Title: Endogenous parvoviral elements in the pit viper (*Protobothrops mucrosuamatus*) and in three members of three mammalian orders: implications for ecology and evolution of genera *Amdoparvovirus* and *Protoparvovirus*.

Authors: Judit J Pénzes[1*], Soledad Marsile-Medun[2,3], Mavis Agbandje-McKenna[1], and Robert James Gifford[3*]

Affiliations:

1. University of Florida McKnight Brain Institute, 1149 Newell Dr, Gainesville, FL 32610, USA
2. Agrocampus Ouest, 65 Rue de Saint-Brieuc, 35000 Rennes, France
3. MRC-University of Glasgow Centre for Virus Research, 464 Bearsden Road, Glasgow, UK

***Corresponding author emails**:

judit.penzes@ufl.edu

robert.gifford@glasgow.ac.uk

ABSTRACT

Amdoparvoviruses (family *Parvoviridae:* genus *Amdoparvovirus*) infect carnivores, and are a major cause of morbidity and mortality in farmed mink. Relatively little is known about amdoparvovirus evolution, partly because so few endogenous parvoviral elements (PVe) derived from amdoparvovirus-like viruses have been identified. In this study, we systematically screened animal genomes to identify PVe disclosing a high degree of similarity to amdoparvoviruses, and investigated their genomic, phylogenetic and protein structural features. We report the first full-length, amdoparvovirus-derived PVe in the genome of the Transcaucasian mole vole (*Ellobius lutescens*). Furthermore, we identify four further PVe in mammal and reptile genomes that are intermediate between amdoparvoviruses and protoparvoviruses (genus *Protoparvovirus*) in terms of their phylogenetic placement and genomic features. In particular, we identified a genome-length PVe in the genome of a pit viper (*Protobothrops mucrosquamatus*) that is protoparvovirus-like in terms of its phylogenetic placement and the structural features of its capsid protein (as revealed by homology modeling),but exhibits characteristically amdoparvovirus-like features including (i) a putative 'middle ORF' gene, and (ii) the lack of the phospholipase A2 (PLA2) domain as well as (iii) the putative transcription of VP1. These findings indicate that either: (i) amdoparvoviruses evolved from protoparvoviruses via a series of transitional forms, or; (ii) there are as yet uncharacterised parvovirus lineages that possess a mixture of proto- and amdoparvovirus-like characteristics. Our investigation also provides evidence that amdoparvovirus host range has extended to rodents in the past, and that reptilian parvoviruses exist outside of genus *Dependoparvovirus*. Finally, we show that PVe in the mole vole and pit viper encode intact, expressible replicase genes, adding to a growing body of evidence that these genes have repeatedly been co-opted or exapted in vertebrate genomes.

BACKGROUND

The virus family *Parvoviridae* comprises small, single-stranded DNA viruses that infect vertebrate (subfamily *Parvovirinae*) and invertebrate (subfamily *Densovirinae*) hosts. The small (4-6 kb) genome is encompassed by characteristic palindromic repeats, which form hairpin-like secondary structures characteristic for each genus [1]. Despite exhibiting a low level of sequence homology, parvovirus genomes are highly conserved in overall structure, containing two large gene cassettes responsible for encoding the non-structural (NS) and the structural (VP) proteins. The N-terminal region of the minor capsid protein VP1 includes a highly-conserved phospholypase A2 (PLA2) motif, which is considered to be essential to escape from the endosomal compartments after entering the host cell [2]. The parvovirus capsid is icosahedral with a T=1 symmetry, displaying a jelly roll fold of conserved β-sheets linked by variable surface loops, designated variable region (VR) I to IX [3].

Amdoparvovirus is a newly defined parvoviral genus in the family *Parvoviridae* [4]. The type species - Aleutian mink disease virus (AMDV) - causes an immune complex-associated progressive syndrome in mink (Family *Mustelidae*: genera *Neovison* and *Mustela*) called Aleutian disease or plasmacytosis. First described in 1956, Aleutian disease is presently considered to be one of the most important infectious diseases affecting farm-raised mink [5, 6]. AMDV infection is known to be widespread in wild mink as well as in farmed animals [5], and amdoparvoviruses have been identified in other carnivore species, including raccoon dogs [7], foxes and skunks [8-11]. However, relatively little is known about the biology of amdoparvovirus infection in the natural environment. For example, it is unclear whether amdoparvoviruses cause disease in wild animals, or whether the host range of the genus extends beyond carnivores.

Furthermore, relatively little is known about the long-term evolutionary history of amdoparvoviruses, as compared to some other parvovirus genera. In particular, the ancient origins of the *Dependoparvovirus* and *Protoparvovirus* genera has been demonstrated through the identification of endogenous parvoviral elements (PVe) at orthologous loci in distantly related mammalian species [12-15]. PVe are sequences homologous to parvovirus genomes that occur sporadically in eukaryotic genomes, and are presumed to derive from ancient parvovirus genomes via a form of horizontal gene transfer in which DNA sequences derived from viruses are incorporated into the germline of the ancestral host species (presumably via infection of germline cells) such that they are inherited as host alleles [7]. PVe provide a rare and useful source of retrospective

information about the biology and evolution of ancient parvoviruses [7, 16]. So far, however, only a single, highly fragmented element has been described that showed homology to amdoparvoviruses, and this provided relatively limited insight [14].

Among other genera within the *Parvovirinae*, amdoparvoviruses are most closely related to protoparvoviruses (genus *Protoparvovirus*). This genus includes viruses that infect a wide range of mammalian hosts, encompassing several mammalian orders, namely rodents, carnivores, chiropterans, ungulates, shrews and primates [17, 18]. Rodent protoparvoviruses are known for their oncolytic properties, whereas carnivore and ungulate protoparvoviruses are significant pathogens of domestic pets, fur industry and livestock [19-21]. While the *Protoparvovirus* and *Amdoparvovirus* genera form sister taxa in phylogenetic trees, the two groups are distinguished by certain features of their genomes and replication strategies. Amdoparvovirus mRNAs are transcribed from one single upstream promoter and are polyadenylated at two polyadenylation signals. To provide the VP1 encoding transcript, an intron is spliced out, leaving a short, three amino acid-encoding exon leader sequence [22]. Genus *Amdoparvovirus* is unique within the *Parvovirinae,* in that all representatives lack a PLA2 domain in their VP unique region (VP1u). Consequently, their trafficking is not fully understood [2, 4].

In this study, we performed a systematic screen of 688 animal genomes to identify PVe disclosing a high degree to amdoparvoviruses. We characterise the genomic, phylogenetic and protein structural characteristics of six such amdoparvovirus-like PVe, revealing new information about the biology and evolution of the *Amdoparvovirus* genus.

METHODS

Genome screening in silico

Vertebrate genome assemblies were obtained from NCBI genomes (**Table S1**). Screening was performed using the database-integrated genome-screening tool (available from http://giffordlabcvr.github.io/DIGS-tool/). The DIGS procedure used to identify PVe comprised two steps. In the first, a parvovirus protein sequence (e.g. Rep or Cap) was used to search a particular genome assembly file using the tBLASTn program [23]. In the second, sequences that produce statistically significant matches to the probe are extracted and classified by tBLASTn-based comparison to a set of virus reference genomes (**Table S2**), this time using the BLASTx program [23]. Results are captured in a MySQL database.

We applied a systematic approach to naming PVe. Each element was assigned a unique identifier (ID) constructed from two components. The first component is the classifier 'PVe'. The second component is itself a composite of two distinct subcomponents separated by a period; (i) the name of the lowest level taxonomic group (i.e. species, genus, subfamily, or other clade) into which the element can be confidently placed by phylogenetic analysis; (ii) a numeric ID that uniquely identifies the insertion (for cases where multiple, closely-related PVe occur in the same species genome).

