671.

559 **Zsák, L., Strother, K. O. & Kisary, J. (2008).** Partial genome sequence analysis of
560 parvoviruses associated with enteric disease in poultry. *Avian Pathol* **37**, 435-441.

561 **Figure legends**

562 **Figure 1** The aligned homologue partial genome sequences of the three novel reptilian
563 parvoviruses; bearded dragon parvovirus (BDPV), corn snake parvovirus (CSPV) and pygmy
564 chameleon parvovirus (PCPV), respectively. The approx. 1.5-kb long sequence of each
565 genome corresponded with the central region including the partial *rep* and *cap* ORFs, two
566 introns spliced from a common donor site and the complete alternative ORF of the assembly-
567 activating protein that is specific for genus *Dependoparvovirus*.

568

569 **Figure 2** Multiple alignment based on the AAP aa sequences of dependoparvoviruses, with
570 each species represented by at least one type. Viruses of diapsid origin are highlighted in bold,
571 while those of reptilian origin are underlined. Conserved regions of the protein are marked by
572 horizontal lines. Continuous lines represent motifs that are preserved throughout the whole
573 genus, dotted lines stand for those that are absent or highly reduced in reptilian parvoviruses.
574 Dashed lines mark regions that are completely absent from all PVs of diapsid origin. The
575 broken line stands for motifs absent in reptilian, but present in a reduced version some avian
576 dependoparvoviruses. Abbreviations: AAV - adeno-associated virus, AAAV - avian adeno-
577 associated virus, BAAV - bovine adeno-associated virus, BtAAV - bat adeno-associated
578 virus, BDPV - bearded dragon parvovirus, CSPV - corn snake parvovirus, CslAAV -
579 California sea lion adeno-associated virus, DPV - duck parvovirus, GPV - goose parvovirus,
580 PCPV - pygmy chameleon parvovirus.
581
582 **Figure 3** Organization of the complete genome of the bearded dragon parvovirus (A) and the
583 secondary hairpin-like structure of its left ITR (B). The dark, thick arrows represent the two
584 main ORFs (*rep*, *cap*) and the only alternative ORF (coding for assembly-activating protein,
585 AAP) is presented in white. The further arrows represent the presumed transcripts of the *rep*
586 and *cap* genes, respectively. Black arrows indicate the positions of the three promoters, while
587 the white boxes stand for the inverted terminal repeats (ITRs). The positions of the putative
588 polyadenilation signals are indicated with gray boxes. The darker the box, the higher is the
589 score supporting the signal according to *in silico* predictions. The total length of the genome
590 is 4590 nt out of which 257 corresponds to each ITR. The secondary structure of the telomeric
591 hairpins represented 130 bases. The side-arms of the telomeric T-structure occurred in two
592 alternative orientations, “flip” and its reverse-complement “flop”.
593

594

595 **Figure 4** Results of phylogeny reconstructions. Diapsid dependoparvoviruses are highlighted
596 in bold, and the novel squamate parvoviruses are underlined. All squamate parvoviruses can
597 be included in the *Dependoparvovirus* genus. The calculations based on amino acid (aa)
598 sequences of the VP protein fragment (204 aa after gap removal) (A), obtained from the PCR
599 screening, proves the short fragment to be suitable for the classification of these parvoviruses
600 at genus level (maximum likelihood, 204 aa, LG+I+G+F with α=1.59, ρinv=0.04).
601 Calculations according to the complete derived aa sequence of the rep ORF (B) provide better
602 resolution of evolutional relationships within genera (maximum likelihood, RtREV+I+G+F,
603 α=1.29, ρinv=0.03), where the monophyletic branch of reptilian parvoviruses appears to be
604 the most basal cluster of genus *Dependoparvovirus*. The tree based on the full aa sequence of
605 the genus-specific alternative ORF, the assembly-activating protein (AAP) (C) supports the
606 monophyly of squamate parvoviruses yet disproves the monophyly of diapsid parvoviruses
607 (maximum likelihood, HIVb+G+F, α=0.95). Abbreviations: AAV - adeno-associated virus,
608 AMDV - Aleutian mink disease virus, AV - amdovirus, ErPV - *Erythroparvovirus,* MV-
609 minute virus and PV - parvovirus.

