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

Running title: Novel parvoviruses in reptiles

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

Addresses

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

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

*__Corresponding author__

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

Phone: +36 1 467-4084

Contents Category: Animal – Small DNA viruses

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

Accession numbers:

Bearded dragon parvovirus – KP733794

Pygmy chameleon parvovirus – KP733796

Corn snake parvovirus – KP733795

Abstract

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

Introduction

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

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

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

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

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

Results

PCR screening

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

Complete and partial genome characterization of novel reptilian parvoviruses

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

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

The complete genome of the BDPV consisted of 4590 nt with ITRs of 257 nt each. The genome organization is presented in Fig 3(A). Within the ITRs, a regular, T-shaped hairpin structure could be predicted. Such secondary structure within the ITRs is typical of members of genus *Dependoparvovirus*, except goose PV (GPV) and duck PV (DPV) (Zádori *et al.*, 1995). The predicted hairpin consisted of 130 nt as shown in Fig. 3(B). The genome core flanked by the ITRs consisted of 4076 nt only, being the shortest out of all dependoparvoviruses. The length of the full genome, including the ITRs, qualified as the second shortest after that of SAAV. The genome contained two major ORFs corresponding to the *rep* and *cap* genes of all PVs known to date (Qiu *et al.*, 2006), with lengths of 1518 and 2178 nt, respectively. The splice donor- and acceptor sites of the Rep-protein-coding genes of the other two reptilian PVs had similar positions (Fig. 1). Three putative promoters, corresponding to the P5, P19 and P40 of AAV2, were identified in the genome at positions of nts 323, 803, and 1852 (TATA boxes are shown in Fig. 1 and Fig. 3(A)). We identified two predicted poly(A) sites in the BDPV genome, out of which the one with a higher score follows the *cap* gene at position 4278. Another, supposedly cryptic poly(A) site was, however, observed within the intron in the middle of the genome analogous to AAV5 at the position of 2076 (Qiu *et al.*, 2006).

As for the proteins, the BDPV genome is predicted to encode at least two Rep proteins. The presence of the two presumed introns and the predicted inner start codon, following the putative promoter homologous to P19, however, suggests that there might be four Reps expressed, homologous to those of AAV2, respectively. In the protein sequence of the Rep1, supposedly homologous to the AAV2 Rep78, the putative replication initiative motif I and II (Ilyna & Koonin, 1992) as well as the putative tripartite helicase superfamily III motifs could be identified (Tijssen & Bergoin, 1995; Smith et al., 1999).The PLA2 motif (Zádori *et al.*, 2001) was present in the N-terminal region of the putative VP1 protein

sequence in all three viruses. The putative AAP binding site was identified close to the C-terminal of the VP proteins (Naumer *et al.*, 2012). It surrounds M689 at the aa sequence of VEMLWEV. The putative Rep protein sequence shared most identity with its homologue in SAAV (58%) and not less than 34% with all other members of the *Dependoparvovirus* genus. The VP protein sequence also displayed the highest identity with that of SAAV (70%) while this value was not lower than 56% in case of other dependoparvoviruses.

Phylogenetic analysis

Phylogenetic tree reconstructions were performed in case of all three proteins. As for the VP protein, only the partial, 226-aa-long fragments obtained during the PCR screening were used. The phylogenetic trees according to the Rep and AAP proteins were based on the entire deduced aa sequence. All three novel reptilian PVs could be included in the *Dependoparvovirus* genus according to the short fragment of the VP protein, presented in Fig. 4(A). The complete Rep protein sequence of the BDPV, clustered with SAAV Rep protein, forming the most basal group within dependoparvoviruses as shown in Fig. 4(B). The AAP-based tree presented in Fig. 4(C) indicated that all reptilian PVs formed a monophyletic cluster, yet this did not apply for all diapsid PVs.

Discussion

Parvovirus infection in lizards implies the ability of autonomous replication of reptilian dependoparvoviruses

A complete clone and sequence were obtained of the bearded dragon lizard PV and an incomplete sequence of the short-tailed pygmy chameleon PV, besides detecting another novel snake PV that is only the third one to date. Our primers (Pénzes & Benkő, 2014) proved 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

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

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

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

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

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

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

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

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

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

is a basal cluster of the same genus as shown in Fig. 4(B). Currently this contradiction is difficult to resolve, however, the less significant role of the reptilian adaptive immune system should also be considered in this issue (Zimmerman *et al.*, 2010). This might suggest the selection pressure to be somewhat lower on antigenic viral proteins compared to that of mammalian ones. As this might lead to the more flexible evolution of reptilian parvoviral capsid proteins, it could also manifest in homoplasy of the mammalian ones. The *rep* gene, which encodes a non-structural protein is probably exempt from this selective pressure and therefore presumably is more suitable for examining evolutional relationships below the genus level, when the examination is based on complete aa sequences.

According to AAP sequences, reptilian PVs display monophyly with anseriform dependoparvoviruses, yet avian AAV (AAAV) clusters with mammalian AAVs. It is worth mentioning that on a Rep-based tree, published earlier in an in silico analysis on endogenous viral elements (Katzourakis & Gifford, 2010), AAAV also appeared split from the anseriform dependoparvoviruses by an integrated sequence found in a mammalian (dolphin) genome. As the protein is specific for the *Dependoparvovirus* genus, the phylogenetic calculations do not provide any information on which clade is more basal. Phylogenetic calculations based on the AAP aa sequence, however, raise further concerns, namely its complete overlap with the *cap*, coding capsid proteins responsible for antigenic traits. When resolving this contradiction, it is important to point out that the *cap* is more conserved within the *Dependoparvovirus* genus (Cotmore et al., 2014). In case of hepadnaviruses, a similar conservation has been detected within overlapping ORFs, explaining their suitability for phylogeny reconstructions (Mizokami *et al.*, 1997). In this case, it is likely that the AAP sequence is under a stronger functional selection pressure than capsid proteins are, hence it might serve as a better subject for eluding phylogenetic relationships than capsid proteins in general.