Sequence analysis

ORFs were inferred by manual comparison of putative peptide sequences to those of closely related exogenous parvoviruses. The characterization and annotation of the revealed endogenous sequences was executed using Artemis Genome Browser [24]. The putative peptide sequences of *Amdoparvovirus*-related EVEs were aligned with NS and VP sequences of representative amdoparvoviruses and protoparvoviruses using MUSCLE [25] PAL2NAL [26] and T-coffee Expresso [27]. Phylogenies were reconstructed from amino acid (aa) alignments incorporating structural data (at least one high-resolution structure from all available genera) using maximum likelihood as implemented in PhyML-3.1 [28]. Model selection was carried out by ProTest [25] selecting the RtEV (NS) and the LG (VP) protein substitution models. Testing the reliability of the topology was carried out by bootstrapping of 100 repeats.

To detect structural homology, we applied the pGenTHREADER and pDomTHREADER algorithms of the PSIPRED Protein Sequence Analysis Workbench [29]. The selected PDB structures were applied as templates for homology modeling, carried out by SWISS-MODEL [30]. Polymers of the acquired capsid monomer models were constructed by the Oligomer Generator feature of the Viper web database (http://viperdb.scripps.edu/) [31]. The generated polymers were rendered as well as ribbon diagrams compared using PYMOL [32] .

RESULTS

Identification and characterisation of amdoparvovirus-like PVe

We screened whole genome sequence (WGS) assemblies of 688 animal species (**Table S1**) for PVe disclosing a high degree of homology to amdoparvoviruses. Similarity searches using the replicase (NS) and capsid (VP) proteins of AMDV identified six such PVe (**Table 1**). These sequences were revealed in five species genomes,

including one reptile – the pit viper (*Protobothrops mucrosquamatus*) - in addition to four mammals including three placental species (one rodent and two afrotherians), and one marsupial species: the Tasmanian devil (*Sarcophilus harrisii*).

In all cases, genomic regions exhibiting similarity to amdoparvoviruses were present within large contigs, and there was little question they represented PVe as opposed to contaminating virus. Further investigation of these loci revealed additional PVe sequences upstream and downstream of the amdoparvovirus-like regions initially identified via DIGS. We generated an overview of the six PVe loci, delineating the locations of promoters, polyadenylation signals, parvovirus genome features, and transposable element insertions (**Figure 1**). Their relative length and location of the preserved fragments compared to the exogenous amdoparvovirus genome is shown in **Figure 2**. Analysis of genomic sequences flanking these PVe established that all six were generated in distinct germline incorporation events. To infer the evolutionary relationships of these PVe to contemporary proto- and amdoparvoviruses we reconstructed maximum likelihood (ML) phylogenies (**Figure 3**) using conserved regions of the putative NS and VP peptide sequences contained within these PVe. The genomic and phylogenetic characteristics of all six elements are described below.

Amdoparvovirus-like PVe in the mole vole germline

We identified two *Amdoparvovirus*-derived sequences in the genome of the Transcaucasian mole vole (*Ellobius lutescens*). The first of these elements spanned a near complete genome containing both the NS and VP genes, while the second spanned the majority of the NS gene, with no identifiable VP present (**Figure 2**). Following the systematic approach described in the methods section, these elements were given the identifiers PVe-Amdo-EllLut.1 and PVe-Amdo-EllLut.2 respectively (**Table 1**). In the interests of brevity, we refer to these elements as EllLut.1 and EllLut.2 in subsequent text.

Amdoparvovirus genomes encode a short middle ORF (M-ORF) of unknown function between the two major open reading frames (ORFs), NS and VP. As shown in **Figure 1**, a region of potentially protein-coding sequence that corresponds to the M-ORF of AMDV is present in the first of these elements. A methionine (M) residue that might represent the start codon of an M-ORF gene product could not be identified. However, this is also the case for several exogenous amdoparvovirus isolates.

The putative NS ORF of EllLut.1 has gaps relative to AMDV - in fact, the complete N-terminal region is absent up to residue 28G of the protein encoded by AMDV NS gene. The gap in NS is apparently due to a deletion (since the corresponding region is present in EllLut.2). A partial VP ORF could be identified downstream of the NS gene, corresponding to the VP1u and the VP2 N-terminal, as well as nucleotides encoding the last 173aa of the C-terminus. Interestingly, the complete missing fragment could be identified downstream of the NS ORF, encompassed by LINE and SINE elements. None of these VP homologues contained a PLA2 motif. The VP1u encoding sequence included a possible intron, hence displaying the typical VP1u transcription pattern of amdoparvoviruses [22]. The putative VP ORF encoded by EllLut.1 has a gap relative to the AMDV VP that spans most of the 5' region of the gene, presumably due to these sequences having been deleted. Frameshifting mutations are present in both the NS and one in the VP pseudogenes of EllLut.1. The element is integrated into a locus that is homologous to mouse chromosome 12.

The EllLut.2 element comprises the NS gene alone (**Figures 1 and 2**). This element is integrated into a locus immediately adjacent to the sequences encoding the MAF BZIP transcription factor G (MAFG) gene, which in the mouse genome is located in the 11qE2 region of chromosome 11. The otherwise intact NS gene lacks the methionine start codon, however, right upstream the three stop codons disrupting the MAFG gene, a conventional ATG start codon could provide translation initiation to express a MAFG-NS fusion product (**Figure 1**). The identification of a potential promoter sequence downstream of the MAFG gene supports the existence of such a fusion protein.

Phylogenies on both VP and NS aa sequences showed that both mole vole elements were relatively closely related and grouped robustly within the clade defined by exogenous amdoparvoviruses (**Figure 3**). We identified empty integration sites in the *E. talpinus* genome at the loci where the EllLut.1 and EllLut.2 elements are integrated in *E.lutescens*. This indicates that both elements were integrated into the *E.lutescens* germline after these two species diverged ~10 million years ago (MYA) [33, 34]. It seems unlikely, however, that either element has been integrated in very recent times. Firstly, the degraded nature of EllLut.1 indicates that it has been resident in the germline for some time. Furthermore, the genomes of two *E.lutescens* individuals have been generated (genomic DNA was obtained from the livers of both a male and a female individual). Both PVe were present in both individuals. The EllLut.2 element in the female animal had a 13-14 bp deletion relative to the one in the male.

The identification of an empty integration in *E. talpinus* facilitated the identification of genomic flanking regions for both PVe. In the case of EllLut.2, genomic flanks occur close to the coding region of NS gene, suggesting the element was derived from an mRNA that was reverse transcribed and integrated into the nuclear genome of an ancestral germline cell. By contrast, EllLut.1 appears to be derived from genome-length nucleic acid, although the partial, otherwise non-disrupted downstream VP ORF might have gone through the same integration process as the NS gene of EllLut.2. In addition to containing regions of both NS and VP, this element includes a region of 3' sequence between the end of VP and the beginning of the genomic flanking sequence that exhibits similarity to the 3' untranslated region of AMDV and contains inverted repeats capable of folding into a stem loop structure (data not shown). We could not, however, identify sequences corresponding to the 5' UTR in the EllLut.1 element. No methionine start codon could be identified at the start of the NS pseudogene encoded by this element.

Amdoparvovirus-like PVe in the pit viper germline

We identified a PVe in the genome of the pit viper (*Protobothrops mucrosquamatus*) genomes that encoded a nearly complete parvovirus genome. The entire PVe sequence was ~4.5 kb in length and was integrated into the complementary strand of the *Probothrops* genome. It comprised two major ORFs and a minor ORF as well as a clearly-identifiable and potentially functional downstream promoter. Furthermore, two polyadenylation signals could be identified. The pit viper PVe was flanked by partial palindromic repeats, resembling the amdoparvoviral hairpin structures (**Figure 4**).

