

# The virus–host interface: Molecular interactions of *Alphacoronavirus-1* variants from wild and domestic hosts with mammalian aminopeptidase N

Ximena A. Olarte-Castillo<sup>1,2</sup>  | Joana F. dos Remédios<sup>3</sup> | Felix Heeger<sup>1,4</sup> |  
Heribert Hofer<sup>1,5,6</sup>  | Stephan Karl<sup>1</sup> | Alex D. Greenwood<sup>1,5</sup>  | Marion L. East<sup>1,2</sup> 

<sup>1</sup>Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany

<sup>2</sup>ZIBI Interdisciplinary Center for Infection Biology and Immunity, Humboldt-Universität zu Berlin, Berlin, Germany

<sup>3</sup>Faculdade de Medicina Veterinária, Universidade de Lisboa, Lisboa, Portugal

<sup>4</sup>Berlin Center for Genomics in Biodiversity Research, Berlin, Germany

<sup>5</sup>Department of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany

<sup>6</sup>Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Berlin, Germany

## Correspondence

Ximena A. Olarte-Castillo, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany.

Email: x.olarte.castillo@gmail.com

## Funding information

Open Access funding enabled and organized by Projekt DEAL.

## Abstract

The *Alphacoronavirus-1* species include viruses that infect numerous mammalian species. To better understand the wide host range of these viruses, better knowledge on the molecular determinants of virus–host cell entry mechanisms in wildlife hosts is essential. We investigated *Alphacoronavirus-1* infection in carnivores using long-term data on Serengeti spotted hyenas (*Crocuta crocuta*) and molecular analyses guided by the tertiary structure of the viral spike (S) attachment protein's interface with the host receptor aminopeptidase N (APN). We sequenced the complete 3'-end region of the genome of nine variants from wild African carnivores, plus the APN gene of 15 wild carnivore species. Our results revealed two outbreaks of *Alphacoronavirus-1* infection in spotted hyenas associated with genetically distinct canine coronavirus type II (CCoVII) variants. Within the receptor binding domain (RBD) of the S gene the residues that directly bind to the APN receptor were conserved in all variants studied, even those infecting phylogenetically diverse host taxa. We identified a variable region within RBD located next to a region that directly interacts with the APN receptor. Two residues within this variable region were under positive selection in hyena variants, indicating that both sites were associated with adaptation of CCoVII to spotted hyena APN. Analysis of APN sequences revealed that most residues that interact with the S protein are conserved in wild carnivores, whereas some adjacent residues are highly variable. Of the variable residues, four that are critical for virus–host binding were under positive selection and may modulate the efficiency of virus attachment to carnivore APN.

## KEYWORDS

aminopeptidase N, carnivores, coronavirus, human APN, spike protein, virus–host interaction

## 1 | INTRODUCTION

Outbreaks of emerging viruses threaten the health of humans and livestock, undermine efforts to conserve biodiversity and are usually the result of human activities (Daszak et al., 2000; Hassell et al., 2017; Johnson et al., 2020). Determining the molecular interactions between viruses and their wildlife hosts is essential for our understanding of spillovers of infection to novel species, host jumps that establish infection in new species, and the evolution of virus–host adaptations (Li, 2013; Woolhouse et al., 2005). Virus–host cell entry mechanisms are key determinants of host species range and involve the virus attachment protein binding to a host cell receptor. Generally, the attachment protein of a virus is more likely to bind to the cell receptor of taxonomically closely related host species than to those in more distantly related ones (Longdon et al., 2014). Even so, some viruses can successfully infect distantly related species (Greenwood et al., 2012; Woolhouse et al., 2005).

Within the family *Coronaviridae*, the subfamily *Orthocoronavirinae* consists of four major genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* (de Groot et al., 2012) that contain coronaviruses (CoVs) that infect distantly related hosts. For example, some betacoronaviruses in the subgenus *Sarbecovirus* that cause severe acute respiratory syndrome (SARS) in humans are closely related to CoVs that infect bats (Lau et al., 2005; Lu et al., 2015). Within the genus *Alphacoronavirus*, the *Alphacoronavirus-1* species, groups viruses of veterinary importance including canine coronavirus (CCoV), feline coronavirus (FCoV), transmissible gastroenteritis virus (TGEV) and its variant porcine respiratory coronavirus (PRCV; de Groot et al., 2012). These CoVs cause enteric, respiratory or systemic infections in a taxonomically diverse range of mammalian species (Le Poder, 2011). *Alphacoronavirus-1* research has mostly focused on CoVs infecting domestic species including domestic dogs (*Canis lupus familiaris*), domestic cats (*Felis catus*) and domestic pigs (*Sus scrofa domestica*). However, there is both genetic (Alfano et al., 2019; Gao et al., 2009; Goller et al., 2013; Heeney et al., 1990; Ma et al., 2008; Pearks Wilkerson et al., 2004; Vijaykrishna et al., 2007) and serological (East et al., 2004; Harrison et al., 2004; Packer et al., 1999; Thalwitzer et al., 2010; Woodroffe et al., 2012) evidence of *Alphacoronavirus-1* infection in several wild species of the order Carnivora (to which the domestic dog and cat belong). This order has suborders Feliformia which includes families Felidae, Hyaenidae and Viverridae and Caniformia which includes families Canidae, Ursidae and Mustelidae (Wozencraft, 2005). Enteric *Alphacoronavirus-1* infection occurs in spotted hyenas (*Crocuta crocuta*, family Hyaenidae), in the Serengeti National Park (Serengeti NP) in Tanzania, with infection predominantly occurring in juveniles (Goller et al., 2013) and a high (74%) seroprevalence of antibody titres in adults (East et al., 2004). Phylogenetic analyses of partial sequences obtained between 2003 and 2008 revealed infection of spotted hyenas with diverse variants and in 2007 genetically distinct variants from sympatric spotted hyenas and one silver-backed jackal (*Canis mesomelas*, family Canidae, Goller et al., 2013), called jackals from hereon. In order to understand how *Alphacoronavirus-1* can infect such

distantly related hosts and to assess which other carnivore species they might infect, studies on the genetic and molecular adaptations of *Alphacoronavirus-1* variants from different carnivores to the receptors of different mammalian hosts are needed.

*Alphacoronavirus-1* possesses a single-stranded, nonsegmented, positive-sense RNA genome. The genomic sequence contains several open reading frames (ORF). Towards the 5′-end are the two ORFs (1a and 1b) that encode the polymerase proteins. Towards the 3′-end of the genome are several ORFs that encode a varied number of nonstructural proteins and four structural proteins: the spike (S), envelope (E), nucleocapsid (N) and matrix (M) proteins (Gallagher & Buchmeier, 2001). The S, M and E proteins form the virion envelope, the N protein is a phosphoprotein that packages the viral RNA genome and form a helical nucleocapsid inside the virion. CCoV and FCoV are divided into genotype I (CCoVI and FCoV I) and genotype II (CCoVII and FCoV II), because FCoV I and CCoV I are genetically and antigenically divergent from FCoV II and CCoV II (Le Poder, 2011) and TGEV and PRCV are genetically and antigenically closely related to CCoV II (Enjuanes, 2000). FCoV II is the result of homologous recombination between FCoV I and CCoV II (Herrewegh et al., 1998). Consequently, the FCoV II S gene is homologous to that of CCoV II and the rest of the genome is homologous to FCoV I (Herrewegh et al., 1998). Thus within the *Alphacoronavirus-1* S gene phylogeny, clade A contains CCoV I and FCoV I and clade B contains CCoV II, FCoV II, TGEV and PRCV (Whittaker et al., 2018).

Coronavirus entry to host cells is mediated by the S protein, considered a critical determinant of viral host range (Kuo et al., 2000; Li, 2016). The S gene is divided into the amino (N)-terminal S1 subunit and the carboxy (C)-terminal S2 subunit (Li, 2012). The S2 subunit mediates membrane fusion with the host upon receptor binding. The S1 subunit is involved in host receptor binding and contains two domains, the N-terminal domain (S1-NTD) and the C-terminal domain (S1-CTD). Within the S1-CTD is the receptor binding domain (RBD) that interacts with the host receptor. The host cell receptor used by FCoV II, CCoV II, TGEV and PRCV is the protein aminopeptidase N (APN; Benbacar et al., 1997; Delmas et al., 1992, 1994; Tresnan et al., 1996). APN has several functions and is highly expressed on intestinal and kidney brush border membranes, mucosal cells in the small intestine, liver cells, immune cells and at synaptic junctions of neurotransmitters (Luan & Xu, 2007; Sjöström et al., 2000). The S protein of CCoV II, FCoV II, TGEV and PRCV bind to overlapping regions at the C-terminal segment of APN which strongly suggests these viruses interact in a similar manner with their respective APN receptor (Delmas et al., 1994; Hegyi & Kolb, 1998; Tusell et al., 2007). This is consistent with observations that TGEV, CCoV II and FCoV II can interchangeably bind to the intestinal brush border membranes of domestic pigs, dogs and cats, but not to that of the house mouse (*Mus musculus*), which is resistant to *Alphacoronavirus-1* infection (Levis et al., 1995). Furthermore, baby hamster kidney cells (BHK) and mouse fibroblast cells (NIH 3T3), which are resistant to infection with FCoV II, CCoV II and TGEV, can be infected by all these viruses when expressing domestic cat APN (feline APN, fAPN, Tresnan et al., 1996). As two domestic cat cell lines (whole foetus,

FCWF and kidney CRFK) and the domestic dog fibroblast cell line A72 all support FCoV, CCoV and TGEV infection (Benbacar et al., 1997; Kolb et al., 1998), there is evidence that fAPN and domestic dog APN (canine APN, cAPN) support infection with a range of *Alphacoronavirus-1* variants.

The crystal structure of PRCV's RBD coupled to porcine APN (pAPN) revealed the crucial residues both in the RBD of the virus (19 residues located in regions  $\beta 1$ – $\beta 2$  and  $\beta 3$ – $\beta 4$ ) and on the host APN (23 residues) that mediate binding (Table S1; Reguera et al., 2012) and thus are essential in the virus–host interaction. Although comparative analysis of receptor protein sequences has been used to predict the efficiency of binding in other CoVs (Damas et al., 2020), the diversity of residues essential for binding of *Alphacoronavirus-1* variants from wild carnivores to APN sequences from wild species of carnivore has not been explored.