Conclusions

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

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

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

length to expand continuously throughout the evolution of dependoparvoviruses, then a Squamata origin of the genus turns out to be more likely. This is further supported by the phylogeny based on the Rep sequence. A similar expanding tendency in case of the AAP conserved motifs is observed, i.e. there are five in total T/S-rich regions in mammalian AAVs, only four in the avian and three in reptilian dependoparvoviruses. There is, however, the exception of the AAAV AAP which has the same organization as the AAPs of mammalian viruses. Even the phylogeny reconstruction does not display any monophyly of this avian PV with anseriform PVs, but with mammalian AAVs instead. This clustering cannot be observed, nevertheless, with the Rep-based phylogenetic tree, yet the monophyly of the most basal clade of squamate PVs is still valid. Earlier results on dependoparvovirus Rep-based phylogeny, however, indicated AAAV to be split from the anseriform dependoparvoviruses by an endogenous viral element originating from a cetacean (Katzourakis &Gifford, 2010).

In summary, our current results suggest that evolution of dependoparvoviruses support rather an exclusive Squamata origin than a common avian-reptilian one. Further research on novel non-squamate reptilian dependoparvoviruses, provided if such viruses exist, would be required to elucidate this enigma.

Methods

Samples and DNA-extraction

162 independent samples originating from various hosts including members of all reptilian orders except for *Rhyncochephalia* were screened. The most represented order was Squamata, from which 110 samples were examined. The samples were collected from deceased individuals obtained from local pet stores or private pet owners.

For nucleic acid extraction, small (15−25 mg) pieces from the internal organs (lungs, liver, intestines, gonads and kidney) were transferred to 2-ml microcentrifuge safe-lock tubes

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

PCR primers and conditions

To check the presence of parvoviral DNA, a consensus primer pair aiming at a conserved region of the *cap* gene was used (forward: GGYGCCGAKGGAGTGGGYAATKCCTC, reverse: TCAAARTTRTTBCCBGTYCTYAGCAT) (Pénzes & Benkő, 2014). As for the PCR program an initial denaturation step at 92°C for 5 minutes was followed by 45 cycles of denaturation at 92°C for 30 seconds, annealing at 46°C for 60 seconds and elongation at 72°C for another 60 seconds. Final elongation was performed at 72°C for 5 minutes, expected to result an approximately 600 base pair(bp)-long fragment amplified from the VP gene. For AdV detection, a very sensitive consensus nested PCR that targets a highly conserved region of the adenoviral DNA-dependent DNA polymerase gene (Wellehan *et al.*, 2004) was applied. Amplification of the herpesviral DNA polymerase gene fragment was carried out by another nested consensus PCR (VanDevanter *et al.*, 1996). Furthermore an expanded-range PCR designed for the DNA polymerase gene of other large DNA viruses was performed (Hanson *et al.*, 2006).

In case of PV positivity, a short fragment from the *rep* gene was targeted for amplification. To this end two different sets of consensus primer pairs were designed, with the sequences of 5'-GTDAAYTGGACYAAYGMRAAC-3' and 5'-AACATNCKBTCYTSYARNGG-3' in case of set A and 5'-TGTGTCARGTMTWTGATGGKAA-3' and 5'-CAATTCAGGRTAACATTCNRWACA-3' for set B. The size of the expected PCR products was approx. 250 or 400 bp, respectively. A PCR program consisting of an initial denaturation 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

utilized while finishing with a final synthesis step at 72°C for 3 min. In order to sequence the full genome, primers facing outwards from the already known sequences were designed, and then submitted to PCR with a reaction mix including only these primers. The PCR products were cloned and sequenced. PCR reactions were in 50 µl volume with AmpliTaq Gold® DNA Polymerase (Life Technologies Corporation®, Carlsbad, CA, USA) or, in case of fragments larger than 1000 bp, with Phusion® High-Fidelity DNA Polymerase (ThermoFisher Scientific®, Waltham, MA, USA) according to the manufacturer's recommendations.

Obtaining the genome end structures

Because of their secondary structure, the ITRs could not be obtained via the single-primer PCR. To solve this problem, a phosphorylated adaptor 5'-ATCCACAACAACTCTCCTCCTC-3' was attached using T4 RNA ligase (New England Biolabs®, Ipswich, MA, USA) to both genome ends. Additional primers were designed specifically to the nearest ClaI restriction sites at both genome ends. A special PCR was applied in 25 µl with OneTaq® Hot Start DNA Polymerase (New England Biolabs®, Ipswich, MA, USA) including GC enhancer and 3 µl of 2 mM EDTA, and using the adaptor reverse primer. The PCR fragments obtained were cloned and sequenced. In case of incomplete ITRs, specific primers were designed for both the flip and flop structures in order to acquire the remaining nucleotides.

Sequencing, molecular cloning and sequence analysis

Sequencing reactions were performed with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Life Technologies Corporation®, Carlsbad, CA, USA), and sent for analysis by a commercial service on an ABI PRISM 3100 Genetic Analyzer (Life Technologies Corporation®, Carlsbad, CA, USA). PCR fragments that were amplified with only one primer

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

Phylogeny reconstructions

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

Acknowledgements

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

References

Abascal, F., Zardoya, R. & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.

Ahne, W. & Scheinert, P. (1989). Reptilian viruses: isolation of parvovirus-like particles from corn snake Elapha guttata (Colubridae). *Zentralbl Veterinarmed B* **36**, 409-412.

Allander, T., Emerson, S. U., Engle, R. E., Purcell, R. H. & Bukh, J. (2001). A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A* **98**, 11609-11614.

Belyi, V. A., Levine, A. J. & Skalka, A. M. (2010). Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old. *J Virol* **84**, 12458-12462.