The first major ORF exhibited a relatively high degree of aa identity to the AMDV NS protein (35% with no deletions). The second major ORF, which was disrupted by nonsense mutations (two stop codons, two frameshifts), was homologous to amdoparvovirus VP encoding genes (36% aa identity with skunk amdoparvovirus VP). This ORF did not possess a conventional Met start codon to express VP1, however, the three aa-long exon leader, revealed upstream, is suspected to provide this, like in case of amdoparvoviruses. Aligned with the EllLut.1and exogenous amdoparvovirus VP aa sequences, deletions of various lengths could be observed, limited almost exclusively to the variable regions (**Figure 5a**). These VRs form surface loops responsible for host-virus interactions, such as immunogenicity and receptor attachment. The only insertion, six-aa-long, was present in VR VIII (**Figure 5a**). In contrast, the EllLut.1 VP displayed

very similar organization to the exogenous amdoparvovirus VPs. Despite of lacking the PLA2 domain, in all exogenous amdoparvovirus VP aa sequences a polyglycine (poly-G) stretch could be revealed, which is suspected to be responsible for externalizing the VP1u, so the PLA2 enzymatic function can be carried out as well as revealing the nuclear localization signal (NLS) [35]. The poly-G was present in the predicted VP sequence, in contrast with the EllLut.1 VP, from which it was absent (**Figure 6**). Interestingly, both PVe VP1u displayed the putative NLS.

The NS gene of the pit viper PVe clustered as an outgroup to a clade containing the mole vole PVe and exogenous amdoparvoviruses (**Figure 3a**). However, this relationship was not strongly supported by bootstrap values. The pit viper PVe locus occurs on a contig that has not been mapped to a specific chromosome. Nevertheless, the pre-integration locus could be identified in WGS data of two other reptilian species: the Burmese python (*Python bivittatus*), and a colubrid, the common garter snake (*Tamnophis sirtalis*). Both genomes contain the LINE transposon of the RTE-BovB class [36] but lack any sequences of parvoviral homologues.

A small ORF was identified between the putative NS and VP genes. The similar position ORF to ORF-M of amdoparvoviruses would suggest this ORF could represent a highly divergent form of the amdoparvovirus ORF-M gene.

Amdoparvovirus-like PVe in marsupial and afrotherian germlines

We identified three additional short, fragmentary matches to amdoparvoviruses in mammalian genomes (**Table 1**). One of these, identified in the Cape hyrax (*Procavia capensis*), has been reported previously [14]. We identified two additional, similar length matches in the Tasmanian devil (*Sarcophilus harrisii*) and aardvark (*Orycteropus afer*) genomes, both of which turned out to contain complete or near complete VP genes (**Figure 1 and 2**). As all three of these sequences were heavily disrupted by stop codons, frameshifts as well as retrotransposable elements, their former coding sequences were present as small fragments. With the exception of the aardvark EVE, which harbored a partial, highly disrupted NS homologue of 343 aa, only minimal trace of the former replication protein encoding genes could be detected (**Figure 1**). At the N-terminal of their putative VP1 sequences a well-preserved calcium-binding loop of the PLA2 domain could be revealed, however, the catalytic core was barely recognizable in the Cape hyrax PVe and completely absent in the aardvark PVe (**Figure 6**). Similarly to

the EllLut.1 VP, the poly-G stretch was absent in all three cases, although only the Cape hyrax PVe lacked the NLS.

In phylogenies based on NS (**Figure 3a**), PVe from the aardvark grouped together with the Mpulungu bufavirus of shrews [18], as a robustly-supported sister group to rodent, ungulate and carnivore protoparvoviruses. VP-based phylogenies unite PVe from the pit viper, Tasmanian devil, hyrax and aarvark in a sister clade to protoparvoviruses (with the exception of the bufavirus), but with low support. The Tasmanian devil PVe clustering as an outgroup to the afrotherian PVe was well-supported.

Structural characterization of capsid proteins via homology modeling

To investigate what the capsid proteins of the ancient parvoviruses that gave rise to amdoparvovirus-like PVe might have looked like, and to compare their predicted structures to those of their contemporary counterparts, we subjected the capsid sequences of more complete and intact elements to homology modelling. Amdoparvovirus-like PVe in the pit viper and the mole vole proved to be structurally most similar to the canine parvovirus (CPV) capsid (PDB ID: 2CAS) according to fold recognition, hence this structure was used as a template. As there are no structural data available on any members of genus *Amdoparvovirus*, we constructed the model of the AMDV capsid as well, based on the CPV template.

The predicted structures of capsid proteins ancestrally encoded by the pit viper and EllLut.1 PVe displayed a rather protoparvovirus-like appearance, unlike the AMDV capsid model (**Figure 4**). In the case of the pit viper, three-fold protrusions were thicker and bulkier than either on CPV or AMDV, while the AmdoPVe-EllLut capsid model displayed spike-like protrusions instead of the slope-like depressions characteristic of the parvovirus two/fivefold wall. These differences could be ascribed to insertions in variable regions, namely VRVIII of the pit viper PVe and VRVII of EllLut.1. Both capsids appeared to contain the canonical β-strand A (βA), an eight-stranded β-barrel core making up the jelly roll fold (βBIDG-CHEF), and an α-helix (αA) (**Figure 4b and c**).

DISCUSSION

In this study, we screened animal genomes *in silico* to identify PVe disclosing a high degree of similarity to amdoparvoviruses. We identified six such sequences, and used comparative approaches to investigate them. Our analysis revealed new information about the biology of amdoparvoviruses and their evolutionary relationships with protoparvoviruses. We also found evidence from comparative analysis that some of the parvovirus replicase genes we identify may have been co-opted or exapted by the host species genomes in which they occur.

The genomic fossil record of amdoparvovirus-like parvoviruses

The most conspicuously amdoparvovirus-like PVe were identified in the genome of the Transcaucasian mole vole. These two elements, which share ~80% identity at the amino acid level, were found to group robustly within the amdoparvovirus clade in ML phylogenies (**Figure 3**). They also exhibit characteristic features that support their grouping within the genus *Amdoparvovirus*, including the absence of the PLA2 domain from the predicted VP protein sequence as well as the presence and location of the intron in their VP1u-encoding gene (**Figures 1 and 6**). No other PVe were identified that grouped within the *Amdoparvovirus* genus. However, we identified several that displayed a mixture of amdoparvovirus and protoparvovirus features. In the pit viper element, certain aspects of genome organization suggested an amdoparvoviral origin - specifically (i) the presence of a single promoter and two polyadenylation signals, and (ii) the attributes of the intron in the VP1u and the PLA2 absence. In phylogenetic terms, however, these PVe appear to be marginally more closely related to protoparvoviruses than to amdoparvoviruses (**Figure 3**).