Our study aims to (i) provide an extensive molecular characterisation of *Alphacoronavirus-1* variants obtained from wild carnivores in the Serengeti NP during a monitoring period of 12 years, (ii) consider fluctuations in infection prevalence over time in relation to variant type, (iii) provide the first genetic analysis of the APN receptor in wild carnivore species in both the Caniformia and Feliformia suborders, (iv) use the published tertiary structure of the *Alphacoronavirus-1* S-APN interface to guide a detailed comparative genetic analysis of *Alphacoronavirus-1* S genes from carnivore species and other mammalian hosts, and (v) investigate the importance of both virus and host residues for the binding of S protein from *Alphacoronavirus-1* variants to mammalian APN receptors, particularly those in wild carnivores.

## 2 | MATERIALS AND METHODS

### 2.1 | Long-term *Alphacoronavirus-1* data collection

Fresh faecal samples were collected from spotted hyenas and jackals in the Serengeti NP between 2001 and 2012. Sampling effort was relatively consistent across years and seasons for spotted hyenas. Spotted hyena samples came from three large, closely watched clans for which all animals were individually known and their life histories recorded (Marescot et al., 2018). Spotted hyenas were considered adult when 24 months old, juvenile when less than 24 months of age (Hofer & East, 2003). After collection, faeces were thoroughly mixed and divided into aliquots.

Previously, Goller et al. (2013) screened 65 faecal samples collected from our study clans and 17 faecal samples from jackals in the Serengeti NP between 2003 and 2008. We extend the data collection period to 12 years (2002–2012) and increased screened sample sizes to 505 faeces from spotted hyenas (397 from juveniles, 108 from adults) and 74 faeces from jackals. Additionally, we opportunistically collected tissue samples from recently dead animals, including five intestines, two livers, nine lungs, 10 lymph nodes, 11 spleens from spotted hyena and one intestine, two livers and three lungs from jackals. For each tissue type we used only one sample per

carcass. Both faecal and tissue samples were stored and transported frozen at  $-80^{\circ}\text{C}$ , or preserved in RNAlater (Sigma-Aldrich Inc), initially stored and transported at  $-10^{\circ}\text{C}$ , and finally stored at  $-80^{\circ}\text{C}$  until analyses.

The screening of faecal and tissue samples for *Alphacoronavirus-1* RNA was performed as described by Goller et al. (2013) using primers targeting the conserved M gene (Pratelli et al., 1999, 2002) and the 3'-end of the more variable S gene (Pratelli et al., 2004). RT-PCR was performed using SuperScript III One-Step RT-PCR System (Life Technologies GmbH) following the manufacturer's instructions. Purification and sequencing of positive samples were done as previously described (OlarTE-Castillo et al., 2016).

Screening produced 19 partial M gene sequences (18 from spotted hyenas, one from a jackal), and 11 partial S gene sequences (10 from spotted hyenas, one from a jackal, Table S2). These sequences were aligned using the MUSCLE algorithm (Edgar, 2004) in GENEIOUS 9.0.2 (Biomatters Ltd) and compared with those obtained by Goller et al. (2013). Sequence alignments were used to construct haplotype networks for each partial gene using the median joining algorithm in Network 5.0 (Bandelt et al., 1999).

### 2.2 | Carnivore APN

Kidney samples were obtained from 15 carnivore species. From Feliformia these included cheetah (*Acinonyx jubatus*), African lion (*Panthera leo*), leopard (*Panthera pardus*), serval (*Leptailurus serval*), tiger (*Panthera tigris*), spotted hyena, aardwolf (*Proteles cristata*), brown hyena (*Hyaena brunnea*), striped hyena (*Hyaena hyaena*), white-tailed mongoose (*Ichneumia albicauda*), African civet (*Civettictis civetta*), from Caniformia African wild dog (*Lycaon pictus*), silver-backed jackal, bat-eared fox (*Otocyon megalotis*) and grey wolf (*Canis lupus*). We obtained kidney samples from spotted hyena, aardwolf, serval, silver-backed jackal, bat-eared fox, African civet, white-tailed mongoose opportunistically from animals that died of predation, disease and car accidents in the Serengeti ecosystem. The cheetah sample was provided by the IZW Cheetah Project (CITES Namibian export permit number 0030491, German import permit E-05403/11). Brown hyena samples were provided by The Brown Hyena Project in Namibia. African lion, leopard, tiger and wolf samples came from dead animals from European zoos supplied to the IZW Department of Wildlife Diseases for pathological and disease examination. Tissue samples were stored at  $-20^{\circ}\text{C}$ , transported frozen and then stored at  $-80^{\circ}\text{C}$ , or preserved in RNAlater, stored and transported at  $-10^{\circ}\text{C}$  and finally stored at  $-80^{\circ}\text{C}$ .

### 2.3 | Hybrid capture

Ten *Alphacoronavirus-1* positive samples (eight from spotted hyenas, two from jackals) were selected for sequencing of the complete 3'-end of the genome (~9 kb). This produced nine complete sequences of the 3'-end of the genome plus one incomplete sequence which

contained the complete sequence of the S, N and 7a genes from a spotted hyena variant from 2008. Kidney samples from 11 carnivore species were used to sequence the complete APN gene (2904 nucleotides). For these two purposes, we used hybrid-capture enrichment introduced by Maricic et al. (2010) with two steps, the production of PCR "baits" from known sequences and the production of pooled cDNA libraries. The baits are used to capture by hybridisation the target sequences from the cDNA libraries. The specific hybridization capture protocol for this study was done as previously described by Tsangaras et al. (2014). Baits for the sequencing of the 3'-end of the genome of *Alphacoronavirus-1* were generated from RT-PCR products from a spotted hyena variant from 2007 (SH110\_2007) using seven overlapping primers (Decaro et al., 2007). Baits for the sequencing of the APN gene were generated from RT-PCR products obtained from a kidney sample from a bat-eared fox using previously described primers (Tresnan et al., 1996). Bait preparation for the hybridization including ligation to a biotin adapter and immobilization on streptavidin magnetic beads was done as previously described (OlarTE-Castillo et al., 2015).

To generate sequencing libraries, cDNA was synthesized, purified and sheared as previously described (OlarTE-Castillo et al., 2015). Libraries were indexed by sample and built as previously described (Meyer & Kircher, 2010). DNA quantity and quality were monitored using the Agilent 2200 TapeStation (Agilent Technologies). Baits and cDNA libraries were constructed in separate laboratories to avoid contamination.

## 2.4 | High throughput sequencing

Enriched libraries resulting from the hybrid capture of the 3'-end of the *Alphacoronavirus-1* genome and the APN gene were sequenced on the Illumina MiSeq platform (Illumina Inc.). Sequence quality was assessed with FASTQC 0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Removal of adapters and trimming of sequences was done as previously published (OlarTE-Castillo et al., 2015). Trimmed reads from the hybrid capture of the 3'-end of *Alphacoronavirus-1* were assembled for each sample by using the software BWA (Li & Durbin, 2009) against seven reference sequences covering all genotypes of CCoV, FCoV and TGEV (accession numbers CCoV1b EU856362, FCoV1 DQ010921, CCoV1 DQ112226, AY342160, TGEV DQ811788, FCoV1 AB088223 and CCoV1 AY307021). With this approach we determined the reference sequence to which each sample had mapped reads. This sequence was then used as a reference for further mapping using the iterative mapping assembler MITOBIEM version 1.5 (Hahn et al., 2013). Sequence gaps found after mapping were filled or confirmed by RT-PCR using SuperScript III One-Step RT-PCR System (Life Technologies GmbH) with primers designed from flanking sequences obtained after mapping. Purification and sequencing of these segments were done as previously described (OlarTE-Castillo et al., 2016).

Trimmed reads from the hybrid capture of the APN gene were mapped against the APN sequence from domestic dog or cat

(accession numbers NM001146034 and NM001009252) using the medium sensitivity method with up to 10 iterations in GENEIOUS 9.0.2 (Biomatters Ltd).

## 2.5 | Genetic analyses of *Alphacoronavirus-1*

The nine complete 3'-end genome sequences obtained in this study (seven spotted hyena variants, two jackal variants) were aligned using the MUSCLE algorithm (Edgar, 2004). This alignment showed deletions in accessory genes 3b, 7a and 7b in our sequences. Nucleotide variability along the six genes that did not have large deletions (S, 3a, 3c, E, M, N) in nine sequences were visualized in SIMPLOT 5.1 (Lole et al., 1999).

An additional alignment of our nine sequences with 44 other complete 3'-end *Alphacoronavirus* sequences (11 CCoV1, 5 FCoV1, 4 CCoV1b, 1 CCoV1, 10 FCoV1, 9 TGEV, 2 PRCV, 2 Mink CoV and 2 Ferret CoV) was performed using the MUSCLE algorithm (Edgar, 2004). We searched for evidence of recombination events and break points in these sequences using the Recombination Detection Program (RDP; Martin & Rybicki, 2000), GENECONV (Padidam et al., 1999), BOOTSCAN (Martin et al., 2005), MAXCHI (Maynard Smith, 1992), CHIMAERA (Posada & Crandall, 2001) and SISCAN (Gibbs et al., 2000) methods implemented in RDP4 (Martin et al., 2015).

We investigated phylogenetic relationships between our variants and other variants using sequences of the subunits of the S gene (S1-NTD, S1-CTD, S2) and the other structural genes (E, M, N). Each alignment included different groups of sequences for two reasons. First, large deletions in PRCV variants in the S1-NTD subdomain precluded its inclusion in the phylogenetic analysis of this region. Second, sequences from wild carnivores other than those from this study are based on sequences from different genes (Table S3). Information about the sequences used in our analyses, including genotype, variant name and species is available in Table S4. Each sequence data set was aligned using a codon alignment in PAL2NAL (Suyama et al., 2006), taking into account the amino-acid sequence aligned using the CLUSTAL OMEGA algorithm (Goujon et al., 2010; Sievers et al., 2011) in the server of the EMBL (McWilliam et al., 2013). Each of these six alignments was used to reconstruct the phylogenetic relationships between our sequences and published sequences. For this purpose, we used the maximum likelihood (ML) and Bayesian Markov chain Monte Carlo phylogenetic inferences, as previously published (OlarTE-Castillo et al., 2015).