Brown, K. E., Green, S. W. & Young, N. S. (1995). Goose parvovirus--an autonomous member of the dependovirus genus? *Virology* **210**, 283-291.

Chiorini, J. A., Kim, F., Yang, L. & Kotin, R. M. (1999). Cloning and characterization of adeno-associated virus type 5. *J Virol* **73**, 1309-1319.

Cotmore, S. F., Agbandje-McKenna, M., Chiorini, J. A., Mukha, D. V., Pintel, D. J., Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Gatherer, D. & Davison, A. J. (2014). The family Parvoviridae. *Arch Virol* **159**, 1239-1247.

Day, J. M. & Zsak, L. (2010). Determination and analysis of the full-length chicken parvovirus genome. *Virology* **399**, 59-64.

Estevez, C. & Villegas, P. (2004). Sequence analysis, viral rescue from infectious clones and generation of recombinant virions of the avian adeno-associated virus. *Virus Res* **105**, 195-208.

Farkas, S.. & Gál, J. (2008). First Hungarian report of inclusion body hepatitis associated with adenovirus and secondary parvovirus infection in an Indonesian pit viper (Parias (Trimeresurus) hageni). *Magy Allatorvosok* **130**, 775-761. [In Hungarian]

Farkas, S. L., Zádori, Z., Benkő, M., Essbauer, S., Harrach, B. & Tijssen, P. (2004). A parvovirus isolated from royal python (Python regius) is a member of the genus Dependovirus. *J Gen Virol* **85**, 555-561.

Galli, A. & Cervelli, T. (2014). Inverted terminal repeats of adeno-associated virus decrease random integration of a gene targeting fragment in Saccharomyces cerevisiae. *BMC Mol Biol* **15**, 5.

Gao, G., Alvira, M. R., Somanathan, S., Lu, Y., Vandenberghe, L. H., Rux, J. J., Calcedo, R., Sanmiguel, J., Abbas, Z. & Wilson, J. M. (2003). Adeno-associated viruses undergo substantial evolution in primates during natural infections. *Proc Natl Acad Sci U S A* **100**, 6081-6086.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

Hanson, L. A., Rudis, M. R., Vasquez-Lee, M. & Montgomery, R. D. (2006). A broadly applicable method to characterize large DNA viruses and adenoviruses based on the DNA polymerase gene. *Virol J* **3**, 28.

Heldstab, A. & Bestetti, G. (1984). Virus associated gastrointestinal disease in snakes. *J Zoo Anim Med* **15**, 118-128.

Jacobson, E. R., Kopit, W., Kennedy, F. A. & Funk, R. S. (1996). Coinfection of a bearded dragon, Pogona vitticeps, with adenovirus- and dependovirus-like viruses. *Vet Pathol* **33**, 343-346.

Ilyina, T. V. & Koonin, E.V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res* **20,** 3279-3285.

Katzourakis, A. & Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genet* **6**, e1001191.

Kim, D. Y., Mitchell, M. A., Bauer, R. W., Poston, R. & Cho, D. Y. (2002). An outbreak of adenoviral infection in inland bearded dragons (Pogona vitticeps) coinfected with dependovirus and coccidial protozoa (Isospora sp.). *J Vet Diagn Invest* **14**, 332-334.

Kotin, R.-M. & Smith, R.-H. (2001). Dependovirus. In *The Springer Index of Viruses* pp. 703-707. Edited by C. A. Tidona, G. Darai & C. Büchen-Osmond. Berlin, Heidelberg: Springer Verlag.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Le Gall-Recule, G. & Jestin, V. (1994). Biochemical and genomic characterization of muscovy duck parvovirus. *Arch Virol* **139**, 121-131.

Li, Y., Ge, X., Hon, C. C., Zhang, H., Zhou, P., Zhang, Y., Wu, Y., Wang, L. F. & Shi, Z. (2010). Prevalence and genetic diversity of adeno-associated viruses in bats from China. *J Gen Virol* **91**, 2601-2609.

Mizokami, M., Orito, E., Ohba, K., Ikeo, K., Lau, J.Y.& Gojobori, T. (1997). Constrained evolution with respect to gene overlap of hepatitis B virus. *J Mol Evol* **44**, 83-90.

Naumer, M., Sonntag, F., Schmidt, K., Nieto, K., Panke, C., Davey, N. E., Popa-Wagner, R. & Kleinschmidt, J. A. (2012). Properties of the adeno-associated virus assembly-activating protein. *J Virol* **86**, 13038-13048.

Ogawa, M., Ahne, W. & Essbauer, S. (1992). Reptilian viruses: adenovirus-like agent isolated from royal python (Python regius). *Zentralbl Veterinarmed B* **39**, 732-736.

Papp, T., Fledelius, B., Schmidt, V., Kajan, G. L. & Marschang, R. E. (2009). PCR-sequence characterization of new adenoviruses found in reptiles and the first successful isolation of a lizard adenovirus. *Vet Microbiol* **134**, 233-240.

Pénzes, J.J. & Benkő, M. (2014). Novel parvovirus from the worm lizard Trogonophis wiegmanni - First virus ever detected in amphisbaenian hosts. *Acta Vet Hung* **62**, 284-292.

Pénzes, J. & Doszpoly, A. (2011). Detection of adenoviral infection in bearded dragons (Pogona vitticeps) in Hungary. *Magy Allatorvosok* **133**, 432-437. [In Hungarian]

Qiu, J. & Pintel, D. (2008). Processing of adeno-associated virus RNA. *Front Biosci* **13**, 3101-3115.

Qiu, J., Yoto, Y., Tullis, G. & Pintel, D. J. (2006). Parvovirus RNA processing strategies. In *Parvoviruses*, pp. 253-273. Edited by J. R. Kerr, S. Cotmore, M. E. Bloom, R. M. Linden & C. R. Parrish. London, UK: Hodder Arnold Publishers.

Ruffing, M., Heid, H. & Kleinschmidt, J. A. (1994). Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD integrin-binding motif. *J Gen Virol* **75 (Pt 12)**, 3385-3392.