Our phylogenies support a common evolutionary origin for; (i) genus Amdoparvovirus; (ii) genus Protoparvovirus; (iii) the recently described 'bufaviruses', which appear to represent highly divergent protoparvoviruses [18, 40]; and (iv) PVe derived from a basal, amdoparvovirus-like protoparvovirus lineage (referred to hereafter as 'AP'). Within the AP lineage, the pit viper PVe consistently groups in a basal position (**Figure 3**), close to the split between the *Amdoparvovirus* and *Protoparvovirus* genera. Furthermore, this PVe harbors the most amdoparvovirus-like characteristics out of the AP stem group. This suggests it might represent an ancestral progenitor of contemporary amdoparvoviruses, potentially implying that genus *Amdoparvovirus* has an origin in lower vertebrate hosts, which might have been reptiles. In this scenario, the

afrotherian and the marsupial PVe of the AP stem group represent transitional forms along a pathway from protoparvovirus-like complete PLA2 domains to the PLA2-absent amdoparvoviral-like VP1u. Thus, the absence of heavy degradation of the catalytic core in these PVe might reflect various stages of PLA2 diminishment during evolution. However, these PVe could also be remnants of an as yet unknown parvovirus lineage, distinct from *Amdoparvovirus*, that contained features of both amdo- and protoparvoviruses and may now be extinct.

Evolution of amdoparvovirus and protoparvovirus capsid proteins

We used homology modelling to infer the structures of the capsid (VP) proteins encoded by the ancestral parvoviruses that gave rise to the mole vole and pit viper PVe. Comparison of predicted VP proteins structures to those of exogenous amdoparvoviruses revealed that most differences are limited to highly variable regions, i.e. VR loops and VP1u. Currently, we cannot be certain these differences reflect changes prior to germline incorporation. However, as the VRs are situated on the virion surface and play an important role in receptor attachment and host immune interactions, this should be amenable to experimental investigation [37]. All known amdoparvoviruses infect carnivore hosts. Since successful infection and replication in a carnivore cell versus a rodent or reptilian cell likely requires very different surface features, we might expect to see a high degree of divergence in these regions. Consistent with this, the number of deletions in the VR regions of the putative pit viper PVe VP protein sequence were found to be relatively high compared to those in the mole vole PVe. Fundamental differences in the reptile anti-viral response (namely the greater reliance on native, aspecific immune processes as opposed to adaptive processes) could account for this [38]. This idea has previously been proposed to explain the smooth surface features of invertebrate-infecting densoviruses [39].

The results of homology modelling should always be interpreted with caution – nevertheless, the more protoparvovirus-like capsid appearance of both PVe is intriguing, as it is consistent with the close evolutionary relationship between the *Amdo*parvovirus and *Protoparvovirus* genera suggested by phylogenies. Possibly, genus amdoparvovirus has changed tropism or receptor specificity during its evolution. In other words, these PVe might share an ancestral tropism or receptor specificty with carnivore protoparvoviruses, in contrast to AMDV, hence their similar appearance.

As in the case of VR regions, we cannot say for certain whether the inferred loss of PLA2 function (see **Figure** 3) occurred before or after integration, however, the otherwise intact, and clearly-recognizable nature of the calcium binding loop suggests it occurred before. Potentially, the loss of the poly-G stretch in capsid (see **Figure 6**) might have occurred in the progenitor virus, and could be related to the loss of PLA2 function. The insertion in the pit viper VP VRVIII is a unique finding of unknown role, however, dependoparvovirus-derived PVe in marsupial genomes have also been found to harbor extended VRVIIs, similar to the VP encoded by AmdoPVe-EllLut [40].

Timescale of amdoparvovirus evolution

Unfortunately, the present data do not allow us to unequivocally demonstrate minimum ages for any of the PVe described here, since we did not identify any orthologous PVe in related species. Nevertheless, the mutational degradation observed in many of these elements suggests they are likely to have similarly ancient origins to other PVe (i.e. extending back many millions of years). Furthermore, in the case of the mole vole, the presence of two EllLut.2 alleles was indicated (one containing a deletion), suggesting that, at the very least, these elements have been present in the species gene pool for multiple generations.

While we cannot be certain how recently in evolutionary history the mole vole and pit viper PVe were generated, we could establish maximum age bounds for these elements, based on the identification of empty integration sites in related species. In the case of the pit viper PVe, identification of an empty orthologous insertion site in colubrid snakes established that incorporation into the viper germline occurred within the last 34-54 Mya [41]. In the case the mole vole, empty insertion sites were identified in the most closely related sister taxon (*E. talpinus*), establishing that the PVe in the *E.lutescens* were incorporated after these species diverged ~10Mya. In reality, the mole vole elements could be much younger than this, but even if we assume they are ~10 MY old, we might still expect to observe the close relationship with contemporary amdoparvoviruses indicated by our phylogenies, despite the rapid rates of evolution observed in the parvoviruses. This is due to the time-dependent bias of viral evolutionary rates, which can vary by several orders of magnitude depending on the timeframe of measurement [42]. In fact, given that the timescale of parvovirus evolution has already been shown to date back many millions of years, we think it is likely that

amdoparvoviruses have similarly ancient origins, and identification of additional PVe will eventually confirm this.

Insights into amdoparvovirus and protoparvovirus host range

Our findings suggest that the apparently restricted host range of amdoparvoviruses (order Carnivora) may simply reflect limited sampling. Recently, a partial amdoparvovirus capsid sequence was identified via metagenomic screening of samples derived from the least horseshoe bat (*Rhinolophus pusillus*), suggesting that bats (order Chiroptera) may also harbour amdoparvoviruses. However, as with all virus sequences recovered via metagenomic sequencing, there remains a degree of uncertainty regarding host associations [43], whereas the identification of very clearly amdoparvovirus-derived PVe in the genome of the Transcaucasian mole vole constitutes unequivocal evidence that amdoparvoviruses have infected rodents in the past.

The pit viper PVe described here is the first to be identified in a reptile genome and provides the first evidence for the existence of reptilian parvoviruses outside of the genus *Dependoparvovirus*. Interestingly, *Dependoparvovirus*, the other genus of *Parvoviridae* known to infect reptiles is also suspected to be of reptilian origin [1].

PVe-encoded replicase genes appear to have been exapted or co-opted

Amongst the various parvovirus-derived EVEs that have been reported so far, several have exhibited such intact or nearly intact replicase genes. [44, 45]. In this report, we identify yet another, relatively intact and independently acquired parvovirus replicase in a rodent genome. Intriguingly, analysis of the locus indicated that this replicase is possibly expressed as a joint protein with the partial MAFG gene product. As MAFG is a transcription factor, this fusion protein might have gained new, beneficial functions as far as the host cell's transcription machinery and gene expression are concerned. Furthermore, the PVe we identify in the pit viper genome also appears to encode an intact, expressible, parvovirus replicase gene, and represents the first non-mammalian example of this phenomenon. Fixation of PVe is extremely unlikely and would be expected to occur over many host generations when it does occur. Thus, the observation that (i) PVe encoding replicase genes have been independently acquired by multiple vertebrate species, and (ii) the capacity of these PVe to express mRNA and protein has apparently been maintained, suggests that these genes are being functionalised and selected for in some as yet undetermined way.

Conclusions

Our analysis of PVe indicates that either; (i) amdoparvoviruses and protoparvoviruses evolved from a common ancestor via a series of transitional forms, or (ii) there are, or have been in the past, parvovirus lineages that possess a mixture of proto- and amdoparvovirus-like characteristics. Our investigation also provides insight into parvovirus host range, suggesting amdoparvoviruses likely infect a broad range of mammalian species, and that a lineage of protoparvovirus-like viruses infects reptiles, and might have given rise to exogenous amdoparvoviruses of today. Finally, we provide further evidence that replicase genes of PVe derived from multiple *Parvovirinae* genera have been co-opted or exapted in vertebrate genomes.

Table 1. Amdoparvovirus-like Pve identified via *in silico* genome screening.