We also analysed the structural genes (S, E, M, N) to identify signatures of positive selection using only sequences from field variants and excluding variants passaged in culture and sequences with ambiguous nucleotide sequences. When identical sequences were present, only one was used and if sequences from different wild carnivores were available these were included in the analysis (Table S3). For the E, M, and N genes, variants from all groups (CCoV1, CCoV1, CCoV1b, FCoV1, FCoV1, TGEV, PRCV) were used. Analysis of the S gene was performed separately for the S1-NTD, S1-CTD and S2 subunits. For these analyses, variants from FCoV1

and CCoV were excluded as they are highly divergent (<50% similarity with FCoV/CCoV/TGEV/PRCV in the S gene), possibly because they use different receptors (Dye et al., 2007; Hohdatsu et al., 1998) and have a low antigenic similarity (Hohdatsu et al., 1992). For the analysis of S1-NTD, sequences from TGEV and CCoVb were excluded because of the high divergence in this region with FCoV/CCoV (Table S4).

As recombination occurs frequently in alphacoronaviruses (Graham & Baric, 2010), we used RDP (Martin & Rybicki, 2000) to find recombinant sequences and excluded them from the analysis of sites under positive selection (Table S4). To detect sites under positive selection for each gene, we used the ML site models implemented in the CODEML program of PAML version 4 (Yang, 1997, 2007) as previously published (Nikolin et al., 2017). Only sites with posterior probabilities >95% were considered. These models assume pervasive positive selection across the entire phylogeny (Murrell et al., 2012). Therefore, we additionally looked for signals of episodic diversifying selection using mixed-effect models of evolution (MEME) which permit the identification of both pervasive and episodic positive selection that affects specific sites along particular lineages (Murrell et al., 2012). For this method we used a cutoff value of .05 and only events of positive selection with an Empirical Bayes Factor of >100 were considered. Finally, using the amino acid sequence of the S1-CTD of the S protein, we performed an analysis to detect directional selection using directional evolution of protein sequences (DEPS; Kosakovsky-Pond et al., 2008). We used this domain because it contains the RBD and therefore is involved in receptor use. For this analysis we rooted the tree using the TGEV group. The tree was constructed using the neighbour joining method with the Jones-Taylor-Thorton model in MEGA version 6.06 (Tamura et al., 2013). The MEME and DEPS analyses were carried out in the DATAMONKEY server (Kosakovsky-Pond & Frost, 2005).

## 2.6 | The APN gene

Using hybrid capture we obtained the complete sequence of the APN gene for eight African carnivores (jackal, bat-eared fox, African wild dog, aardwolf, brown hyena, spotted hyena, cheetah and white-tailed mongoose). Using RT-PCR and previously published primers (Tresnan et al., 1996), three additional complete sequences (2904 to 2907 nucleotides long) were obtained from grey wolf, African lion and striped hyena and four incomplete sequences (1941 nucleotides long, missing the first 963 nucleotides) from African civet, leopard, serval and tiger. The purified products were sequenced bidirectionally using the BIG DYE TERMINATOR CYCLE sequencing kit 1.1 (Applied Biosystems) following the manufacturer's instructions. A 3130 GENETIC ANALYZER (ABI) was used for sequencing. The nucleotide and translated amino-acid sequences were aligned using the MUSCLE (Edgar, 2004) and the CLUSTAL OMEGA (Goujon et al., 2010; Sievers et al., 2011) algorithms, respectively, with the sequences of the APN from the domestic pig, three carnivore, four bat, seven rodent and

four primate species including humans (details in Table S5). With this nucleotide alignment (34 species in total) we constructed phylogenetic relationships using the ML method with 1000 bootstrap replicates to estimate the statistical support of branches. MEGA 5.0 (Tamura et al., 2013) was used to obtain the nucleotide substitution model and construct the phylogenetic tree used to detect sites under positive selection within the APN gene, using CODEML and MEME as explained above for the analysis of *Alphacoronavirus-1* structural genes. Average nucleotide and amino-acid similarities were calculated using DISCOVERY STUDIO VISUALIZER 4.0 (Accelrys Software Inc).

## 2.7 | Tertiary structure visualisation

The tertiary structure of the RBD in the S1-CTD of PRCV coupled with pAPN (PDB ID 4F5C, Reguera et al., 2012) was visualized using SWISS-PDB VIEWER 4.1.0 (Guex & Peitsch, 1997). Variable sites and sites detected as under positive selection in host and virus were mapped with different colours in the tertiary structure to see their relative position in the interface.

## 3 | RESULTS

### 3.1 | *Alphacoronavirus-1* in wild carnivores in the Serengeti NP

*Alphacoronavirus-1* RNA detected by RT-PCR in spotted hyena faecal samples indicated an overall infection prevalence of 8.9% (45/505 samples) and a significantly higher infection prevalence in juveniles of 11.3% (42/397) than adults of 2.8% (3/108; log likelihood ratio test,  $G = 6.41$ ,  $df = 1$ ,  $p = .01$ ,  $n = 505$ ). Infection prevalence in juveniles fluctuated between 2002 and 2012 (Figure S1), with evidence of two outbreaks of infection in 2003 when infection prevalence was 25.0% (5/20), and 2007, when infection prevalence was 21.4% (9/42 samples). In 2009, *Alphacoronavirus-1* RNA was not detected in any faeces from juveniles (0/37 samples).

During the same study period, overall infection prevalence in jackal faeces was 1.4% (1 /74 samples). The one *Alphacoronavirus-1* RNA positive sample was from 2007 (see Goller et al., 2013). One sample of intestines from 2011 was also positive for *Alphacoronavirus-1* RNA.

### 3.2 | Haplotype networks for M and S gene fragments

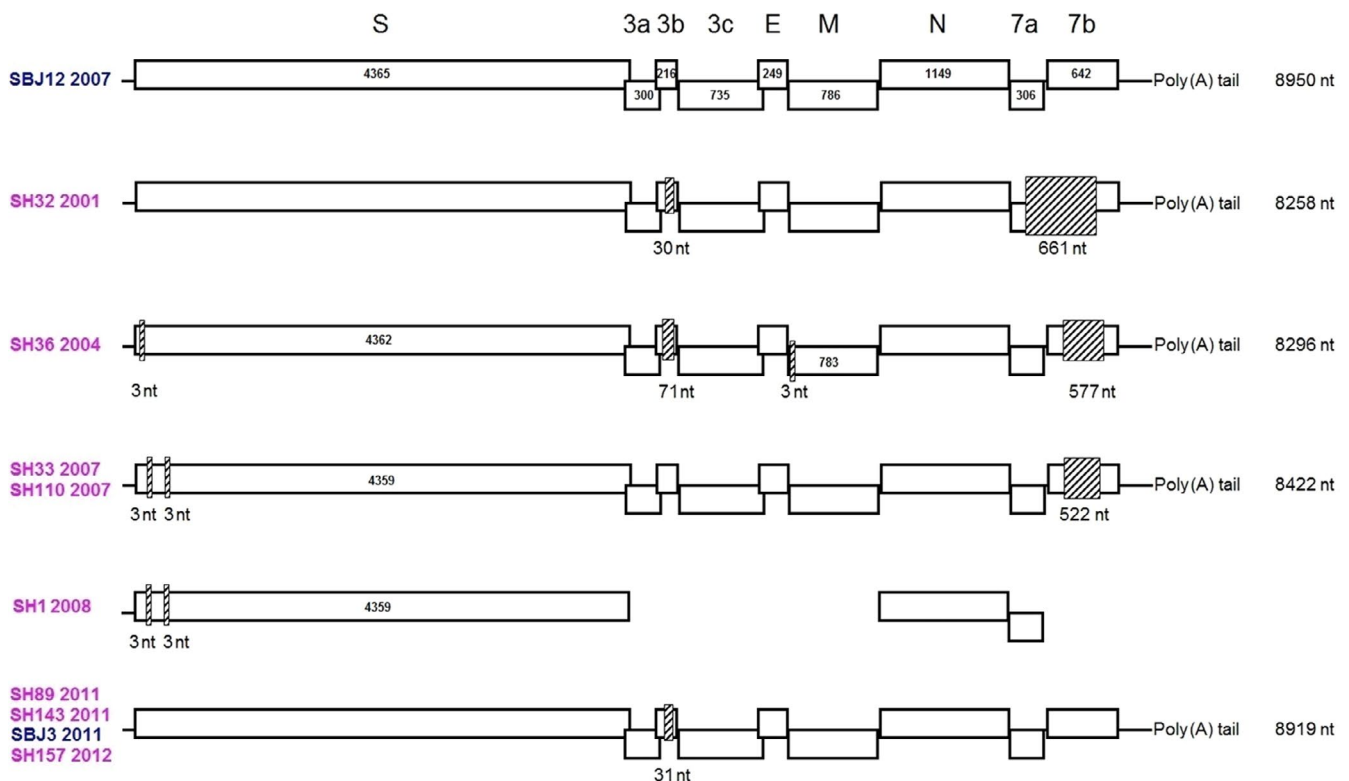
The haplotype network (Figure S2A) based on 33 partial M gene sequences (Table S2) from Serengeti NP (31 from spotted hyenas, two from jackals, including 14 sequences from Goller et al., 2013) revealed that all partial M gene haplotypes were connected to one hypothetical node, and that haplotypes grouped by year and not by

host species. One spotted hyena variant from 2004 (SH42\_2004) differed by 10 nucleotides from all other variants (Figure S2).

The haplotype network (Figure S2B) based on 25 partial S gene sequences (Table S2) from Serengeti NP (23 from spotted hyenas, two from jackals, including 14 sequences from Goller et al., 2013), revealed a substantially larger number of nucleotide differences between partial S gene haplotypes than partial M gene haplotypes. Variants from spotted hyenas in 2004 grouped together and were separated from those from other years by a long branch (39 nucleotide differences to the closest median node, Figure S2B). All 10 variants from spotted hyenas in 2007 were identical and grouped closely with those from 2006 but were separated by a long branch from the 2007 jackal variant (Figure S2B). All five spotted hyena variants from 2010, 2011 and 2012 grouped closely, but were not identical, and the jackal variant in 2011 (SBJ3-2011) was identical to the spotted hyena variant from 2012 (SH157-2012, Figure S2B). S gene haplotypes also grouped by year and not by species, as nucleotide differences between groups of variants from adjacent years were smaller than those from widely different years.