Smith, D.H.,Ward, P. & Linden, R.M. (1999). Comparative Characterization of Rep Proteins from the Helper-Dependent Adeno-Associated Virus Type 2 and the Autonomous Goose Parvovirus. *J Virol* **73** 2930-2937.

Sonntag, F., Kother, K., Schmidt, K., Weghofer, M., Raupp, C., Nieto, K., Kuck, A., Gerlach, B., Bottcher, B., Muller, O. J., Lux, K., Horer, M. & Kleinschmidt, J. A. (2011). The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. *J Virol* **85**, 12686-12697.

Staden, R., Beal, K. F. & Bonfield, J. K. (2000). The Staden package, 1998. *Methods Mol Biol* **132**, 115-130.

Tijssen, P., Agbandje-McKenna, M., Almendral, J. M., Bergoin, M., Flegel, T. W., Hedman, K., Kleinschmidt, J. A., Li, Y., Pintel, D. J. & Tattersall, P. (2011). Parvoviridae. In *Virus taxonomy: classification and nomenclature of viruses: Ninth report of the International Committee on Taxonomy of Viruses*, pp. 375-395. Edited by A. M. Q. King, M. J. Adams, E. Carstens & E. J. Lefkowitz. San Diego: Elsevier.

Tijssen, P., Bergoin, M. (1995). Densonucleosis viruses constitute an increasingly diversified subfamily among the parvoviruses. *Semin Virol* **6**, 347-355.

Tse, H., Tsoi, H. W., Teng, J. L., Chen, X. C., Liu, H., Zhou, B., Zheng, B. J., Woo, P. C., Lau, S. K. & Yuen, K. Y. (2011). Discovery and genomic characterization of a novel ovine partetravirus and a new genotype of bovine partetravirus. *PLoS One* **6**, e25619.

VanDevanter, D. R., Warrener, P., Bennett, L., Schultz, E. R., Coulter, S., Garber, R. L. & Rose, T. M. (1996). Detection and analysis of diverse herpesviral species by consensus primer PCR. *J Clin Microbiol* **34**, 1666-1671.

Wellehan, J. F., Johnson, A. J., Harrach, B., Benko, M., Pessier, A. P., Johnson, C. M., Garner, M. M., Childress, A. & Jacobson, E. R. (2004). Detection and analysis of six lizard adenoviruses by consensus primer PCR provides further evidence of a reptilian origin for the atadenoviruses. *J Virol* **78**, 13366-13369.

Wozniak, E.-J., DeNardo, D.-F., Brewer, A., Wong, V. & Tarara, R.-P. (2000). Identification of adenovirus and dependovirus-like agents in an outbreak of fatal gastroenteritis in captive born California mountain kingsnakes, Lampropeltis zonata multicincta. *J Herpet Med Surg* **10**, 4-7.

Wu, P., Xiao, W., Conlon, T., Hughes, J., Agbandje-McKenna, M., Ferkol, T., Flotte, T. & Muzyczka, N. (2000). Mutational analysis of the adeno-associated virus type 2

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

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

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

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

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

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

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

Figure legends

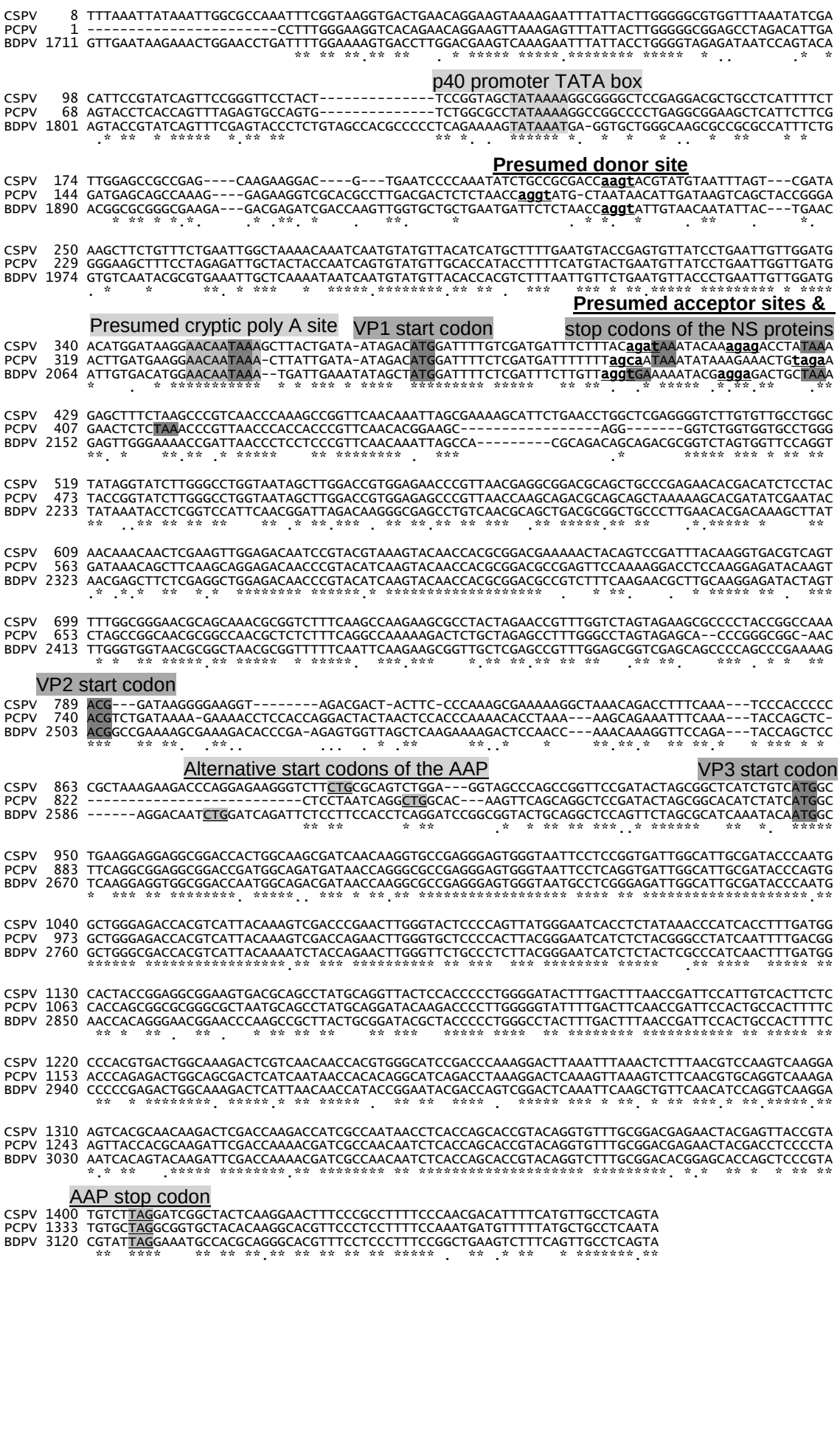
Figure 1 The aligned homologue partial genome sequences of the three novel reptilian parvoviruses; bearded dragon parvovirus (BDPV), corn snake parvovirus (CSPV) and pygmy chameleon parvovirus (PCPV), respectively. The approx. 1.5-kb long sequence of each genome corresponded with the central region including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