610

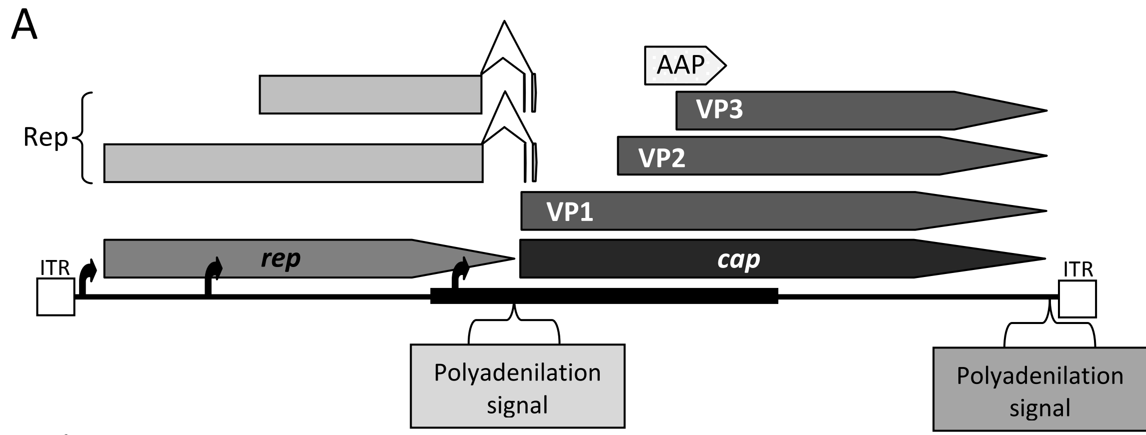
611 **Table 1** Intron lengths of squamate (dark gray background), avian (light gray background),
612 and mammalian (white background) parvoviruses. A general expanding tendency can be
613 observed especially in the length of the first introns. Abbreviations: AAV - adeno-associated
614 virus, PV - parvovirus.