Species in which identified		Genus/Clade	Element ID [a]	Genes present
Common name	*Latin binomial*			
Mole vole	*Ellobius lutescens*	*Amdoparvovirus*	PVe-Amdo-EllLut.1	NS-M-VP
Mole vole	*Ellobius lutescens*	*Amdoparvovirus*	PVe-Amdo-EllLut.2	NS
Pit viper	*Protobothrops mucrosquamatus*	AP	PVe-AP-ProMuc.1	VP
Cape hyrax	*Procavia capensis*	AP	PVe-AP-ProCap.1	M-VP
Tasmanian devil	*Sarcophilus harrisii*	AP	PVe-AP-SarHar.1	VP
Aardvark	*Orycteropus afer*	AP	PVe-AP-OryAfe.1	VP

Footnote: * Genbank accession numbers are shown for genomic scaffolds Abbreviations: NS=nonstructural proteins (replicase); VP=viral capsid; M=middle ORF; AP=amdoparvovirus-like protoparvovirus lineage.

Acknowledgements

RJG was funded by the Medical Research Council of the United Kingdom (MC_UU_12014/12). JJP and MA-K were supported by a grant from the National Institutes of Health (NIH R01 GM109524). We thank Andrew Davison and Joseph Hughes for their comments and feedback on the manuscript.

References

1. Penzes, J.J., et al., *Novel parvoviruses in reptiles and genome sequence of a lizard parvovirus shed light on Dependoparvovirus genus evolution.* J Gen Virol, 2015. **96**(9): p. 2769-79.
2. Zadori, Z., et al., *A viral phospholipase A2 is required for parvovirus infectivity.* Dev Cell, 2001. **1**(2): p. 291-302.
3. Chapman, M.S. and M. Agbandje-McKenna, *Atomic structure of viral particles*, in *Parvoviruses*, J.R. Kerr, et al., Editors. 2006, Hodder Arnold, Ltd;. p. 107–123.
4. Cotmore, S.F., et al., *The family Parvoviridae.* Arch Virol, 2014. **159**(5): p. 1239-47.
5. Canuti, M., H.G. Whitney, and A.S. Lang, *Amdoparvoviruses in small mammals: expanding our understanding of parvovirus diversity, distribution, and pathology.* Front Microbiol, 2015. **6**: p. 1119.
6. Canuti, M., H.G. Whitney, and A.S. Lang, *Corrigendum: Amdoparvoviruses in small mammals: expanding our understanding of parvovirus diversity, distribution, and pathology.* Front Microbiol, 2016. **7**: p. 264.
7. Holmes, E.C., *The evolution of endogenous viral elements.* Cell Host Microbe, 2011. **10**(4): p. 368-77.
8. Pennick, K.E., et al., *Aleutian disease in two domestic striped skunks (Mephitis mephitis).* Vet Pathol, 2007. **44**(5): p. 687-90.
9. Shao, X.Q., et al., *Novel amdoparvovirus infecting farmed raccoon dogs and arctic foxes.* Emerg Infect Dis, 2014. **20**(12): p. 2085-8.
10. LaDouceur, E.E., et al., *Aleutian Disease: An Emerging Disease in Free-Ranging Striped Skunks (Mephitis mephitis) From California.* Vet Pathol, 2015. **52**(6): p. 1250-3.
11. Canuti, M., et al., *Full genetic characterization and epidemiology of a novel amdoparvovirus in striped skunk (Mephitis mephitis).* Emerg Microbes Infect, 2017. **6**(5): p. e30.
12. Belyi, V.A., A.J. Levine, and A.M. Skalka, *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, 2010. **84**(23): p. 12458-62.
13. Kapoor, A., P. Simmonds, and W.I. Lipkin, *Discovery and characterization of mammalian endogenous parvoviruses.* J Virol, 2010. **84**(24): p. 12628-35.
14. Katzourakis, A. and R.J. Gifford, *Endogenous viral elements in animal genomes.* PLoS Genet, 2010. **6**(11): p. e1001191.
15. Liu, H., et al., *Widespread endogenization of densoviruses and parvoviruses in animal and human genomes.* J Virol, 2011. **85**(19): p. 9863-76.
16. Feschotte, C. and C. Gilbert, *Endogenous viruses: insights into viral evolution and impact on host biology.* Nat Rev Genet, 2012. **13**(4): p. 283-96.
17. Tijssen, P., et al., *Family Parvoviridae*, in *Virus taxonomy—Ninth Report of the International Committee on Taxonomy of Viruses.*, A.M.Q. King, et al., Editors. 2011, Elsevier/Academic Press: London. p. 405–425.
18. Sasaki, M., et al., *Distinct Lineages of Bufavirus in Wild Shrews and Nonhuman Primates.* Emerg Infect Dis, 2015. **21**(7): p. 1230-3.

19. Wolf, V.H., et al., *Molecular basis for porcine parvovirus detection in dead fetuses.* Genet Mol Res, 2008. **7**(2): p. 509-17.
20. Allison, A.B., et al., *Host-specific parvovirus evolution in nature is recapitulated by in vitro adaptation to different carnivore species.* PLoS Pathog, 2014. **10**(11): p. e1004475.
21. Marchini, A., et al., *Oncolytic parvoviruses: from basic virology to clinical applications.* Virol J, 2015. **12**: p. 6.
22. Qiu, J., et al., *The transcription profile of Aleutian mink disease virus in CRFK cells is generated by alternative processing of pre-mRNAs produced from a single promoter.* J Virol, 2006. **80**(2): p. 654-62.
23. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nuc. Acids Res., 1997. **25**: p. 3389-3402.
24. Carver, T., et al., *Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data.* Bioinformatics, 2012. **28**(4): p. 464-9.
25. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
26. Suyama, M., D. Torrents, and P. Bork, *PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W609-12.
27. Armougom, F., et al., *Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W604-8.
28. Guindon, S., et al., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.* Syst Biol, 2010. **59**(3): p. 307-21.
29. Lobley, A., M.I. Sadowski, and D.T. Jones, *pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination.* Bioinformatics, 2009. **25**(14): p. 1761-7.
30. Biasini, M., et al., *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W252-8.
31. Carrillo-Tripp, M., et al., *VIPERdb2: an enhanced and web API enabled relational database for structural virology.* Nucleic Acids Res, 2009. **37**(Database issue): p. D436-42.
32. Schrödinger, L.L.C., *The PyMOL Molecular Graphics System*.
33. Fabre, P.H., et al., *A glimpse on the pattern of rodent diversification: a phylogenetic approach.* BMC Evol Biol, 2012. **12**: p. 88.
34. Pisano, J., et al., *Out of Himalaya: the impact of past Asian environmental changes on the evolutionary and biogeographical history of Dipodoidea (Rodentia).* Journal of Biogeography, 2015. **42**(5): p. 856-870.
35. Chapman, M.S. and M.G. Rossmann, *Structure, sequence, and function correlations among parvoviruses.* Virology, 1993. **194**(2): p. 491-508.
36. Kordis, D. and F. Gubensek, *Molecular evolution of Bov-B LINEs in vertebrates.* Gene, 1999. **238**(1): p. 171-8.
37. Huang, L.Y., S. Halder, and M. Agbandje-McKenna, *Parvovirus glycan interactions.* Curr Opin Virol, 2014. **7**: p. 108-18.
38. Zimmerman, L.M., L.A. Vogel, and R.M. Bowden, *Understanding the vertebrate immune system: insights from the reptilian perspective.* J Exp Biol, 2010. **213**(5): p. 661-71.