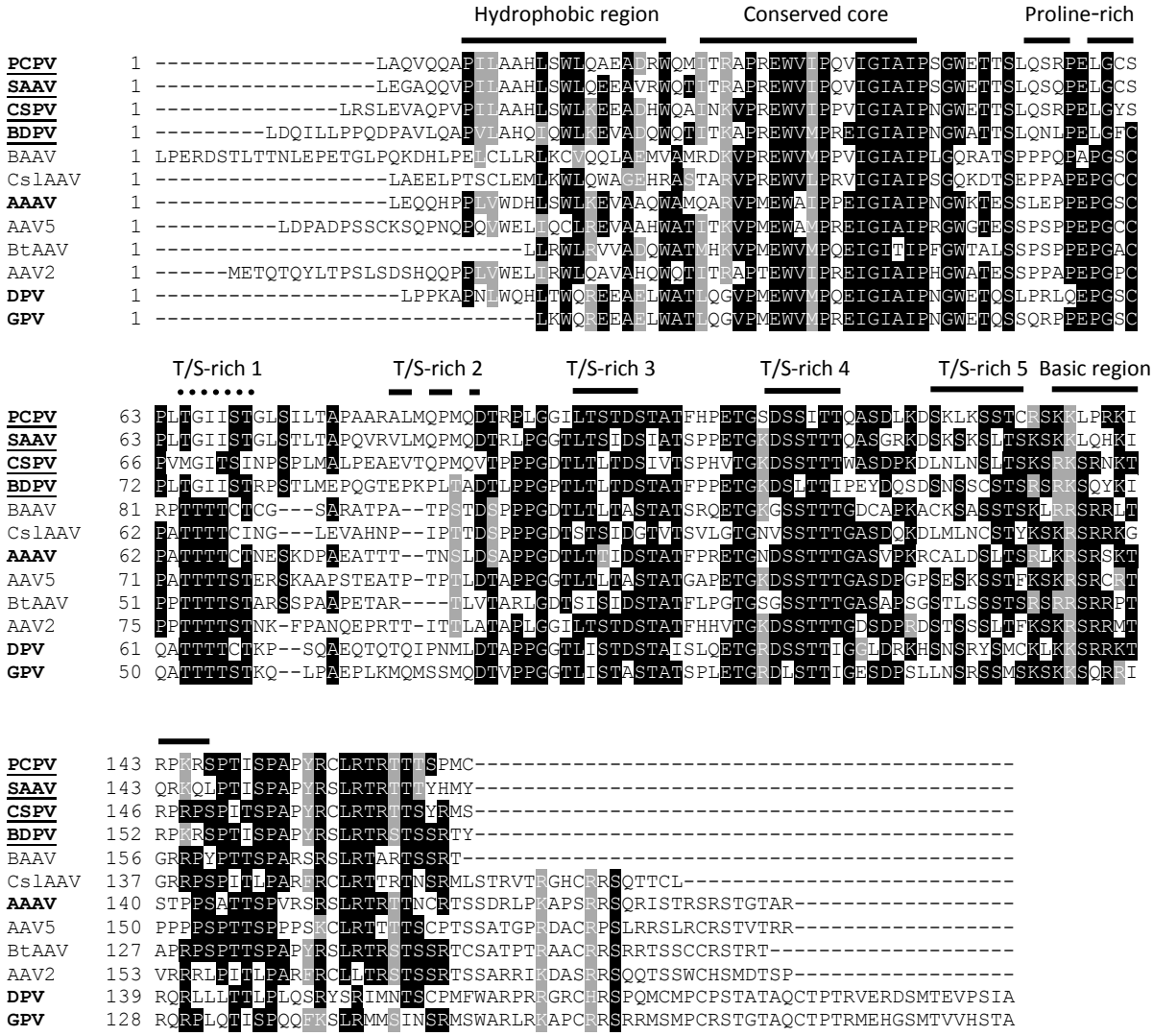
Figure 2 Multiple alignment based on the AAP aa sequences of dependoparvoviruses, with each species represented by at least one type. Viruses of diapsid origin are highlighted in bold, while those of reptilian origin are underlined. Conserved regions of the protein are marked by horizontal lines. Continuous lines represent motifs that are preserved throughout the whole genus, dotted lines stand for those that are absent or highly reduced in reptilian parvoviruses. Dashed lines mark regions that are completely absent from all PVs of diapsid origin. The broken line stands for motifs absent in reptilian, but present in a reduced version some avian dependoparvoviruses. Abbreviations: AAV - adeno-associated virus, AAAV - avian adeno-associated virus, BAAV - bovine adeno-associated virus, BtAAV - bat adeno-associated virus, BDPV - bearded dragon parvovirus, CSPV - corn snake parvovirus, CslAAV - California sea lion adeno-associated virus, DPV - duck parvovirus, GPV - goose parvovirus, PCPV - pygmy chameleon parvovirus.