### 3.3 | Molecular characterisation of the complete 3'-end of the *Alphacoronavirus-1* genome

We obtained nine complete 3'-end genome sequences, seven from spotted hyenas, two from silver-backed jackals (accession numbers MF095847-MF095855), plus the complete sequence of the S, N and 7a genes (accession numbers KX156832, MW505904, MW505905, respectively) from a spotted hyena variant from 2008 (Figure 1). The 3'-end of the genome of all variants varied in size, from 8258 bp in spotted hyena (SH) variant SH32-2001 to 8950 bp in the jackal (SBJ) variant SBJ12-2007. Minor deletions in the S and M genes and major deletions in the nonstructural genes 3b, 7a and 7b accounted for size differences (Figure 1). Jackal variant SBJ12-2007 was the only one with all complete structural (S, E, M, N) and nonstructural (3a, 3b, 3c, 7a, 7b) genes and therefore used as reference to compare deletions in other variants (Figure 1). The S gene in variant SH36-2004 had a deletion of codon 4, and variants SH33-2007, SH110-2007, and SH1-2008 had two deletions (codons 5, 19). For the M gene, variant SH36-2004 had a deletion of codon 24. For the 3b gene different deletions were observed, one of 30 nucleotides (nt) in SH32-2001, one of 71 nt in SH36-2004 resulting in a premature stop at codon



**FIGURE 1** Schematic representation of the complete 3'-end of the *Alphacoronavirus-1* genome of variants from spotted hyena (SH, in pink) and silver-backed jackal (SBJ, in blue) from 2001 to 2012. Variant SBJ12 2007 (on top) had all nine genes and therefore was set as a reference variant. Each box represents a gene and the name of each gene is indicated above the reference variant. The number inside each box represents the length of different genes in the reference variant. In other variants, gene length is only presented for genes that differed in size to genes in the reference variant. Hatched boxes represent deletions, the number of nucleotides (nt) deleted is indicated below each hatched box. The total size of each 3'-end genome fragment, from the start of the S gene to the end (excluding the poly(A) tail), is indicated on the right of each schematic representation. The two SH variants from 2007 (SH33-2007 and SH110-2007) had the same genome arrangement. Three SH variants from 2011 and 2012 and one SBJ variant from 2011 (SH89-2011, SH143-2011, SH157-2012 and SBJ3-2011) had the same genome arrangement. Only the complete S, N and 7a genes were obtained from variant SH1-2008

24 and one of 31 nt in variants from 2011 and 2012 (SH89-211, SH143-2011, SH157-2012 and SBJ3-2011) resulting in a frame shift and a premature stop at codon 40. For the 7a gene, all variants had an intact gene except for SH32-2001 which had a 661 nt deletion spanning the 3'-end of the 7a gene and the 5'-end of the 7b gene. Deletions were also present in the 7b gene in variants SH36-2004 (577 nt) and SH33-2007, SH110-2007 (522 nt).

A similarity plot (Figure 2) based on 3'-end genome sequences of six variants including only the genes without major deletions (S, 3a, 3c, E, M, N) revealed that the S1-NTD domain of the S protein is the most variable region. Two other regions of high variability included the RBD within the S1-CTD, and the beginning of the M gene, encompassing the first 111 nucleotides (37 amino acids). In this region, variant SH36-2004 differed substantially from other Serengeti variants (Figure 2, in blue).

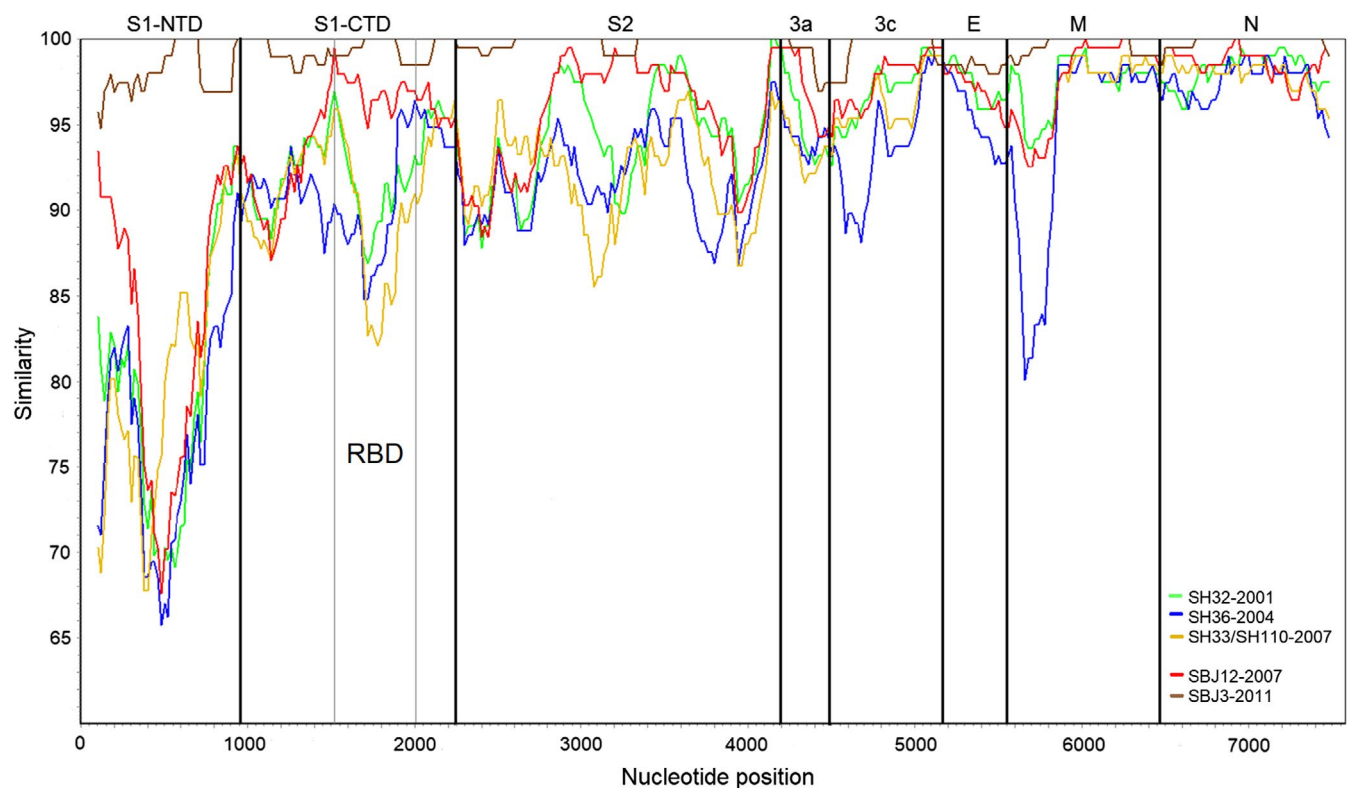
A phylogenetic analysis based on the S1-CTD of the S gene (Figure 3) of *Alphacoronavirus-1* variants from a range of hosts worldwide and Serengeti variants and analyses using other complete structural genes (S1-NTD, S2, E, M, N, Figure S3A-E, respectively), showed that Serengeti variants from spotted hyenas and jackals belonged to the FCoV/CCoV group. As we found no evidence of recombination throughout the 3'-end of the genome of our nine Serengeti variants these variants lack the FCoV genotype which has a recombinant origin. These findings indicate that all variants

from spotted hyenas and jackals belonged to the CCoV genotype. Variants from spotted hyenas did not cluster together in a single group within the FCoV/CCoV clade but were placed in different groups according to the year of collection (bootstrap values greater than 80, Figures 3 and S3A-E).

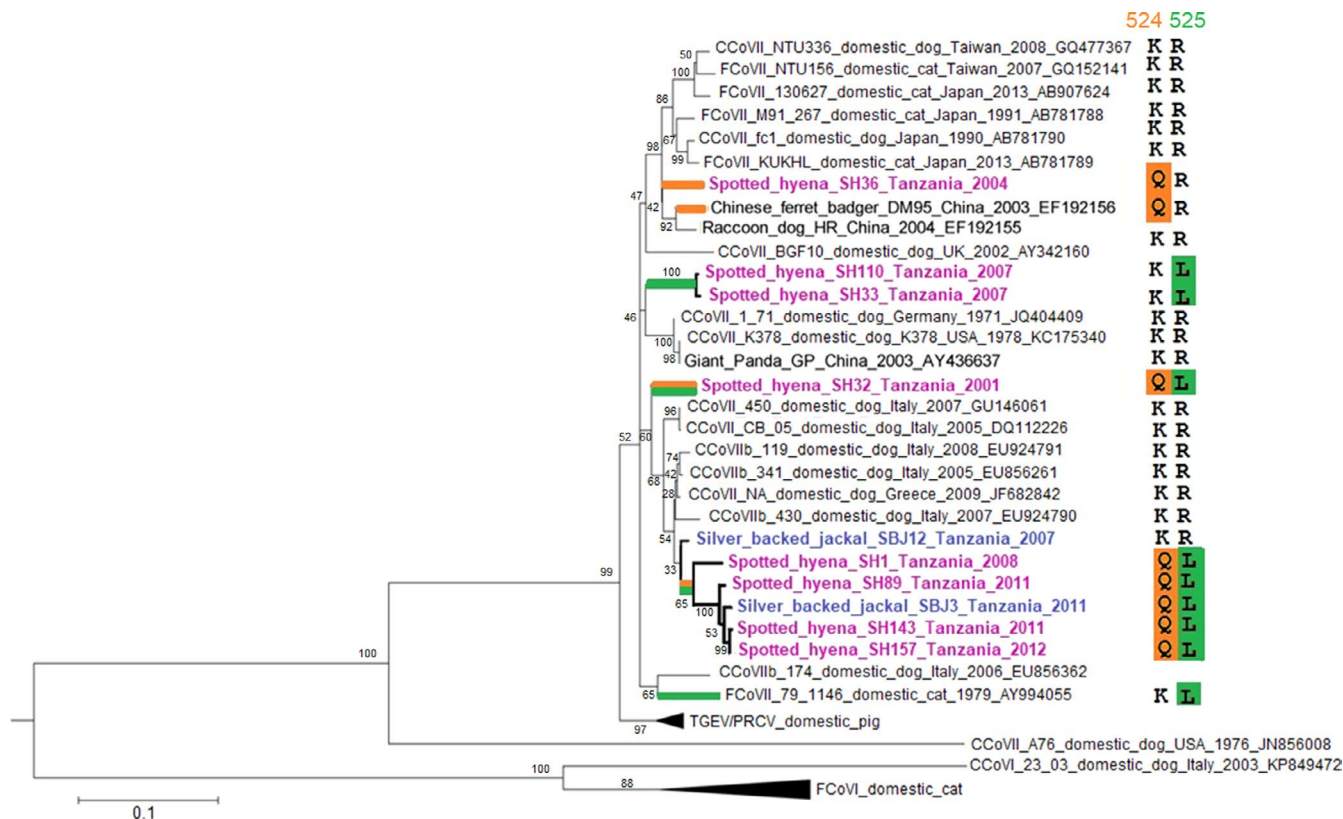
In the phylogenetic trees of the S1-CTD (Figure 3), S1-NTD domains and the S2 subunit of the S gene (Figure S3A-B), spotted hyenas variants from 2007 (SH110, SH33) were placed in a different group to that of the jackal variant (SBJ12) from that year whereas these three variants were placed in the same group in the phylogenetic trees of the E, M and N genes (Figure S3C-E). In all six phylogenetic trees the spotted hyena variants from 2011 and 2012 and the jackal variant from 2011 grouped together (Figures 3 and S3A-E).