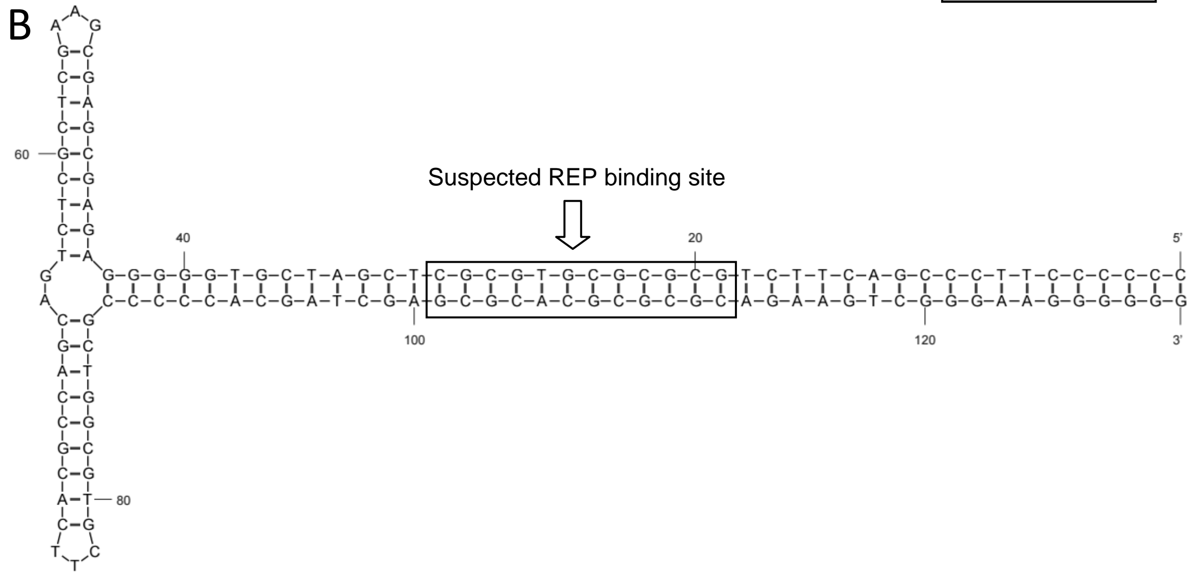
615

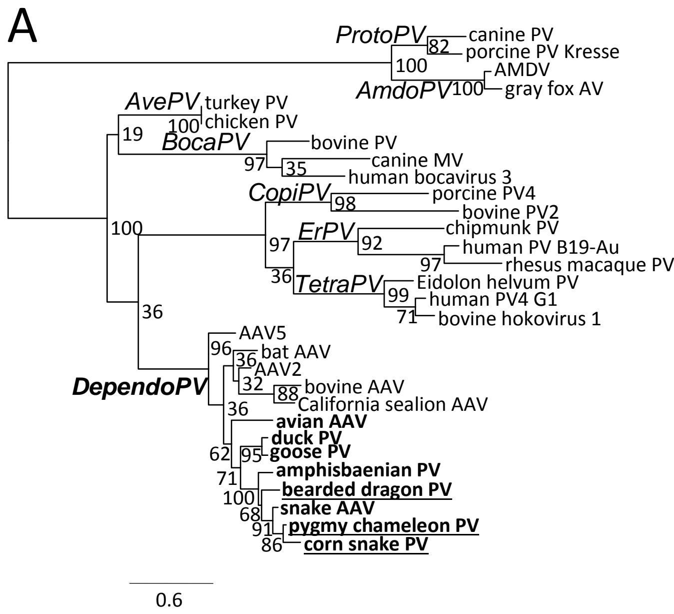
616

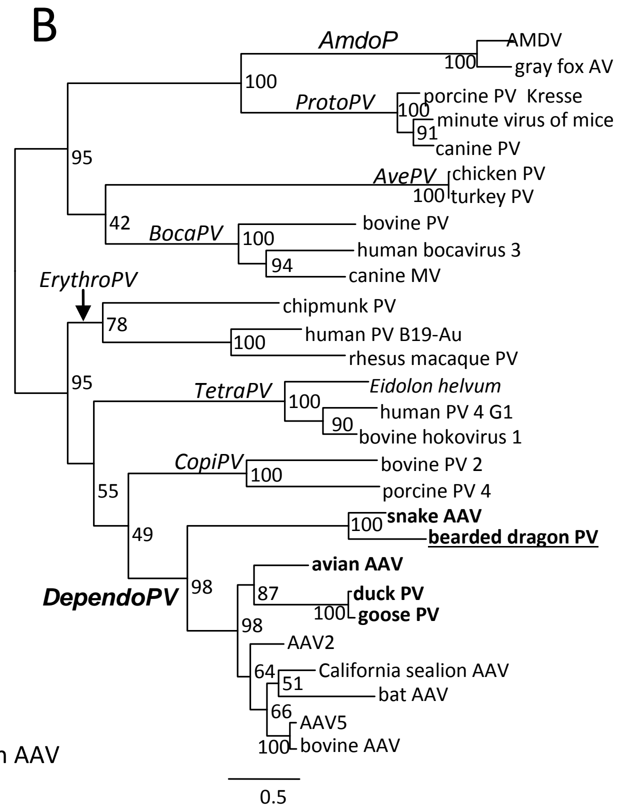
```
CSPV    8 TTTAAATTATAAATTGGCGCCAAATTTCGGTAAGGTGACTGAACAGGAAGTAAAAGAATTTATTACTTGGGGGCGTGGTTTAAATATCGA
PCPV    1 -----------------------CCTTTGGGAAGGTCACAGAACAGGAAGTTAAAGAGTTTATTACTTGGGGGCGGAGCCTAGACATTGA
BDPV 1711 GTTGAATAAGAAACTGGAACCTGATTTTGGAAAAGTGACCTTGGACGAAGTCAAAGAATTTATTACCTGGGGTAGAGATAATCCAGTACA
                                 ** ** **.** **    . * ***** *****.******** *****  * ..        .*  *

                                                   p40 promoter TATA box
CSPV   98 CATTCCGTATCAGTTCCGGGTTCCTACT--------------TCCGGTAGCTATAAAAGGCGGGGCTCCGAGGACGCTGCCTCATTTTCT
PCPV   68 AGTACCTCACCAGTTTAGAGTGCCAGTG--------------TCTGGCGCCTATAAAAGGCCGGCCCCTGAGGCGGAAGCTCATTCTTCG
BDPV 1801 AGTACCGTATCAGTTTCGAGTACCCTCTGTAGCCACGCCCCCTCAGAAAAGTATAAATGA-GGTGCTGGGCAAGCGCCGCGCCATTTCTG
           .* **  * *****  *.** **                  ** *. .  ****** *.  *  *   * ..  *  **    * *

                                                        Presumed donor site
CSPV  174 TTGGAGCCGCCGAG----CAAGAAGGAC----G---TGAATCCCCAAATATCTGCCGCGACCaagtACGTATGTAATTTAGT---CGATA
PCPV  144 GATGAGCAGCCAAAG----GAGAAGGTCGCACGCCTTGACGACTCTCTAACCaggtATG-CTAATAACATTGATAAGTCAGCTACCGGGA
BDPV 1890 ACGGCGCGGGCGAAGA---GACGAGATCGACCAAGTTGGTGCTGCTGAATGATTCTCTAACCaggtATTGTAACAATATTAC---TGAAC
            * ** * *.*.     .* .**. *    .    **.      *                .  * *.  *    . **     .      *.

CSPV  250 AAGCTTCTGTTTCTGAATTGGCTAAAACAATCAATGTATGTTACATCATGCTTTTGAATGTACCGAGTGTTATCCTGAATTGTTGGATG
PCPV  229 GGGAAGCTTTCCTAGAGATTGCTACTACCAATCAGTGTATGTTGCACCATACCTTTTCATGTACTGAATGTTATCCTGAATTGGTTGATG
BDPV 1974 GTGTCAATACGCGTGAAATTGCTCAAAATAATCAATGTATGTTACACCACGTCTTTAATTGTTCTGAATGTTACCCTGAATTGTTGGATG
          . *    *      **. * ***   *  *****.********.** **  .   ***    *** * **.***** ********* * ****