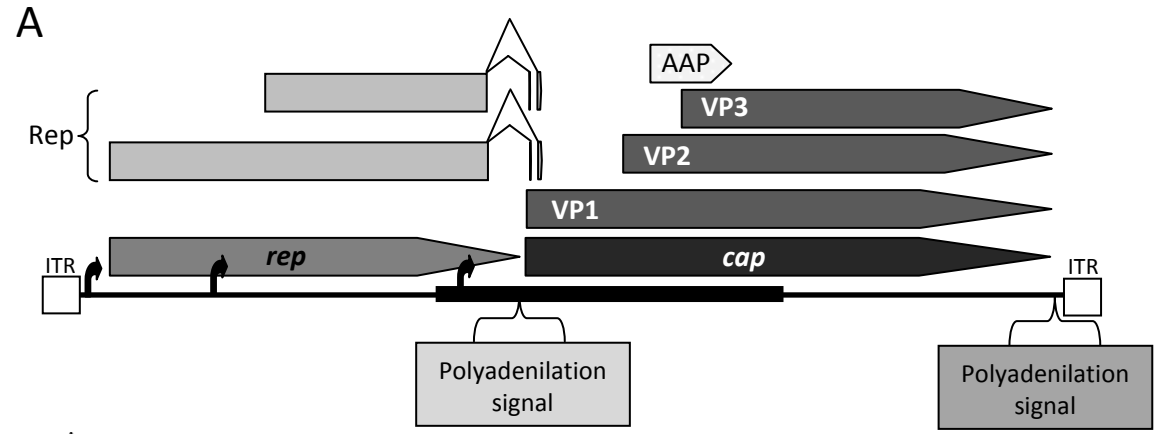
Figure 3 Organization of the complete genome of the bearded dragon parvovirus (A) and the secondary hairpin-like structure of its left ITR (B). The dark, thick arrows represent the two main ORFs (*rep, cap*) and the only alternative ORF (coding for assembly-activating protein, AAP) is presented in white. The further arrows represent the presumed transcripts of the *rep* and *cap* genes, respectively. Black arrows indicate the positions of the three promoters, while the white boxes stand for the inverted terminal repeats (ITRs). The positions of the putative polyadenilation signals are indicated with gray boxes. The darker the box, the higher is the score supporting the signal according to *in silico* predictions. The total length of the genome is 4590 nt out of which 257 corresponds to each ITR. The secondary structure of the telomeric hairpins represented 130 bases. The side-arms of the telomeric T-structure occurred in two alternative orientations, “flip” and its reverse-complement “flop”.

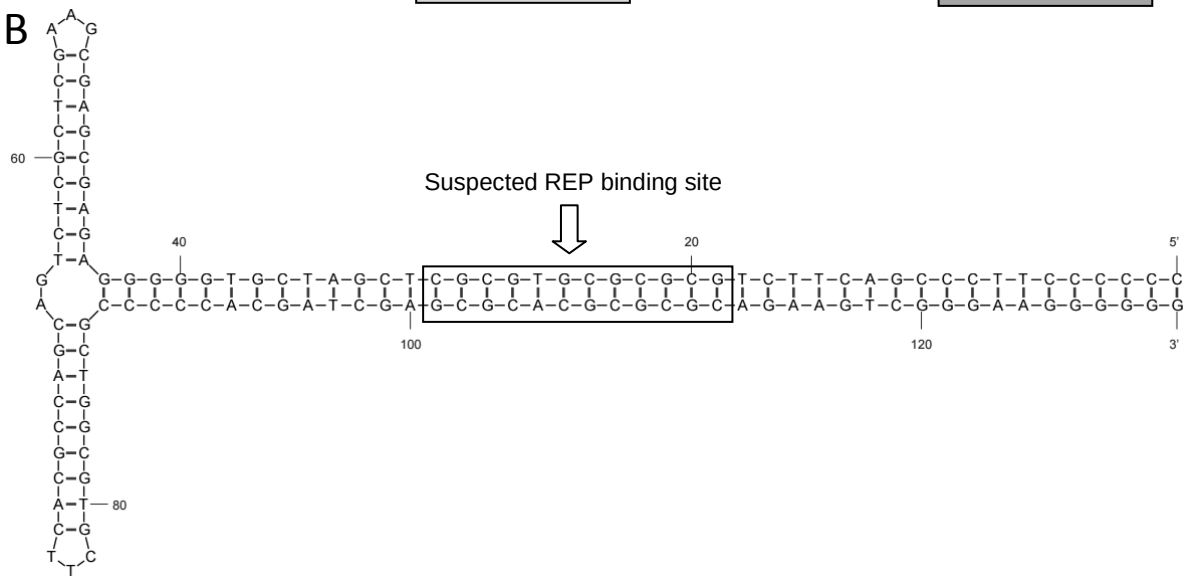
Figure 4 Results of phylogeny reconstructions. Diapsid dependoparvoviruses are highlighted in bold, and the novel squamate parvoviruses are underlined. All squamate parvoviruses can be included in the *Dependoparvovirus* genus. The calculations based on amino acid (aa) sequences of the VP protein fragment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