### 3.4 | Genetic analyses of the APN gene of wild carnivores

The phylogenetic analysis of 15 partial sequences of the APN gene (1941 nt) we obtained from various carnivore species plus 19 publicly available sequences revealed that the APN of each species was correctly grouped by suborder (Caniformia vs. Feliformia, Figure 4). Four bat sequences clustered together and were closer to carnivore



**FIGURE 2** Similarity plot of the 3'-end of the genome of six "Serengeti" variants. Four variants from spotted hyena (SH) and two from silver-backed jackal (SBJ) against one spotted hyena variant from 2011 (SH143-2011). SH variants include one from 2001 in green, one from 2004 in blue, and two from 2007 in yellow. The SBJ variants include one from 2007 in red and one from 2011 in brown. The locations of gene regions (S1-NTD, S1-CTD, S2, 3a, 3c, E, M, N) are indicated above the plot. The location of the receptor binding region (RBD) is indicated above the X axis. The plot was constructed with genes that did not have major deletions (see Figure 1)



**FIGURE 3** The phylogenetic relationships of “Serengeti” variants and other *Alphacoronavirus-1* variants based on the S1-CTD domain of the S gene (1161 nt) and the distribution across lineages of two residues (524, 525) under positive selection. The “Serengeti” variants from spotted hyenas (SH) are in pink and those from silver-backed jackals (SBJ) are in blue. For each variant, the coronavirus genotype, variant name, host, country of origin, year of collection and Genbank accession number are quoted. Numbers at the branches indicate bootstrap percentage values from 1000 replicates. Branch colour indicate evidence of significant episodic positive selection (EBF >50) for site 524 (in orange) and selective sweeps (R → L, EBF >100) for site 525 (in green)

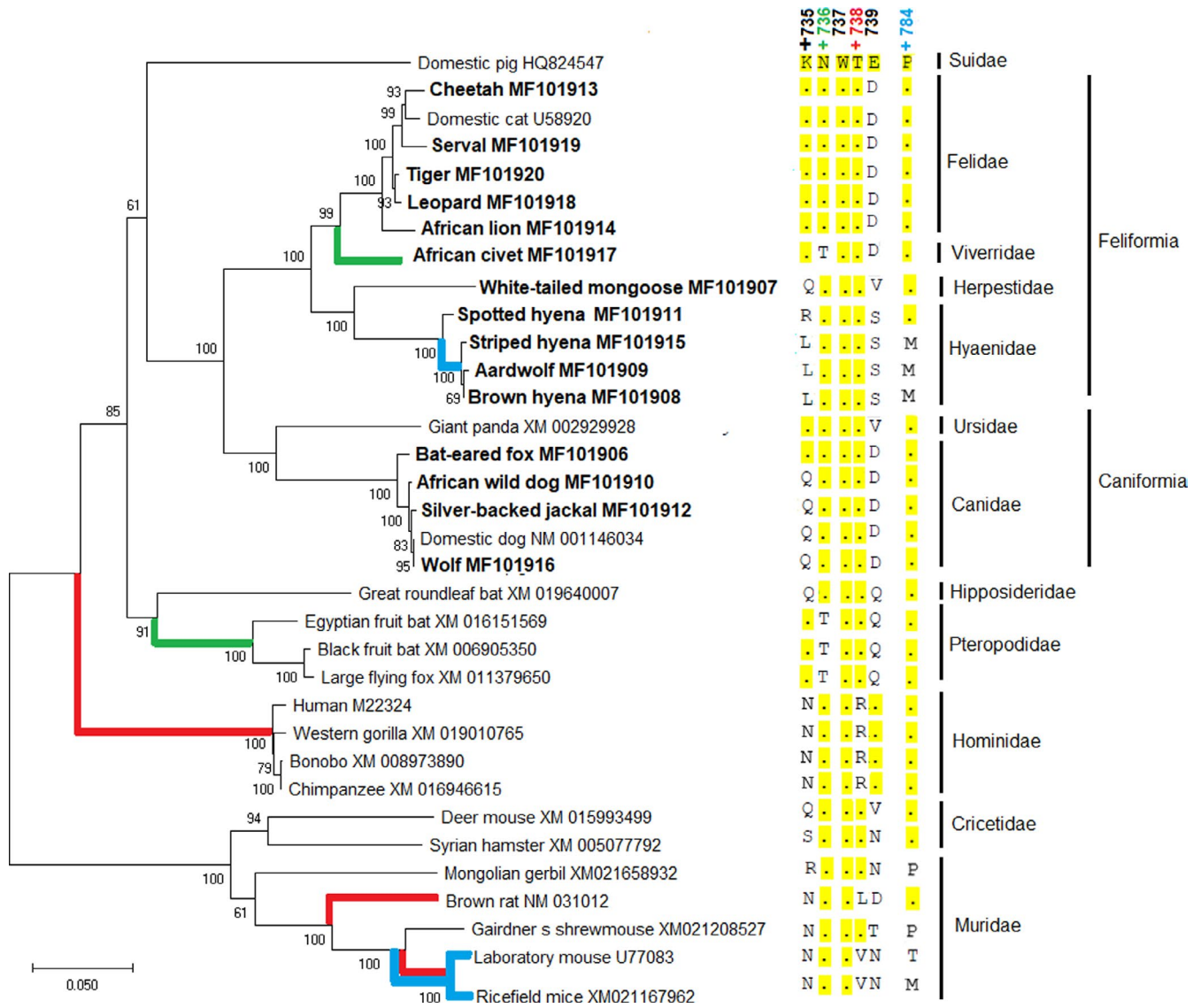
than primate sequences, and seven rodent sequences clustered together and separate from carnivores, bats and primates (Figure 4).

Pairwise alignment of APN amino acid sequences from carnivores (15 sequences obtained from this study plus publically available sequences for giant panda, domestic dog and domestic cat) revealed an overall similarity of 92.7%. The similarity of APN sequences within carnivore families was high (Felidae 97.6%, Canidae 99.2%, Hyaenidae 99.6%). The similarity between the spotted hyena and jackal complete APN was 88%. The similarity between APN sequences from carnivores and domestic pig was 86.7%. A comparison of the 23 residues in pAPN that directly bind with PRCV (Table S1, Reguera et al., 2012) with the APN of carnivores showed that 11 of these residues were identical in all species (Figure 5). Of the 12 residues that differed between carnivore species, two were highly variable (residues 735 and 739 with four different residue variants, Figures 5 and 6a). Most sites (18/23) that interact with the viral S protein were identical in spotted hyenas and jackals (Figures 5 and 6a).

The analyses of sites under positive selection in the APN protein from 34 mammalian species revealed that eight sites within the virus binding region were under positive selection (Table 1).

Four of these sites (728, 732, 742, 770) were adjacent to the 23 residues known to directly bind with the virus RBD and all differed between domestic pigs and carnivores (Figure 5 in cyan blue). The other four sites (735, 736, 738, 784) were among the 23 residues that directly interact with PRCV. Residue 735 was the most variable between carnivores (four residues were observed, Figure 5) and the only residue under positive selection that differed between spotted hyenas (Arginine, R) and jackals (Glutamine, Q). Residue 736 was under episodic diversifying selection (Figure 4) in the African civet and in the branch leading to Old World fruit bats (family Pteropodidae). Both had a different residue (Threonine, T) from the other species (Asparagine, N). Residue 738 was T in all species except in lineages where it was under episodic diversifying selection, including the Hominidae group where it encoded R, and three species in the family Muridae where it encoded Leucine (L) in the brown rat and Valine (V) in the laboratory and ricefield mouse (Figure 4). Residue 784 was under episodic diversifying selection and encoded proline (P) in all species except for the branch to the aardwolf, striped hyena and brown hyena, where it encoded Methionine (M), and the branch leading to the laboratory (T) and the ricefield mouse (M).