                                                                Presumed acceptor sites &
     Presumed cryptic poly A site   VP1 start codon            stop codons of the NS proteins
CSPV  340 ACATGGATAAGGAACAATAAAGCTTACTGATA-ATAGACATGGATTTTGTCGATGATTTCTTTACagatAAATACAAagagACCTATAAA
PCPV  319 ACTTGATGAAGGAACAATAAA-CTTATTGATA-ATAGACATGGATTTTCTCGATGATTTTTTTagcaATAAATATAAAGAAACTGtagaA
BDPV 2064 ATTGTGACATGGAACAATAAA--TGATTGAAATATAGCTATGGATTTTCTCGATTTCTTGTTaggtGAAAAATACGaggaGACTGCTAAA
          *    .   * **********  * * *** * ****  ********* *****    ** **  .  .* *****  .*.**.**    .**

CSPV  429 GAGCTTTCTAAGCCCGTCAACCCAAAGCCGGTTCAACAAATTAGCGAAAAGCATTCTGAACCTGGCTCGAGGGGTCTTGTGTTGCCTGGC
PCPV  407 GAACTCTCTAAACCCGTTAACCCACCACCCGTTCAACACGGAAGC-----------------AGG-------GGTCTGGTGGTGCCTGGG
BDPV 2152 GAGTTGGGAAAACCGATTAACCCTCCTCCCGTTCAACAAATTAGCCA---------CGCAGACAGCAGACGCGGTCTAGTGGTTCCAGGT
          **. *    **.** .* *****    ** ********  .  ***                  .*          ***** *** * ** **

CSPV  519 TATAGGTATCTTGGGCCTGGTAATAGCTTGGACCGTGGAGAACCCGTTAACGAGGCGGACGCAGCTGCCCGAGAACACGACATCTCCTAC
PCPV  473 TACCGGTATCTTGGGCCTGGTAATAGCTTGGACCGTGGAGAGCCCGTTAACCAAGCAGACGCAGCAGCTAAAAAGCACGATATCGAATAC
BDPV 2233 TATAAATACCTCGGTCCATTCAACGGATTAGACAAGGGCGAGCCTGTCAACGCAGCTGACGCGGCTGCCCTTGAACACGACAAAGCTTAT
          **   ..** ** ** **    ** .* **.***  . ** **.** ** ***  .** *****.** **    .*.***** *    **

CSPV  609 AACAAACAACTCGAAGTTGGAGACAATCCGTACGTAAAGTACAACCACGCGGACGAAAAACTACAGTCCGATTTACAAGGTGACGTCAGT
PCPV  563 GATAAACAGCTTCAAGCAGGAGACAACCCGTACATCAAGTACAACCACGCGGACGCCGAGTTCCAAAAGGACCTCCAAGGAGATACAAGT
BDPV 2323 AACGAGCTTCTCGAGGCTGGAGACAACCCGTACATCAAGTACAACCACGCGGACGCCGTCTTTCAAGAACGCTTGCAAGGAGATACTAGT
          .* .*.*  **  *.*  ******** ******.* ******************     .  * **.     .  * ***** **  .  ***