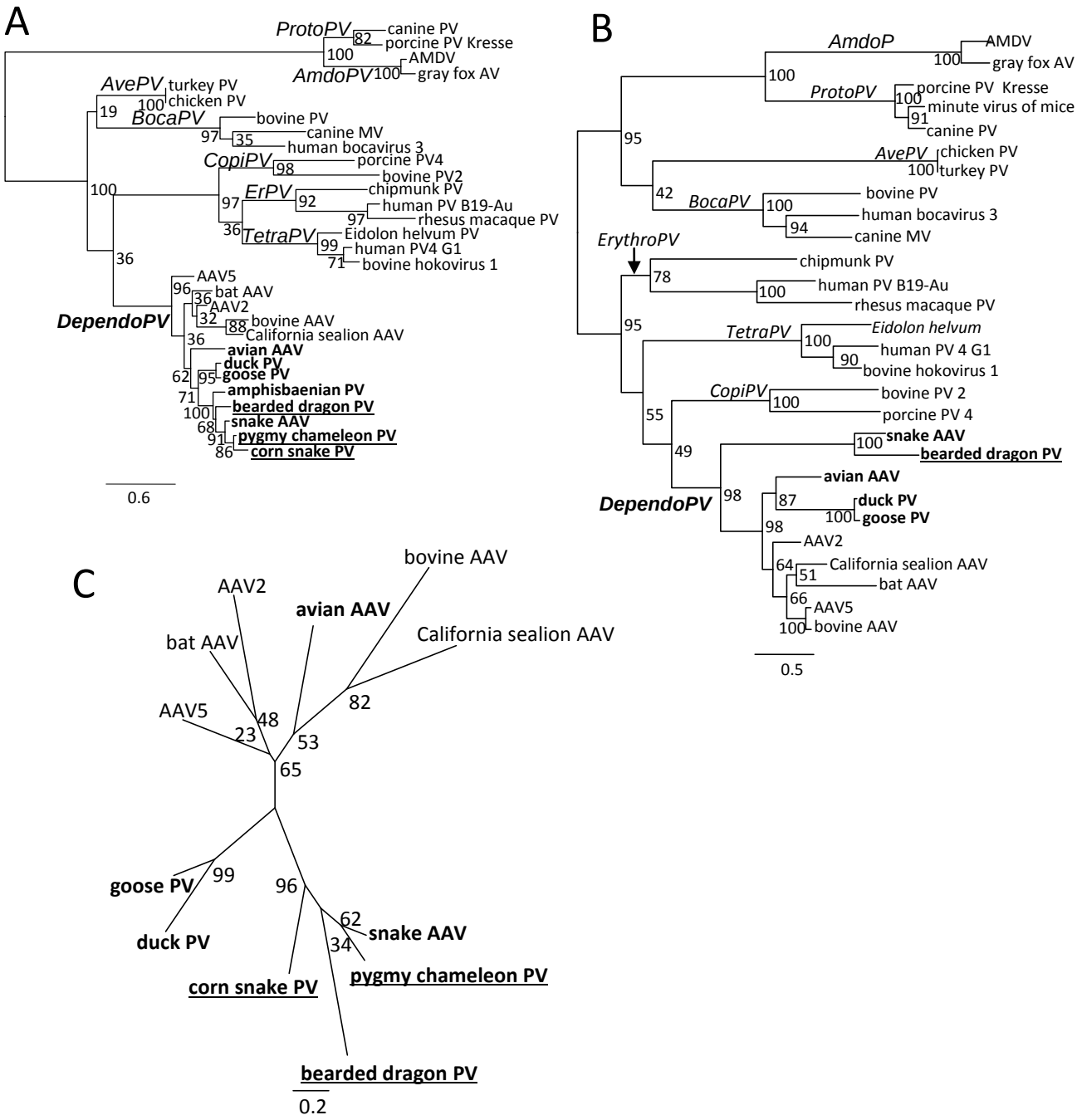
Table 1 Intron lengths of squamate (dark gray background), avian (light gray background), and mammalian (white background) parvoviruses. A general expanding tendency can be observed especially in the length of the first introns. Abbreviations: AAV - adeno-associated virus, PV - parvovirus.


p40 promoter TATA box
Presumed donor site
Presumed cryptic poly A site
VP1 start codon
Presumed acceptor sites &
stop codons of the NS proteins
VP2 start codon
Alternative start codons of the AAP
VP3 start codon
AAP stop codon