39. Simpson, A.A., et al., *The structure of an insect parvovirus (Galleria mellonella densovirus) at 3.7 A resolution.* Structure, 1998. **6**(11): p. 1355-67.
40. Smith, R.H., et al., *Germline viral "fossils" guide in silico reconstruction of a mid-Cenozoic era marsupial adeno-associated virus.* Sci Rep, 2016. **6**: p. 28965.
41. Head, J.J., et al., *First report of snakes (Serpentes) from the Late Middle Eocene Pondaung Formation, Myanmar.* Journal of Vertebrate Paleontology, 2005. **25**(1): p. 246-250.
42. Duchene, S., E.C. Holmes, and S.Y. Ho, *Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates.* Proc Biol Sci, 2014. **281**(1786).
43. Bexfield, N. and P. Kellam, *Metagenomics and the molecular identification of novel viruses.* Vet J, 2011. **190**(2): p. 191-8.
44. Arriagada, G. and R.J. Gifford, *Parvovirus-derived endogenous viral elements in two South American rodent genomes.* J Virol, 2014. **88**(20): p. 12158-62.
45. Yuki, K., et al., *An endogenous adeno-associated virus element in elephants.* Virus Res, 2018.

Figure captions

Figure 1. The genetic structure of six amdoparvovirus-like PVe loci identified in this study. The structure of loci is shown in all six reading frames.. Green arrows represent amino acid (aa) strings homologue to the non-structural (NS) proteins of parvoviruses, whereas capsid protein (VP) homologues are shown in cyan. Vertical black lines mark stop codons. The aa string marked by brown in case of the Transcaucasian mole vole 2 element represents the potentially expressed, NS-fused region of the MafG transcription factor. Abbreviations: EVE – endogenous viral element, LINE – long interspaced nuclear element, LTR – long terminal repeat, PLA2

– phospholipase A2 domain, SINE – short interspaced nuclear element, VP1u – VP1 unique region.

Figure 2. Genetic structure of the PVe identified in this study, shown in relation to a generalized amdoparvovirus genome. Abbreviations: AmPVe-Ellut – endogenous amdoparvovirus element of the Transcaucasian mole vole (*Ellobius lutescens*), PVe – endogenous parvoviral element, NS – non-stroctural protein, ORF-M – a small additional ORF of amdoparvovirus genomes, VP – capsid protein.

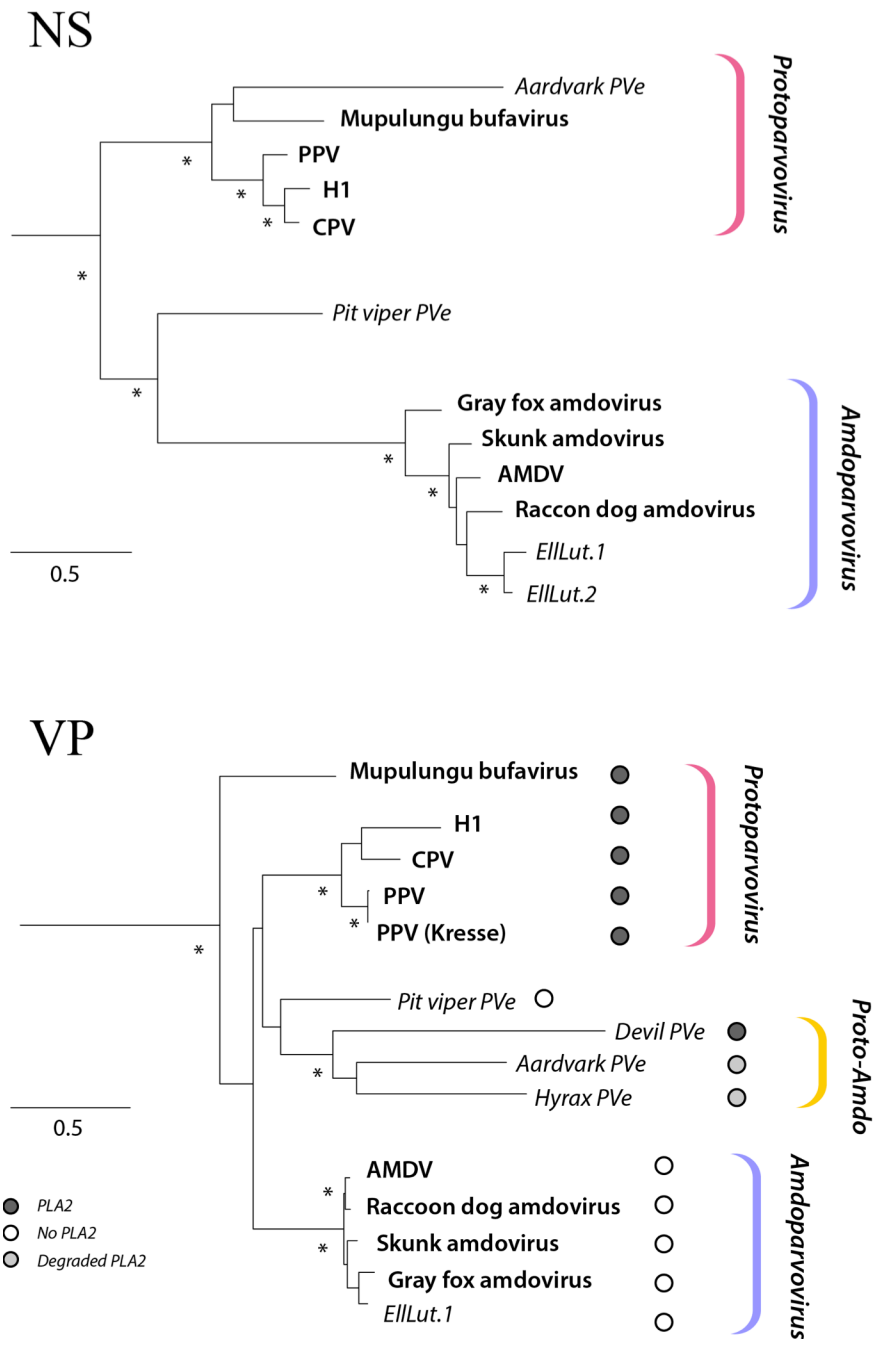


Figure 3. Maximum likelihood phylogenies based on the amino acid sequences of the (a) non-structural (NS) and (b) capsid protein (VP) of subfamily *Parvovirinae*. Alignments spanned XXX and XXX amino acids in NS and VP respectively. Brackets indicate genera (*Amdoparvovirus*, *Protoparvovirus*) and clades (Proto-Amdo). Taxa labels are shown in bold for taxa derived from parvoviruses, and in italics for taxa derived from PVe. Asterisks indicate node with bootstrap support >90%. The scale bar indicates number of changes in