**FIGURE 4** The phylogenetic relationships of the APN gene from 34 mammalian species and the distribution across lineages of five residues (735–738, 784) critical for the interaction between APN and the *Alphacoronavirus-1* virus receptor binding region, four of which are under positive selection. Maximum likelihood tree under the HKY85 + G + I model of a segment of 1941 nucleotides (nt) of the APN nt sequence. The fifteen carnivore APN sequences obtained by this study are in bold letters. The family for each species is indicated to the right of the tree, followed by the suborder to which each carnivore species belongs. Numbers at the branches indicate bootstrap percentage values from 1000 replicates. Accession numbers of each sequence are specified after each species name. Residue 735 is under positive selection. Residues 736, 738 and 784 are under episodic diversifying selection. Branches in green, red and light blue indicate where in the tree are residues 736, 738 and 784 under episodic diversifying selection, respectively. On the right of the tree is the alignment of these five residues (735–738, 784). At the top of the alignment is the sequence of these residues in porcine APN (pAPN). Residues from other mammalian species APN which are identical to those in the domestic pig APN are highlighted in yellow, residues that differed are labelled with the respective amino acid. The plus signs indicate the four residues found under positive selection

### 3.5 | Positive selection in *Alphacoronavirus-1* structural genes

The RBD of *Alphacoronavirus-1* includes 147 amino acids within the S1-CTD of the S protein (residues 507 to 654). Average similarity of RBD sequences from our 10 Serengeti variants was 89.3%. Average similarity of RBD sequences of these Serengeti variants with 55 variants worldwide (Table S4), including FCoVII, CCoVII, TGEV, PRCV and CCoVIIb variants, was 86.5%. Using the 19 residues in PRCV

that interact with pAPN (Table S1) as reference, our 10 complete S protein sequences had 11 residues identical to those of the reference sequence, which included the four residues (527, 528, 530, 571) considered essential for receptor binding (Figure 7). These four residues were also identical in the other 55 variants from different hosts (Figure 7). The eight remaining residues differed from the reference sequence and included residues 524 (Table 2) and 525 which were under some type of selection. Residue 524 was under episodic diversifying selection (Table 2) and an analysis of directional

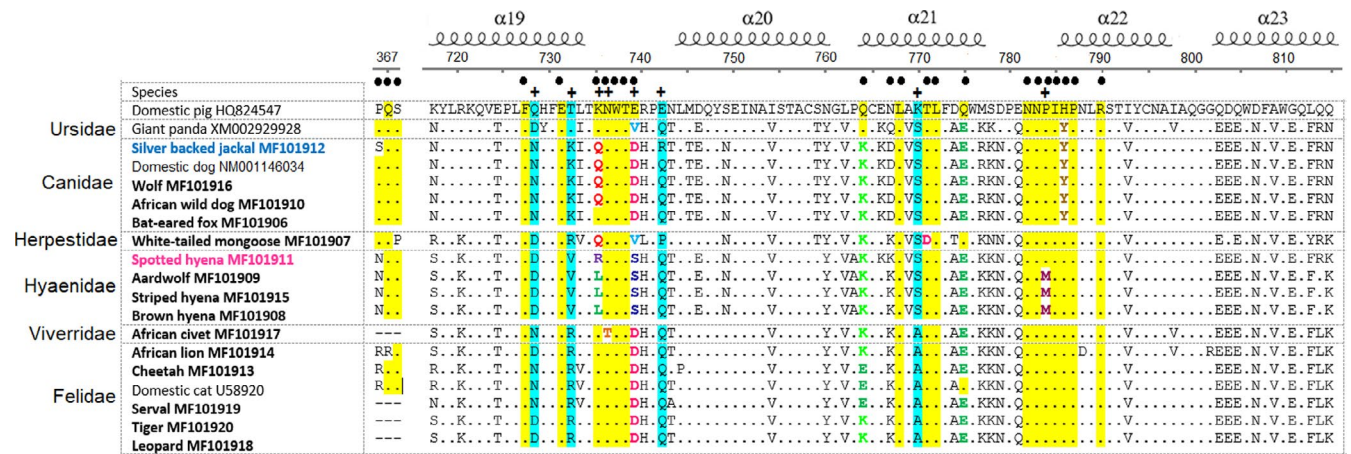


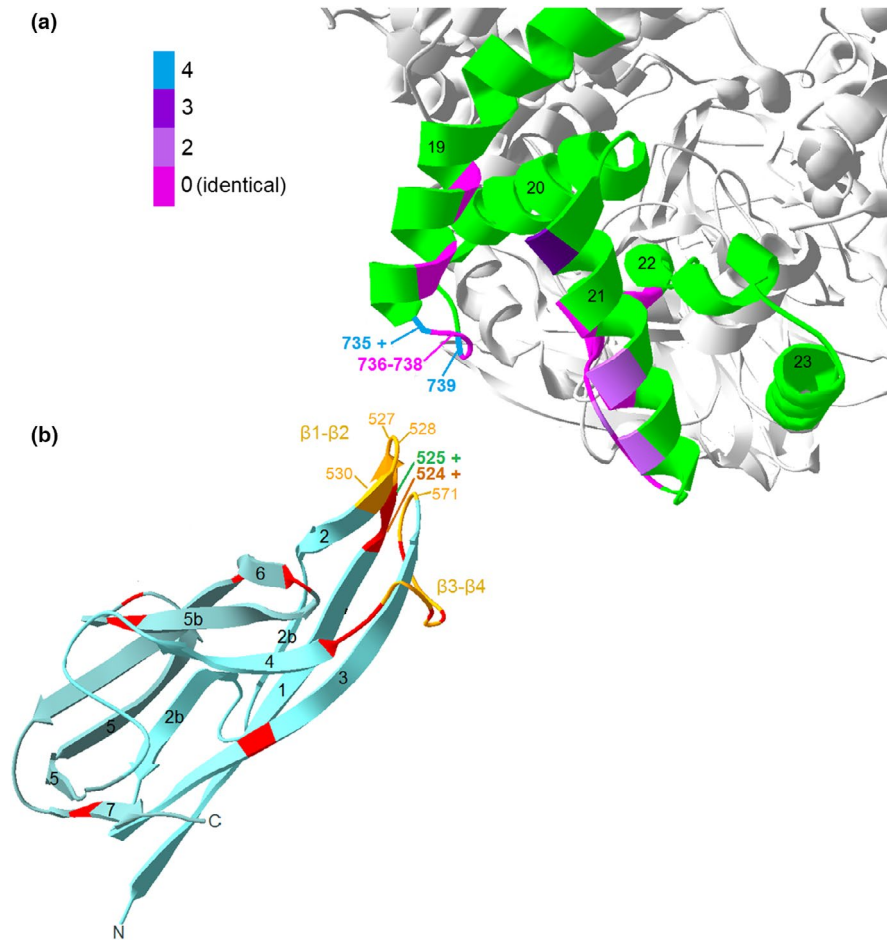
FIGURE 5 Alignment of the region in the porcine APN (pAPN) protein known to bind to *Alphacoronavirus-1* of APN sequences from 18 carnivore species and the domestic pig. In bold are the 15 sequences obtained in this study, including four that are incomplete (African civet, serval, tiger, leopard) and those of the spotted hyena (in pink) and the silver-backed jackal (in blue). Above the alignment is the representation of the tertiary structure (alpha helices) of pAPN, the black circles below indicate the 23 residues in pAPN that directly contact PRCV. Residues highlighted in yellow are identical to the ones observed in pAPN and residues that differ from those in pAPN are in a different colour to increase clarity. Plus signs indicate the eight residues under positive selection. Highlighted in cyan are four residues under positive selection that are not among the 23 residues in pAPN that interact with the porcine *Alphacoronavirus-1* PRCV

evolution in protein sequences detected an elevated substitution rate towards L in residue 525 (R → L, EBF >100). Mapping these amino acid substitutions to the *Alphacoronavirus-1* S1-CTD phylogeny showed that most substitutions occurred in branches which included spotted hyena variants (Figure 3). Substitution of Lysine (K) to Q in residue 524 (K524Q) occurred in all “Serengeti” variants apart from variants from both carnivore species in 2007. This substitution also occurred in a Chinese ferret badger variant in China in 2003 (EF192156). Substitution of R to leucine (L) in residue 525 (R525L) occurred in all Serengeti variants except variants SB12-2007 and SH36-2004 (Figure 3). This substitution was also present in variant FCoV7 79-1146 from the USA in 1979 (AY994055). Eleven additional residues were detected under episodic positive selection within the C-terminal region of the RBD (Table 2) but none occurred in the region that directly interacts with the receptor (Figures 6b and 7). Five of these residues (574, 577, 580, 584) were highly variable (Figure 7) and located in the β3–β4 loop region in the tertiary structure of the RBD (Figure 6b).

The analysis of sites under positive selection in S domains S1-NTD and S2 and structural genes M, N, E revealed additional sites under positive or episodic selection (Table 2). In the E protein only one residue (32) was under episodic diversifying selection (Table 2) and identical in all “Serengeti” variants (L) except for variant SH36-2004 (M). All eight sites under positive selection in the M gene were located in the hypervariable region of the gene (the first 110 nucleotides encoding 37 amino acids, Figure 2). Three of these residues (23K, 26S, 28T) were identical in all Serengeti variants except for SH36-2004 (23Q, 26S, 28I). Three other residues (27D, 35T, 36N) were identical in all Serengeti variants except for variants SH36\_2004 (27G, 35V, 36T) and SH32\_2001 (27E, 35A, 36A).

## 4 | DISCUSSION

Long-term, noninvasive monitoring of enteric *Alphacoronavirus-1* infection in spotted hyenas in the Serengeti NP revealed that infection prevalence fluctuated across years (Figure S1). In line with Goller et al. (2013), we found that virus is primarily shed by juvenile spotted hyenas which therefore play a key role for within-clan transmission, as is the case in canine distemper virus (Nikolin et al., 2017) but not sapovirus (OlarTE-Castillo et al., 2016). The rarity of *Alphacoronavirus-1* infection in adult spotted hyenas, coupled with their high seroprevalence (74%) against infection (East et al., 2004) suggests high herd immunity among adult spotted hyena against future infection. Immunity following recovery from infection may persist throughout life and be boosted by repeated contact with infected juveniles or virus contaminated faeces, particularly at communal dens. We detected two peaks of infection prevalence in juveniles (Figure S1), consistent with the dynamics of immunizing infections in which an outbreak of infection declines as the level of herd immunity rises (Kuiken et al., 2006). Although spotted hyenas give birth throughout the year (Hofer & East, 2003), susceptible individuals do not apparently enter the population sufficiently frequently to prevent the burn-out of specific variants and their replacement by other variants. Our results (Figures S2 and S3A–E) suggest that *Alphacoronavirus-1* variants persist in Serengeti spotted hyena population for one to several years before replacement. High herd immunity among adults probably protects susceptible young at communal dens against the spread of infection from other clans, but this may not prevent infection between jackals and juvenile spotted hyenas, and possibly other wild carnivores.