Hydrophobic region
Conserved core
Proline-rich
T/S-rich 1
T/S-rich 2
T/S-rich 3
T/S-rich 4
T/S-rich 5
Basic region
PCPV 1 ----------LAQVQQAPTLAAHLSWLQAEADRWQMTTRAPREWVIPQVIGIAIPSGWETTSLQSRPELGCS
SAAV 1 ----------LEGAQQVPILAAHLSWLQEEAVRWQTTTRAPREWVIPQVIGIAIPSGWETTSLQSQPELGCS
CSPV 1 ----------LRSLEVAQPVPILAAHLSWLKEEADHWQAINKVPREWVIPPVIGIAIPNGWETTSLQSRPELGYS
BDPV 1 ---------LDQILLPPQDPAVLQAPYLAHQIQWLKEVADQWQTTTKAPREWVMPREIGIAIPNGWATTSLQNLPELGFC
BAAV 1 LPERDSTLTTNLEPETGLPQKDHLPELCLLRLKCWQQLAEMVAMRDKVPREWVMPPVIGIAIPLGQRATSPPPQPAPGSC
CslAAV 1 ----------LAEELPTSCLEMLKWLQWAGEHRASTARVPREWVLPRVIGIAIPSGQKDTSEPPAPEPGCC
AAAV 1 ----------LEQQHPPLVWDHLSWLKEVAAQWAMQARVPMEWAIPPEIGIAIPNGWKTESSLEPPEPGSC
AAV5 1 ----------LDPADPSSCKSQPNQPQVWELIQCLREVAAHWATTTKVPMEWAMPREIGIAIPRGWGTESSPSPPEPGCC
BtAAV 1 ----------LLRWLFVVADQWATVHKVPMEWVMPQEIGITIPFGWTALSSPSPPEPGAC
AAV2 1 ------METQTQYLTPSLSDSHQQPPLVWELIRWLQAVAHQWQTTTKAPTEWVIPREIGIAIPHGWATESSPPAPEPGPC
DPV 1 ----------LPPKAPNLWQHLTWQREEADLWATLQGVPMEWVMPQEIGIAIPNGWETQSLPRLQEPGSC
GPV 1 ----------LKWQREEAELWATLQGVPMEWVMPREIGIAIPNGWETQSSQRPEPEGSC
PCPV 63 PLTGIISTGLSILTAPAARALMQPMQDTRPLGGILTSTDSTATFHPETGSDSSITTQASDLKDSKLKSSTCHSKKLPRKI
SAAV 63 PLTGIISTGLSTLTAPQVRVLMQPMQDTRLPGGTLTSIDSIATSPPETGKDSSTTTQASGRKDSKSKSLTSKSKKLQHKI
CSPV 66 PVMGITSINPSPLMALPEAEVTQPMQVTPPPGDTLTLTDSIVTSPHVTGKDSSTTTWASDPKDLNLNSLTSKSRKSRNKT
BDPV 72 PLTGIISTRPSTLMEPQGTEPKPLTADTLPPGPTLTLTDSTATFPPETGKDSLTTIPEYDQSDSNSSCSTSPSRKSQYKI
BAAV 81 RPTTTTTCTCG---SARATPA--TPSTDSPPPGDTLTLTASTATSRQETGKGSSTTTGDCAPKACKSASSTSKLRRSRRLT
CslAAV 62 PATTTTCING---LEVAHNP--IPITDSPPPGDTSTSIDGTVTSVLGTGNVSSTTTGASDQKDLMLNCSTYKSKRSRRKG
AAAV 62 PATTTTCTNESKDPAEATTT--TNSLDSAPPGDTLTTIDSTATFPRETGNDSSTTTGASVPKRCALDSSLTSPLKRSRSKT
AAV5 71 PATTTTSTERSKAAPSTEATP-TPILDTAPPGGTLTLTASTATGAPETGKDSSTTTGASDPGPSESKSSTPKSKRSRCRT
BtAAV 51 PPTTTTSTARSSPAAPETAR----TLVTARLGDTSISIDSTATFLPGTGSGSSTTTGASAPSGSTLSSSTSPSRRSRRPT
AAV2 75 PPTTTTSTNK-FPANQEPRTT-ITTLATAPLGGILTSTDSTATFHHVTGKDSSTTTGDSDPRDSTSSSLTPKSKRSRRMT
DPV 61 QATTTTCTKP--SQAEQTQTQIPNMLDTAPPGGTLTSTDSTAISLQETGRDSSTTIGLDRKHSNSRYSMCKLKKSRRKT
GPV 50 QATTTTSTKQ--LPAEPLKMQMSSMQDTVPPGGTLISTASTATSPLETGRDLSTTIGESDPSLLNSRSSMSKSKKSQRRI
PCPV 143 RPKRSPTISPAPVRCLRTRTTTSPMC----------
SAAV 143 QRKQLPTISPAPVRSLRTRTTTYHMY----------
CSPV 146 RPRPSPITSPAPVRCLRTRTTSYRMS----------
BDPV 152 RPKRSPTISPAPVRSLRTRSTSSRTY----------
BAAV 156 GRRPYPTTSPARSRSLRTARTSSRT----------
CslAAV 137 GRRPSPITLPARPRCLRTRTRTNSRMLSTRVTRGHCRRSQTTCL----------
AAAV 140 STPPSATTSPVRSRSLRTRTTTNCRTSSDRLPKAPSRRSQRISTRSRSTGTAR----------
AAV5 150 PPPPSPTTSPPPSKCLRTTTTSCPTSSATGPRDACRPSLRRSLRCRSTVTRR----------
BtAAV 127 APRPSPTTSPAPVRSLRTRSTSSRTCSATPTRAACRRSRRTSSCCRSTRT----------
AAV2 153 VRRRLPITLPARPRCLLTRSTSSRTSSARRIKDASCRRSQQTSSWCHSMDTSP----------
DPV 139 RQRLLLTTLPLQSRYSRIMNTSCPMFWARPRRGRCCRRSPQMCMPCPSTATAQCTPTRVERDSMTEVPSIA
GPV 128 RQRPLQTISPQQPKSLRMMSINSRMSWARLRKAPCCRRSRRMSMPCRSTGTAQCTPTRMEHGSMTVVHSTA


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
porcine PV Kresse
AMDV
gray fox AV
AmdoPV
AvePV
turkey PV
chicken PV
BocaPV
bovine PV
canine MV
human bocavirus 3
CopiPV
porcine PV4
bovine PV2
ErPV
chipmunk PV
human PV B19-Au
rhesus macaque PV
TetraPV
Eidolon helvum PV
human PV4 G1
bovine hokovirus 1
DependoPV
AAV5
bat AAV
AAV2
bovine AAV
California sealion AAV
avian AAV
duck PV
goose PV
amphisbaenian PV
bearded dragon PV
snake AAV
pygmy chameleon PV
corn snake PV
0.6
B
AmdoP
AMDV
gray fox AV
ProtoPV
porcine PV Kresse
minute virus of mice
canine PV
AvePV
chicken PV
turkey PV
BocaPV
bovine PV
human bocavirus 3
canine MV
ErythroPV
chipmunk PV
human PV B19-Au
rhesus macaque PV
TetraPV
Eidolon helvum
human PV 4 G1
bovine hokovirus 1
CopiPV
bovine PV 2
porcine PV 4
snake AAV
bearded dragon PV
DependoPV
avian AAV
duck PV
goose PV
AAV2
California sealion AAV
bat AAV
AAV5
bovine AAV
0.5
C
bovine AAV
AAV2
avian AAV
bat AAV
California sealion AAV
AAV5
goose PV
duck PV
snake AAV
corn snake PV
pygmy chameleon PV
bearded dragon PV
0.2