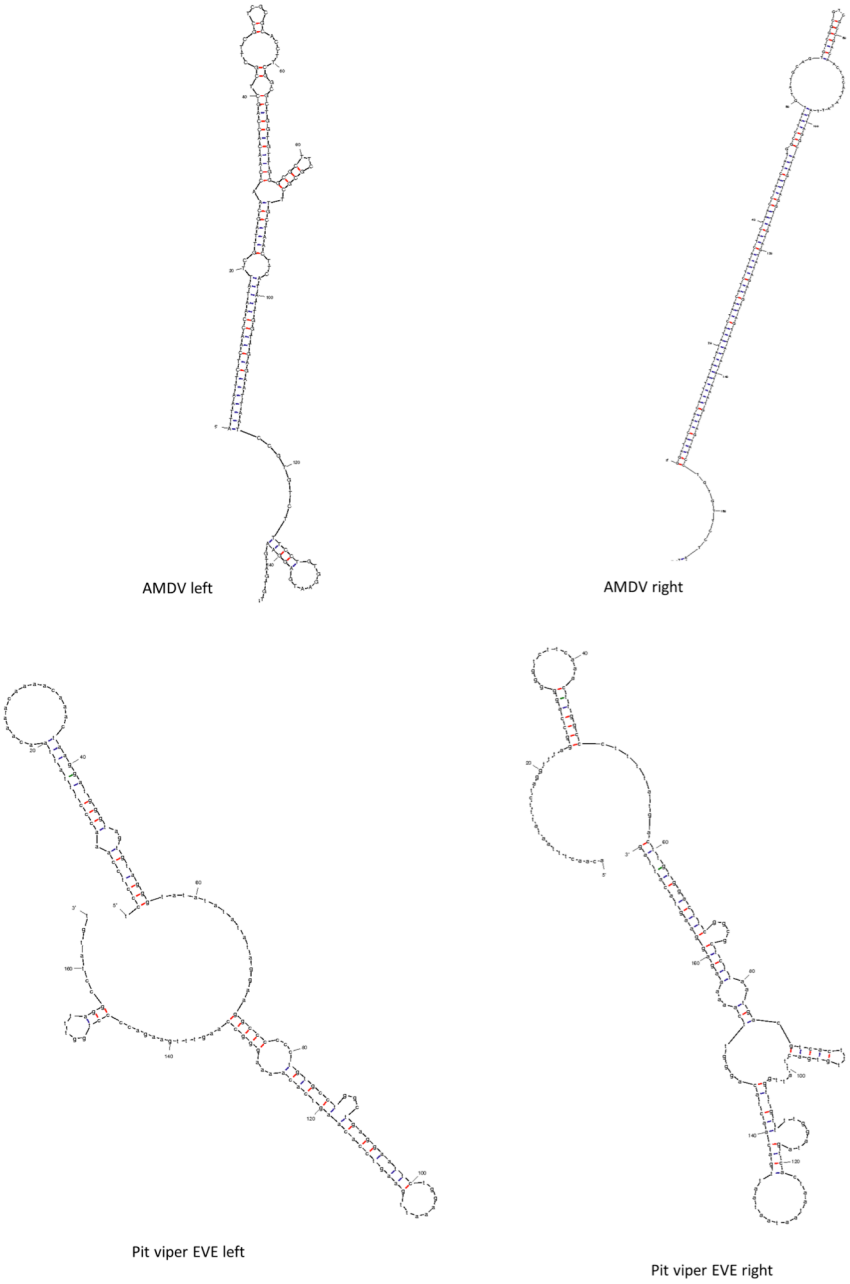
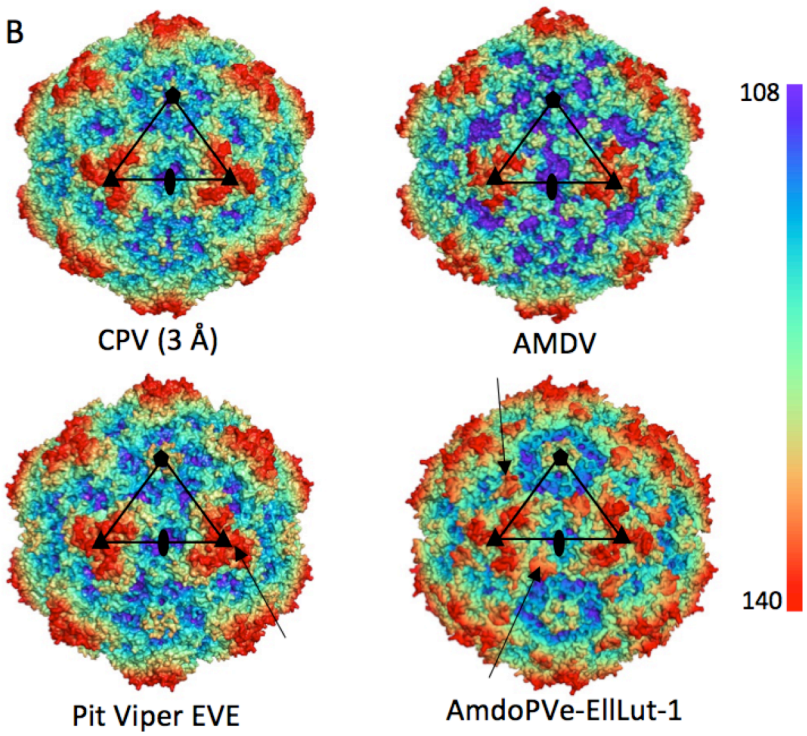


Figure 4. Hairpin-like structures identified directly downstream of the predicted promoter sequence (left hairpin) and directly upstream of the last polyadenylation signal (right hairpin) in a genome-length PVe identified in the pit viper genome. The hairpin like structures found in AMDV are shown for comparison..

A
racoond
skunk
mink
gfox
vole
Pit_viper
cons
VRI
VRII
VRIII
VRIV
VRV
VRVI
VRVII
VRVIII