CSPV  699 TTTGGCGGGAACGCAGCAAACGCGGTCTTTCAAGCCAAGAAGCGCCTACTAGAACCGTTTGGTCTAGTAGAAGCGCCCCTACCGGCCAAA
PCPV  653 CTAGCCGGCAACGCGGCCAACGCTCTCTTTCAGGCCAAAAAGACTCTGCTAGAGCCTTTGGGCCTAGTAGAGCA--CCCGGGCGGC-AAC
BDPV 2413 TTGGGTGGTAACGCGGCTAACGCGGTTTTTCAATTCAAGAAGCGGTTGCTCGAGCCGTTTGGAGCGGTCGAGCAGCCCCAGCCCGAAAAG
           * *  ** *****.** *****  * *****.  ***.***     *.** **.** ** **   .** **.     *** . * *  **

VP2 start codon
CSPV  789 ACG---GATAAGGGGAAGGT--------AGACGACT-ACTTC-CCCAAAGCGAAAAAGGCTAAACAGACCTTTCAAA---TCCCACCCCC
PCPV  740 ACGTCTGATAAAA-GAAAACCTCCACCAGGACTACTAACTCCACCCAAAACACCTAAA---AAGCAGAAATTTCAAA---TACCAGCTC-
BDPV 2503 ACGGCCGAAAAGCGAAAGACACCCGA-AGAGTGGTTAGCTCAAGAAAAGACTCCAACC---AAACAAAGGTTCCAGA---TACCAGCTCC
          ***   ** **.  .**..           ... . * .**      **..*    *     **.**.*  ** **.*   * *** * *

                  Alternative start codons of the AAP                               VP3 start codon
CSPV  863 CGCTAAAGAAGACCCAGGAGAAGGGTCTTCTGCGCAGTCTGGA---GGTAGCCCAGCCGGTTCCGATACTAGCGGCTCATCTGTCATGGC
PCPV  822 --------------------------CTCCTAATCAGGCTGGCAC---AAGTTCAGCAGGCTCCGATACTAGCGGCACATCTATCATGGC
BDPV 2586 ------AGGACAATCTGGATCAGATTCTCCTTCCACCTCAGGATCCGGCGGTACTGCAGGCTCCAGTTCTAGCGCATCAAATACAATGGC
                                    ** **       * **         .*  * ** ** ***..* ******   **  *.  *****

CSPV  950 TGAAGGAGGAGGCGGACCACTGGCAAGCGATCAACAAGGTGCCGAGGGAGTGGGTAATTCCTCCGGTGATTGGCATTGCGATACCCAATG
PCPV  883 TTCAGGCGGAGGCGGACCGATGGCAGATGATAACCAGGGCGCCGAGGGAGTGGGTAATTCCTCAGGTGATTGGCATTGCGATACCCAGTG
BDPV 2670 TCAAGGAGGTGGCGGACCAATGGCAGACGATAACCAAGGCGCCGAGGGAGTGGGTAATGCCTCGGGAGATTGGCATTGCGATACCCAATG
          *  *** ** ********. *****.. *** * **.** ***************** **** ** ******************.**

CSPV 1040 GCTGGGAGACCACGTCATTACAAAGTCGACCCGAACTTGGGTACTCCCCAGTTATGGGAATCACCTCTATAAACCCATCACCTTTGATGG