**FIGURE 6** Model of the tertiary structure of porcine APN (pAPN) coupled with the RBD of the porcine *Alphacoronavirus-1* PRCV. The alpha helices and beta barrels are numbered in each structure. (a) For the pAPN structure, the region known to interact with PRCV is highlighted in green and the rest of the structure is in light grey. The specific residues known to interact with PRCV are coloured according to the number of different residues observed between APN of carnivore species included and the domestic pig (19 sequences in total), i.e., 0 indicates that the residue at a site was identical in all species and four indicates four different residues occurred in the species examined. Most residues were identical in all sequences (0, in magenta), including residues 736 to 738 (indicated in the structure) which are essential for the interaction with PRCV. Residue 735 was detected to be under positive selection and together with residue 739 was the most variable residue (4, light blue). (b) For the RBD structure, the two protruding regions that directly interact with pAPN ( $\beta 1$ - $\beta 2$ ,  $\beta 3$ - $\beta 4$ ) are marked in yellow. Within these two regions the four most important residues that interact with pAPN (527, 528, 530, 571) are shown and are identical in all variants studied. Residues detected under positive selection are coloured in red and most are outside the regions that directly interact with the receptor. Residues 524 (orange) and 525 (green) which were under some kind of selection in most Serengeti hyena variants are also shown

#### 4.1 | Genetic diversity of Serengeti variants

Genetic diversity in CCoV variants obtained between 2001 and 2012 from spotted hyenas in the Serengeti NP was considerable. The genetically most distinct variants were from spotted hyenas in 2004 (SH36 and SH42, Figure S2). We could only sequence the complete 3'-end of SH36-2004, not SH42-2004. This variant did not group with other Serengeti variants (Figures 3 and S3A-E). In the S1-NTD and S1-CTD phylogenetic trees, SH36-2004 was the only Serengeti variant grouped with FCoV variants, close to variants from wild carnivores in China (Chinese ferret badger and raccoon dog, Figures 3 and S3A). In the M gene phylogenetic tree it grouped with CCoV8b variants (Figure S3D). Considering the other structural genes (S2 domain, E, N) SH36-2004 was not closely related

to any CCoV or FCoV variants (Figure S3B, C, E), and regions within the 3c and M genes were highly dissimilar to those in other Serengeti variants (Figure 2). It was also the only Serengeti variant with a deletion in the M gene. Most amino acid sites under positive selection in the structural genes, including most sites detected in the M gene, were identical in all spotted hyena variants except for SH36-2004 (Table 2). The deletion in the M gene and the sites under positive selection were located in the hypervariable region of the M protein (Figure 2), a hydrophilic tail exposed on the surface of the virion (Locker et al., 1992; Risco et al., 1995; Rottier, 1995). Mutations in this region of the M protein affects variant virulence (Charley & Laude, 1988; Pratelli et al., 2002; Sánchez-Morgado et al., 2004). The highly divergent SH36-2004 variant was collected in the year following a substantial peak of infection in spotted hyenas in

Model	G	$\omega$ values	Sites detected to be under positive selection
M0 versus M3	1505,18	0.4059	325N*, 334P*, 527R**, 591V**, 602S*, 685G*, <b>728Q**</b> , <b>732T*</b> , <b>735*</b>
M1a versus M2a	195,22	0.4105	
M7 versus M8	338,96	0.3113	
MEME	NA	NA	325N**, 334P**, 455E*, 469S*, 503A**, 510L**, 530D*, 592M* 594D**, 595H*, 601V* 604A**, 628Q*, 641H**, 645T**, 669H*, 684N*, 701Y**, <b>728Q*</b> , <b>736N*</b> , <b>738 T*</b> , <b>742E**</b> , <b>770K**</b> , <b>784P**</b> , 827S*, 850L*, 858S**, 871P*

TABLE 1 Sites in the APN protein under positive selection according to the nested maximum likelihood site models and the MEME model

Only sites with posterior probabilities (P) >95% (\*) and >99% (\*\*) were considered. For the nested maximum likelihood models, only sites under positive selection by all three alternative maximum likelihood models are shown. The respective residue seen in the domestic dog sequence is shown next to the residue number. In bold and underlined are the residues that are within the virus-binding region. The numbering of sites detected under positive selection is based on the pAPN sequence.

G, log-likelihood ratio test statistic;  $\omega$  values, average values.



FIGURE 7 Sequence "logo" plots of the 19 residues (in bold) within the *Alphacoronavirus-1* RBD region known to interact with the receptor and the additional 11 residues detected to be under episodic positive selection (indicated with red plus signs). (a) A logo based on the 10 CCoV S gene sequences obtained in this study (eight variants from the spotted hyena and two from the silver-backed jackal). (b) A logo based on 55 *Alphacoronavirus-1* sequences including TGEV, PRCV, CCoV, CCoV2b and FCoV (Table S4). The reference porcine *Alphacoronavirus-1* PRCV sequence is shown between the plots. In bold and underlined are the residues that directly interact with the host APN. The four residues indicated with yellow triangles are considered center residues for binding of the virus to its host receptor and these were identical in all variants. Red crosses indicate sites under positive selection. The overall height of each letter is proportional to sequence conservation as measured in bits. In each position the residue letters are ordered from the most to the least frequent. Polar residues are in black, acidic residues are in blue, basic residues are in green and nonpolar residues are in red. The sequence logos were drawn using software WEBLOGO (weblogo.berkeley.edu)

2003 (Figure S1) and may have contributed to this outbreak. Further research might indicate which functions are served by the unique residues under positive selection in this variant. In addition, the high

infection prevalence in 2003 may also be the consequence of the cocirculation of two different variants, SH36-2004 and SH42-2004, with very different M genes (Figure S2).

**TABLE 2** Sites in the structural genes of *Alphacoronavirus-1* proteins under positive selection according to the nested maximum likelihood models and MEME model

Gene	Model	G	$\omega$ values	Sites under positive selection
S1-NTD	M0 versus M3	631.8	0.22	129N**, 150S**, 153T**, 154N**, 158D*, 225H*, 272V**
	M1a versus M2a	44.13	0.29	
	M7 versus M8	60.30	0.23	
S1-NTD	MEME	NA	NA	5V**, 6T**, 18A**, 23D**, 79Y*, 94E*, 132H*, 144T*, 153T**, 154N**, 158D**, 218S**, 272V**
S1-CTD	M0 versus M3	295.2	0.14	No sites under positive selection by all three alternative models
	M1a versus M2a	0	0.19	
	M7 versus M8	0	0.17	
S1-CTD	MEME	NA	NA	368S*, 438D**, 485Q*, 492A**, 508S**, 510Y**, 514S**, <b>524K*</b> , <b>562 V**</b> , <b>574I**</b> , <b>577R**</b> , <b>580T**</b> , <b>584D**</b> , <b>585A*</b> , <b>617V**</b> , <b>622K**</b> , <b>631T**</b> , <b>636V**</b> , <b>650V**</b>
S2	M0 versus M3	241.5	0.067	No sites under positive selection by all three alternative models
	M1a versus M2a	0	0.082	
	M7 versus M8	0		
S2	MEME	NA	NA	14V*, 339T*, 450T*
E	M0 versus M3	11.2	0.16	No sites under positive selection by all three alternative models
	M1a versus M2a	0	0.17	
	M7 versus M8	0	0.16	
E	MEME	NA	NA	32L*
N	M0 versus M3	304.53	0.17	74I**
	M1a versus M2a	5.88	0.20	
	M7 versus M8	24.48	0.17	
N	MEME	NA	NA	2A*, 3N**, 29N*, 43K**, 45S*, 75R**, 147V*, 170S*, 182S*, 217N**, 251S**, 265S*, 342S*, 345A*, 348S*
M	M0 versus M3	736.5	0.2	23K**, 26S**, 27D*, 28T**, 29S**, 35T**, 36N**, 37G**
	M1a versus M2a	50.3	0.24	
	M7 versus M8	75.45	0.2	
M	MEME	NA	NA	3I*, 29S*

Only sites with posterior probabilities (P) >95% (\*) and >99% (\*\*) were considered. For the nested maximum likelihood models, only sites under positive selection by all three alternative maximum likelihood models are shown. All sites under positive selection are numbered according to the reference variant SBJ12-2007 and the respective residue observed in this variant is shown. Bold and underlined, residues within the receptor-binding region; G, log-likelihood ratio test statistic.  $\omega$  values, average values.

Our comparative analysis of the 3'-end of the virus genome showed that the 2007 jackal variant was the only Serengeti variant without major deletions, whereas regardless of year, all spotted hyena variants had major deletions in the nonstructural genes 3b, 7a and 7b (Figure 1). Spotted hyena variants in 3 years (2001, 2004, 2007) had three different large deletions in the 7b gene. As most reported FCoV field variants either have an intact 7b gene or only small deletions (Lin et al., 2009), our results provide, to our

knowledge, the first evidence of major deletions in the 7b gene in field variants from free-ranging wild carnivores. We also report the occurrence of three large deletions in the 3b gene in spotted hyena variants from 2001, 2004, 2011 and 2012, two of which resulted in premature stops (in variants from 2004, 2011, 2012, Figure 1). The effect of deletions in nonstructural genes 3b and 7b on virulence and pathogenesis is variable, as the emergence of virulent or attenuated CCoV, FCoV, or TGEV variants is not always associated with

such deletions (Chang et al., 2010; Herrewegh et al., 1995; Lin et al., 2009; McGoldrick et al., 1999; Pedersen et al., 2009). Variants with deletions in nonstructural genes circulated in years of high and low prevalence, so these deletion events were not linked to infection prevalence in the spotted hyena population. As the 3c genes in all our variants were intact, replication in the intestine and hence faecal transmission was ensured (Pedersen et al., 2009, 2012).

## 4.2 | Lack of recombination in Serengeti variants

Although recombination has been widely described in *Alphacoronavirus-1* variants infecting domestic animals (Baric et al., 1997; Decaro et al., 2009; Herrewegh et al., 1998; Lai et al., 1985; Terada et al., 2014), our extensive search for recombination breakpoints revealed that none of our variants had a recombinant origin. The sequencing method we applied would have detected the concurrent presence of viral sequences with a divergence as great as 13% from a known sequence within a sample (Maricic et al., 2010; Meyer & Kircher, 2010). Thus, despite the ability to detect sequences from different variants within one sample we found no evidence of coinfection. Recombination events require concurrent infection with more than one coronavirus variant and it may be that this rarely occurs in wild juvenile spotted hyenas.