B
CPV (3 Å)
AMDV
Pit Viper EVE
AmdoPVe-EllLut-1
108
140

C

Figure 5. (A) Alignment of the most complete endogenous viral elements (EVEs) with exogenous amdoparvoviruses. The variable regions (VR) are indicated by the horizontal lines. Abbreviations: racoond – racoon dog amdoparvovirus, gfox – grey fox amdoparvovirus, skunk – skunk amdoparvovirus, mink – Aleutian mink disease virus, vole – Transcaucasian mole vole endogenous amdoparvovirus, pit viper – pit viper endogenous parvoviral element. (B) Results of homology modeling; the capsid structure of canine parvovirus (CPV) served as a reference structure for all the three further models. The bar shows the distance from the capsid center in Ångströms and the structures are colored accordingly. The pentagon marks the five-fold, the triangles the three-fold and the two-fold is indicated by an ellipse. The arrows mark the VRIII region of the pit viper EVE and the VRVII of the Transcaucasian mole vole EVE (AmdoPVe-EllLut-1) capsids, which contain the only insertions compared to amdoparvovirus VR regions. (C) Ribbon diagrams of the VRVIII (left) and VRVIII (right) loops of CPV (blue), Aleutian mink disease virus (AMDV) (black), AmdoPVe-EllLut-1 (pink) and the pit viper EVE (yellow) capsid models and structures.

```
Mpulungu bufavirus   1 M-PLLKGF-----------------------KDGLTLPGYNYLGPGNSDFTKKPTNASDAAARKHDLAYGSYLKKGHNPYINFNNADKTFIKDT---------   70
h1                   1 MAPPAKRA----------------------KRGWVPPGYKYLGPGNSLDQGEPTNPSDAAAKEHDEAYDQYIKSGKNPYLYFSAADQRFIDQTK--------   72
canine parvovirus    1 MAPPAKRARRGKGVLVKWGEGKDLITXLSMCFFIGLVPPGYKYLGPGNSLDQGEPTNPSDAAAKEHDEAYAAYLRSGKNPYLYFSPADQRFIDQTK--------   96
porcine parvovirus   1 MAPPAKRA------------------------RGLTLPGYKYLGPGNSLDQGEPTNPSDAAAKEHDEAYDKYIKSGKNPYFYFSAADEKFIKETE--------   71
Tas. Devil EVE       1 ------------------------------MQNYTLPGYNYLRPGNTGHEQKLTSQCDKAAKLHVQAYXQ-IDQGHNPYLYFNKTDKNFIEATKETSXRLPPE   72
Aardvark EVE         1 F-----------------------------IGYATSGYNYLGPGNTLQSQKKKKKKPNPCD-QLQKH-KMRKKGKNLYIFHNVTDDKLTQATE---------   62
Rock hyrax EVE       1 MIIYSYNIL---------------------IYCKLYVLPGYYYLVPGNILEGKKTTKPPDEAVKKHDQGYILLTEQDLNPYLLHNESDHKLIKATN---------   75
AmdoPVe-EllLut-1     1 M-----------------------------------------------------------------------------------------------------    1
Pit_viper            1 M-----------------------------------------------------------------------------------------------------    1
racoondog amdoPV     1 M-----------------------------------------------------------------------------------------------------    1
skunk amdoPV         1 M-----------------------------------------------------------------------------------------------------    1
AMDV                 1 M-----------------------------------------------------------------------------------------------------    1
Grey fox amdoPV      1 M-----------------------------------------------------------------------------------------------------    1

cons                 1                                                                                                          105

Mpulungu bufavirus  71 ----NQAK----------DWGGWLGNKFFKIKEHIAPKLPEEPP-TKKPRPSTSKDPKYSWRHLKEGT----KKPKPFWLFVNKARQKKS---MDGGTST---DT  150
h1                  73 -----DA----------KDWGGKVGHYFFRTKRAFAPKLATD-S-----EPGTSGV-S-----R-AGK----RTRPPAYIFINQARAKKK---LTSSAAQQSSQT  142
canine parvovirus   97 -----DA----------KDWGGKIGHYFFRAKKAIAPVLTDTPD-----HPSTSRP-T-----K-PTK----RSKPPPHIFINLAKKKKA---GAGQVKRDNLAP  167
porcine parvovirus  72 -----HA----------KDYGGKIGHYFFRAKRAFAPKLSETDSPTTSQQPEVRRSPR-----KHPGSKPPGKRPAPRHIFINLAKKKAK---GTSNTNSNSMSE  153
Tas. Devil EVE      73 NLYTNQLQWLNTLXRFQQQATGEVALKVFKVKELLLPKLTPPI---KKAKT------QFHLTHF-GKT----NKQVPRHVFIKQARKIXSRKLSQSNKEN---PS  160
Aardvark  EVE       63 -----DW-------------GGKVAHCVFKTKRTLVPKTPTSLKIKXINKXYKQNYFYPRCRSY-RGI----KQVAPKHIFVNLAKRRVQ---QVNXQNTKQLLL  141
Rock hyrax EVE      76 -----NA-----------------------------------------------------------------------------------------------    77
AmdoPVe-EllLut-1     2 ------------------------------------PRVPQ--------------------H-YPGK----KRSAPRFVFVQQAKKKKQ---SNPAVSH-EELT   40
Pit_viper  EVE       2 ------------------------------------RRLGK--------------------R--PG----------KHIFINLARKKSKQ--SSQQTSE-DPLE   34
racoondog amdoPV     2 ------------------------------------SKIPQ--------------------H-FPGK----KRSAPRHVFIQQAKKRKQ---TNPAVFH-GEDT   40
skunk amdoPV         2 ------------------------------------SKIPQ--------------------H-YPGK----KRSAPRHVFIQQAKKKKQ---TNPAVYH-GPDT   40
AMDV                 2 ------------------------------------SKIPQ--------------------H-YPGK----KRSAPRHVFIQQAKKKKQ---TNPAVYH-GEDT   40
Grey fox amdoPV      2 ------------------------------------SKIPQ--------------------H-YPGK----KRSAPRHVFIQQAKKKKQH--SNPAVSH-QEDT   41

cons               106                                                                                                          210

Mpulungu bufavirus151 QDE--------PNDAAATASNTGT-------TGGGGGGGSGVGHS-TGNFDNRTDFRYE-NGEVTIICRATRLVHVKMSDSEEYRIFQTNNGDQFPP-STTRNEA  237
h1                 143 MSD-GTS-QPDSGNAVHSAARVERAADGPGGSGGGGSGGGGVGVS-TGSYDNQTHYRFLGDGWVEITALATRLVHLNMPKSENYCRIRVHNTTDTSV-------K  237
canine parvovirus  168 MSD-GAV-QPDGGQPA-VRNERAT-GSGNGSGGGGGGGSGGVGIS-TGTFNNQTEFKFLENGWVEITANSSRLVHLNMPESENYRRVVVNNMDKTAV-------N  260
porcine parvovirus154 NVE-QHN-PINAGTEL-SA--TGN-ESGGGGGGGGGRGAGGVGVS-TGTFNNQTEFQYLGEGLVRITAHASRLIHLNMPEHETYKRIHVLNSE-SGV-------A  243
Tas. Devil EVE     161 MTE--------QMQQLPLRENSQE-------QAGLGNGGRRVGSRYLNRLHQCSPXNYN-KG------HASRYINQSMPKNERYTYYKKVETQFTGT-GAT----  238
Aardvark EVE       142 NKQ-WRP-KTLQMMQ--VVSKLLX--DL---R-ALVGEERRVSLN-IGDFVNTTEWKYN-NGHVYITCHTSRLIHLNMPENEEYTLQQITTTF-T-T-------I  225
Rock hyrax EVE      78 ----------------------K--DG---E-I-------KKKQ-TLFFTKTIKWNFD-NGCVYVTCHASRGVHLNMPGSEGYTSPKMPKTL-S-A-------T  135
AmdoPVe-EllLut-1    41 VEEME---EPEQMDTG-------------------------EQ-ATNFNNTTEFKVT-NNEVTITCHATRVVHINQADTDEYLIFNAGRTVEDKPVYRDSLDL  113
Pit_viper  EVE      35 GTSQDTHHSPTVMDNAPAQTTMEETAAGLG-GASGGMGGGGVGLS-TGGLRSTIEWKHLGGGEILITCHSSRLIHLNMPESEEYKMLTINNNARPPG----IEYI  133
racoondog amdoPV    41 IEEMDST-EPEQMDTEQATNQTAEAG----------GGGGGVGNS-TGGFNNTTEFKVI-NNEVYITCHATRMVHINQADTDEYVIFNAGRTTDTKT-HLKKLNL  131
skunk amdoPV        41 IEEMDPA-EPEQMDTEQATNQTAEAG-------------GGVGNS-TGSFNNTTEFKVI-NGEVYITCHSTRMVHINQADTDEYLVFNAGRTTDTTT-THDKLDL  128
AMDV                41 IEEMDST-EAEQMDTEQATNQTAEAGGGGGGGGGGGGGGGGGVGNS-TGGFNNTTEFKVI-NNEVYITCHATRMVHINQADTDEYLIFNAGRTTDTKT-HQQKLNL  141
Grey fox amdoPV     42 IEEMDSN-ETEQMDISEQ-QQVAEHG-------------GGVGKS-TGGFNNTTEFKVT-NNEVIITCNATRVVHINQASTDEYLIFNAGREIDKTP--KGSLNL  127

cons               211                                                                                                          315
```

Figure 6. Alignment of the minor capsid protein VP1 N-terminals of representative members from genera *Protoparvovirus* (yellow) and *Amdoparvovirus* (cyan), including the VP1 unique region, the location of the phospholipase A2 (PLA2) domain, if present (bold). The calcium binding site is marked with the blue arrows and black arrows highlight the catalytic core. A glycine-rich stretch is present at the beginning of the major capsid protein VP2-overlapping region (framed), which is suspected to be responsible for conformational changes during endosomal escape. Similarly to amdoparvoviruses, the pit viper and vole PVe completely lack the PLA2, whereas the catalytic core absent in case of the aardvark EVE. Remnants of the G-rich region are recognizable in case of the Tasmanian devil PVe, but entirely absent in the aardvark, rock hyrax and vole PVe. The putative nucleus localization signal (underlined) is present in all endogenous sequences with the exception of the rock hyrax PVe. Abbreviations:EllLut-1 - Transcaucasian mole vole endogenous amdoparvovirus, PV – parvovirus