PCPV  973 GCTGGGAGACCACGTCATTACAAAGTCGACCAGAACTTGGGTGCTCCCCACTTACGGGAATCATCTCTACGGGCCTATCAATTTTGACGG
BDPV 2760 GCTGGGCGACCACGTCATTACAAAATCTACCAGAACTTGGGTTCTGCCCTCTTACGGGAATCATCTCTACTCGCCCATCAACTTTGATGG
          ****** *****************.** *** ********** ** ***   *** ******** *****    .** ****   ***** **

CSPV 1130 CACTACCGGAGGCGGAAGTGACGCAGCCTATGCAGGTTACTCCACCCCCTGGGGATACTTTGACTTTAACCGATTCCATTGTCACTTCTC
PCPV 1063 CACCAGCGGCGCGGGCGCTAATGCAGCCTATGCAGGATACAAGACCCCTTGGGGGTATTTTGACTTCAACCGATTCCACTGCCACTTTTC
BDPV 2850 AACCACAGGGAACGGAACCCAAGCCGCTTACTGCGGATACGCTACCCCCTGGGCCTACTTTGACTTTAACCGATTCCACTGCCACTTTTC
           ** *  **  .  **  .   * ** ** **    ** ***   ***** ****   ** ******** ********** ** ***** **

CSPV 1220 CCCACGTGACTGGCAAAGACTCGTCAACAACCACGTGGGCATCCGACCCAAAGGACTTAAATTTAAACTCTTTAACGTCCAAGTCAAGGA
PCPV 1153 ACCCAGAGACTGGCAGCGACTCATCAATAACCACACAGGCATCAGACCTAAAGGACTCAAAGTTAAAGTCTTCAACGTGCAGGTCAAAGA
BDPV 2940 CCCCCGAGACTGGCAAAGACTCATTAACAACCATACCGGAATACGACCAGTCGGACTCAAATTCAAGCTGTTCAACATCCAGGTCAAGGA
           **  * ********. *****.* ** *****  .   ** **   ****  .   ***** *** * **. * ** ***.* **.*****.**

CSPV 1310 AGTCACGCAACAAGACTCGACCAAGACCATCGCCAATAACCTCACCAGCACCGTACAGGTGTTTGCGGACGAGAACTACGAGTTACCGTA
PCPV 1243 AGTTACCACGCAAGATTCGACCAAAACGATCGCCAACAATCTCACCAGCACCGTACAGGTGTTTGCGGACGAGAACTACGACCTCCCCTA
BDPV 3030 AATCACAGTACAAGATTCGACCAAAACGATCGCCAACAATCTCACCAGCACCGTACAGGTCTTTGCGGACACGGAGCACCAGCTCCCGTA
          *.* **    .***** ********.** ******** ** ****************** ********. *.*   ** *   * ** **

   AAP stop codon
CSPV 1400 TGTCTTAGGATCGGCTACTCAAGGAACTTTCCCGCCTTTTCCCAACGACATTTTCATGTTGCCTCAGTA
PCPV 1333 TGTGCTAGGCGGTGCTACACAAGGCACGTTCCCTCCTTTTCCAAATGATGTTTTTATGCTGCCTCAATA
BDPV 3120 CGTATTAGGAAATGCCACGCAGGGCACGTTTCCTCCCTTTCCGGCTGAAGTCTTTCAGTTGCCTCAGTA
           **  ****    ** ** **.** ** ** ** ** *****  .   ** .* **    * *******.**
```

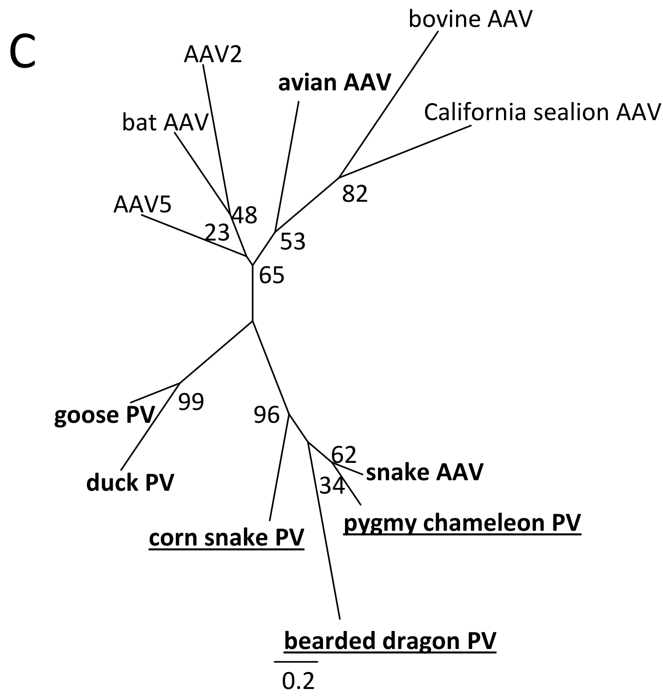
```
                                                  Hydrophobic region      Conserved core              Proline-rich

PCPV     1 ------------------LAQVQQAPILAAHLSWLQAEADRWQMITRAPREWVIPQVIGIAIPSGWETTSLQSRPELGCS
SAAV     1 ------------------LEGAQQVPILAAHLSWLQEEAVRWQTITRAPREWVIPQVIGIAIPSGWETTSLQSQPELGCS
CSPV     1 ---------------LRSLEVAQPVPILAAHLSWLKEEADHWQAINKVPREWVIPPVIGIAIPNGWETTSLQSRPELGYS
BDPV     1 ---------LDQILLPPQDPAVLQAPVLAHQIQWLKEVADQWQTITKAPREWVMPREIGIAIPNGWATTSLQNLPELGFC
BAAV     1 LPERDSTLTTNLEPETGLPQKDHLPELCLLRLKCVQQLAEMVAMRDKVPREWVMPPVIGIAIPLGQRATSPPPQPAPGSC
CslAAV   1 -------------------LAEELPTSCLEMLKWLQWAGEHRASTARVPREWVLPRVIGIAIPSGQKDTSEPPAPEPGCC
AAAV     1 -------------------LEQQHPPLVWDHLSWLKEVAAQWAMQARVPMEWAIPPEIGIAIPNGWKTESSLEPPEPGSC
AAV5     1 ----------LDPADPSSCKSQPNQPQVWELIQCLREVAAHWATITKVPMEWAMPREIGIAIPRGWGTESSPSPPEPGCC
BtAAV    1 ------------------------------LLRWLRVVADQWATMHKVPMEWVMPQEIGITIPFGWTALSSPSPPEPGAC
AAV2     1 ------METQTQYLTPSLSDSHQQPPLVWELIRWLQAVAHQWQTITRAPTEWVIPREIGIAIPHGWATESSPPAPEPGPC
DPV      1 --------------------LPPKAPNLWQHLTWQREEAELWATLQGVPMEWVMPQEIGIAIPNGWETQSLPRLQEPGSC
GPV      1 -------------------------------LKWQREEAELWATLQGVPMEWVMPREIGIAIPNGWETQSSQRPPEPGSC

            T/S-rich 1          T/S-rich 2     T/S-rich 3       T/S-rich 4       T/S-rich 5  Basic region

PCPV    63 PLTGIISTGLSILTAPAARALMQPMQDTRPLGGILTSTDSTATFHPETGSDSSITTQASDLKDSKLKSSTCRSKKLPRKI
SAAV    63 PLTGIISTGLSTLTAPQVRVLMQPMQDTRLPGGTLTSIDSIATSPPETGKDSSTTTQASGRKDSKSKSLTSKSKKLQHKI
CSPV    66 PVMGITSINPSPLMALPEAEVTQPMQVTPPPGDTLTLTDSIVTSPHVTGKDSSTTTWASDPKDLNLNSLTSKSRKSRNKT
BDPV    72 PLTGIISTRPSTLMEPQGTEPKPLTADTLPPGPTLTLTDSTATFPPETGKDSLTTIPEYDQSDSNSSCSTSRSRKSQYKI
BAAV    81 RPTTTTCTCG---SARATPA--TPSTDSPPPGDTLTLTASTATSRQETGKGSSTTTGDCAPKACKSASSTSKLRRSRRLT
CslAAV  62 PATTTTCING---LEVAHNP--IPTTDSPPPGDTSTSIDGTVTSVLGTGNVSSTTTGASDQKDLMLNCSTYKSKRSRRKG
AAAV    62 PATTTTCTNESKDPAEATTT--TNSLDSAPPGDTLTTIDSTATFPRETGNDSSTTTGASVPKRCALDSLTSRLKRSRSKT
AAV5    71 PATTTTSTERSKAAPSTEATP-TPTLDTAPPGGTLTLTASTATGAPETGKDSSTTTGASDPGPSESKSSTFKSKRSRCRT
BtAAV   51 PPTTTTSTARSSPAAPETAR----TLVTARLGDTSISIDSTATFLPGTGSGSSTTTGASAPSGSTLSSSTSRSRRSRRPT
AAV2    75 PPTTTTSTNK-FPANQEPRTT-ITTLATAPLGGILTSTDSTATFHHVTGKDSSTTTGDSDPRDSTSSSLTFKSKRSRRMT
DPV     61 QATTTTCTKP--SQAEQTQTQIPNMLDTAPPGGTLISTDSTAISLQETGRDSSTTIGGLDRKHSNSRYSMCKLKKSRRKT
GPV     50 QATTTTSTKQ--LPAEPLKMQMSSMQDTVPPGGTLISTASTATSPLETGRDLSTTIGESDPSLLNSRSSMSKSKKSQRRI

PCPV   143 RPKRSPTISPAPYRCLRTRTTTSPMC--------------------------------------------
SAAV   143 QRKQLPTISPAPYRSLRTRTTTYHMY--------------------------------------------
CSPV   146 RPRPSPITSPAPYRCLRTRTTSYRMS--------------------------------------------
BDPV   152 RPKRSPTISPAPYRSLRTRSTSSRTY--------------------------------------------
BAAV   156 GRRPYPTTSPARSRSLRTARTSSRT---------------------------------------------
CslAAV 137 GRRPSPITLPARFRCLRTTRTNSRMLSTRVTRGHCRRSQTTCL---------------------------
AAAV   140 STPPSATTSPVRSRSLRTRTTNCRTSSDRLPKAPSRRSQRISTRSRSTGTAR------------------
AAV5   150 PPPPSPTTSPPPSKCLRTTTTSCPTSSATGPRDACRPSLRRSLRCRSTVTRR------------------
BtAAV  127 APRPSPTTSPAPYRSLRTRSTSSRTCSATPTRAACRRSRRTSSCCRSTRT--------------------
AAV2   153 VRRRLPITLPARFRCLLTRSTSSRTSSARRIKDASRRSQQTSSWCHSMDTSP------------------
DPV    139 RQRLLLTTLPLQSRYSRIMNTSCPMFWARPRRGRCHRSPQMCMPCPSTATAQCTPTRVERDSMTEVPSIA
GPV    128 RQRPLQTISPQQFKSLRMMSINSRMSWARLRKAPCRRSRRMSMPCRSTGTAQCTPTRMEHGSMTVVHSTA
```


A
AAP
VP3
Rep
VP2
VP1
rep
cap
ITR
ITR
Polyadenilation signal
Polyadenilation signal


B
Suspected REP binding site
60
40
20
5'
100
120
3'
80


A
ProtoPV
canine PV
82
porcine PV Kresse
100
AMDV
AmdoPV
100
gray fox AV
AvePV
turkey PV
100
chicken PV
19
BocaPV
bovine PV
97
canine MV
35
human bocavirus 3
CopiPV
porcine PV4
98
bovine PV2
100
ErPV
chipmunk PV
97
92
human PV B19-Au
97
rhesus macaque PV
36
TetraPV
Eidolon helvum PV
99
human PV4 G1
71
bovine hokovirus 1
36
AAV5
96
bat AAV
36
AAV2
32
bovine AAV
DependoPV
88
California sealion AAV
36
avian AAV
duck PV
62
95
goose PV
71
amphisbaenian PV
100
bearded dragon PV
68
snake AAV
91
pygmy chameleon PV
86
corn snake PV
0.6


B
AmdoP
AMDV
100
gray fox AV
100
ProtoPV
porcine PV Kresse
100
minute virus of mice
91
canine PV
95
AvePV
chicken PV
100
turkey PV
42
BocaPV
bovine PV
100
human bocavirus 3
94
canine MV
ErythroPV
78
chipmunk PV
100
human PV B19-Au
rhesus macaque PV
95
TetraPV
Eidolon helvum
100
human PV 4 G1
90
bovine hokovirus 1
CopiPV
bovine PV 2
100
porcine PV 4
55
snake AAV
100
bearded dragon PV
49
avian AAV
87
duck PV
100
goose PV
DependoPV
98
98
AAV2
64
California sealion AAV
51
bat AAV
66
AAV5
100
bovine AAV
0.5


C
bovine AAV
AAV2
avian AAV
California sealion AAV
bat AAV
82
AAV5
48
23
53
65
goose PV
99
96
duck PV
62
snake AAV
34
corn snake PV
pygmy chameleon PV
bearded dragon PV
0.2