## 4.3 | *Alphacoronavirus-1* RBD

Virus attachment to the host cell receptor requires direct binding between specific viral residues and those of the host cell receptor, plus other interactions between residues situated close to these direct binding sites (Allison et al., 2012; Govindasamy et al., 2003; Li, 2016). Phylogenetic analyses have been used to discover residues under positive or episodic diversifying selection associated with virus–host adaptations for host cell entry, and other factors such as virulence (McCarthy et al., 2007; Nikolin et al., 2017). We used the crystal structure of the RBD region of PRCV bound to pAPN (Reguera et al., 2012) as the foundation for our molecular investigation of *Alphacoronavirus-1* adaptations to carnivore APN. Reguera et al. (2012) reported that two regions,  $\beta 1$ – $\beta 2$  and  $\beta 3$ – $\beta 4$ , of the PRCV RBD directly bind to pAPN, and that within these regions three residues in  $\beta 1$ – $\beta 2$  (527, 528, 530) and one in  $\beta 3$ – $\beta 4$  (571) are essential. The RBD in our wild carnivore variants was the most variable region within the S1-CTD (Figure 2). However, the four essential residues (G527, Y528, Q530, W571) were identical in all 65 *Alphacoronavirus-1* sequences from various hosts (Table S4), including the 10 Serengeti variants (Figure 7a). As substitutions of these four essential amino acids completely abolishes virus binding to the host receptor (Reguera et al., 2012), these four residues should be under strong negative selection. In comparison, most sites under episodic diversifying selection (Figures 6b and 7) occurred in a highly variable region in the RBD, which is not in direct contact with the receptor and was located in  $\beta 3$ – $\beta 4$  (Figure 6b). The RBD is the most

antigenic region of *Alphacoronavirus-1* (Reguera et al., 2012) and thus its variability could result from immune-mediated positive selection (Reguera et al., 2011). Further research is needed to uncover the extent of selection pressure generated by immune processes on this region of the RBD. Additionally, in the region adjacent to  $\beta 1$ – $\beta 2$ , one site was under episodic (524), another (525) under directional selection (Figure 6b). Most amino acid substitutions at these sites map to variants from spotted hyenas (Figure 3), indicating that these changes at receptor-binding sites may be associated with the adaptation of these variants to the spotted hyena receptor. Similarly, in SARS-CoV, natural mutations in two residues under positive selection enhanced the binding affinity of the SARS-CoV RBD to the human receptor, which was important for civet to human transmission (Li, 2013; Wu et al., 2012).

## 4.4 | Virus–host interface

Our results demonstrated a high similarity (92.7%) in sequences of the host receptor APN gene across both suborders of Carnivora. The three most important residues in pAPN for attachment to PRCV are N736, W737, T738 (residues 740–742 in fAPN) which form an N-linked glycosylation site (Reguera et al., 2012; Tusell et al., 2007). Once this glycosylation site was mutated in fAPN to avoid the glycosylation consensus sequence (N/X/T or S, where N is asparagine, X is any amino acid, T is threonine and S is serine), the resulting mutated fAPN did not have receptor activity for FCoV, CCoV or TGEV (Tusell et al., 2007). In all carnivore APNs (included those reported in this study) except for the African civet, these residues were identical (N736, W737, T738) to pAPN and fAPN (Figure 5). This could explain why the four interacting residues in the virus were conserved even in variants from widely different hosts. This is not the case for APN in all mammalian species, as residue 736 (740 in fAPN) was under episodic diversifying selection in the APN of Old World fruit bats and the APN of the African civet, all of which encoded T instead of N (Figure 4). This suggests that Old World fruit bats and the African civet may not be susceptible to infection by FCoV/CCoV/TGEV/PRCV. This suggestion is consistent with the finding that kidney cell lines from the Egyptian fruit bat (*Rousettus aegyptiacus*) and other Old World fruit bats are not susceptible to infection with TGEV (Hoffmann et al., 2013). Thus the substitution of amino acids at a critical site (N736T) in APN may reduce susceptibility to infection by TGEV and possibly other genetically similar alphacoronaviruses.

We found evidence that residue T738 in pAPN (T742 in fAPN) is under positive selection in human APN (R738), and also in both the brown rat APN and the laboratory and ricefield mouse APNs (V738, Figure 4). These results are in line with experimental results in which fAPN residue T742 was substituted for V742 or R742 (as in the laboratory mouse APN and human APN, respectively) which inhibits TGEV, FCoV and CCoV binding to the mutated fAPN receptors (Tusell et al., 2007). This also helps to explain why mouse and human APN do not naturally support infection with these viruses (Wentworth & Holmes, 2001). In *Betacoronavirus*,

signatures of positive selection in key residues that modulate virus–host interactions have been reported in a bat lineage for the receptor of SARS-CoV, the ACE2 protein (Demogines et al., 2012), and the receptor for Middle East Respiratory Syndrome (MERS)-CoV, the DPP4 protein (Cui et al., 2013). The occurrence of R742 in hominids and V742 in two rodent species may indicate positive selection on the APN gene as a result of past interactions of these APNs with an ancestral virus similar to those in the CCoV/FCoV/TGEV/PRCV group. Thus, the results of our study provide further evidence that CoV-driven selection pressures can drive genetic changes in host genes that control susceptibility to infection (Sironi et al., 2015).

Most residues in APN involved in virus binding were conserved in carnivore APNs (Figure 5), only two residues were highly variable (735, 739, Figure 6a). One of these residues (735) was under positive selection. As residues 735 and 739 are highly variable, and located next to the three residues most important for virus binding (736–738) we suggest that residues 735 and 739 may affect the binding affinity between carnivore *Alphacoronavirus-1* variants infecting carnivores and carnivore APN. Even if changes in both residues may not completely abrogate receptor activity, they may affect binding efficacy of different variants (Tusell et al., 2007).

Considering in more detail the binding of *Alphacoronavirus-1* to spotted hyena and jackal APNs, we note that most APN amino acids which interact with *Alphacoronavirus-1* were conserved (18 of 23 were identical) in both hosts. However, both residues 735 and 739 differed between spotted hyena (735R, 739S) and jackal (735Q, 739D) APNs (Figure 5). These residues were identical in all four extant species in the Hyaenidae (735R, 739S) and as 735Q, 739D in four of five canids examined, the exception being bat-eared foxes (735 K, 739D). Virus adaptation to these differences in host APNs may partly explain the detection of two spotted hyena variants (SH110-2007, SH33-2007) in 2007 encoding L at site 525 (under positive selection, Figure 3) in the RBD and one jackal variant (SBJ12-2007) encoding R at this site, normally present in domestic dog variants (Figure 3). Together with the substantial differences in the 3'-end of the genome of the hyena and jackal variants in 2007 (Figure 1, Figure S2) and the separate placement of the 2007 jackal variant from the 2007 hyena variants (Figure 3), our results support the idea that genetically distinct variants circulated in spotted hyenas and jackals in the Serengeti in 2007 (Goller et al., 2013).

Despite evidence of variant adaptations to specific carnivore hosts, we also have evidence that genetically similar variants circulated in Serengeti spotted hyenas (SH89-2011, SH143-2011) and jackals (SBJ3-2011) in 2011 (Figures 1 and S1). All three variants encoded 524Q, 525L rather than the commonly encoded 524 K, 525R in CCoV variants worldwide (Figure 3). These two substitutions (524Q, 525L) may be adaptations to spotted hyena APN, in which case a variant with these substitutions was able to infect a jackal. Alternatively, these substitutions could be associated with generalist traits that permit *Alphacoronavirus-1* variants to infect a taxonomically wide range of carnivore species but not necessarily with high

efficacy. A low binding efficacy between circulating variants and spotted hyena APN may have contributed to low infection prevalence in the spotted hyena population in 2011 and 2012 (Figure S1). Whether substitutions 524Q, 525L influence the ability of variants to infect specific carnivore species requires further study.

Our study substantially extends knowledge of the APN receptor in a taxonomically broad range of wild carnivores, provides a comprehensive genetic analysis of the virus–host entry mechanism and details important residues that may be required for an optimal interaction between the viral S protein RBD and the APN receptor in wild carnivores. Long-term monitoring of spotted hyenas revealed outbreaks of infection associated with genetically distinct variants, the importance of juveniles in virus spread and likely immunity against infection in adults. Despite long-term monitoring, the number of *Alphacoronavirus-1* variants we describe is relatively small, which limits the interpretation of our molecular data.

## ACKNOWLEDGEMENTS

We are grateful to the Commission for Science and Technology of Tanzania (COSTECH), the Tanzania Wildlife Research Institute (TAWIRI) and Tanzania National Parks (TANAPA) for their support of our research, and three anonymous reviewers for their insightful comments. We thank the Deutsche Forschungsgemeinschaft (GRAKO 1121, Genetic and Immunological Determinants of Pathogen-Host Interactions) and the Leibniz-Institute for Zoo and Wildlife Research (IZW) for financial support. We thank Dagmar Thierer, Katja Goller, Malvina Andris, Nelly Boyer, and Annie Francis for assistance and Robert Fyumagwa (TAWIRI), Oliver Höner, Claudia Szentiks, Bettina Wachter (IZW), and Ingrid Wiesel (Brown Hyena Research Project, Namibia), for providing samples. This study was approved by the Internal Committee for Ethics and Animal Welfare of the Leibniz-Institute for Zoo and Wildlife Research, approval number 2011-12-02. Open access funding enabled and organized by Projekt DEAL.

## AUTHOR CONTRIBUTIONS

X.A.O.-C., M.L.E., H.H. conceived and designed the study. M.L.E., H.H. contributed to fieldwork. X.A.O.-C., J.F.R., A.D.G., S.K. contributed to laboratory analyses, X.A.O.-C., F.H. and S.K. curated data. X.A.O.-C. undertook data analysis, M.L.E., H.H. raised the student-ship funding for X.A.O.-C. X.A.O.-C., M.L.E. wrote the manuscript. All authors discussed the results and contributed to the manuscript.

## DATA AVAILABILITY STATEMENT

Nucleotide sequence data for CCoV variants and APN sequences have been deposited in GenBank: complete 3'-end CCoV sequences (MF095847–MF095855), complete sequence of the S, N and 7a genes (KX156832, MW505904, MW505905, respectively), partial M gene sequences (KX156803–KX156821), partial S gene sequences (KX156822–KX156831), complete APN genes (MF101906–MF101916), incomplete APN genes (MF101917–MF101920) (accession numbers are also quoted on Figures S3–S5 and 3, Table S2). The variants of *Alphacoronavirus-1* used in the different analyses are quoted in Table S3.

## ORCID

Ximena A. Olarte-Castillo  <https://orcid.org/0000-0001-5310-9622>

Heribert Hofer  <https://orcid.org/0000-0002-2813-7442>

Alex D. Greenwood  <https://orcid.org/0000-0002-8249-1565>

Marion L. East  <https://orcid.org/0000-0002-0391-8365>

## REFERENCES

- Alfano, F., Dowgier, G., Valentino, M. P., Galiero, G., Tinelli, A., Decaro, N., & Fusco, G. (2019). Identification of pantropic canine coronavirus in a wolf (*Canis lupus italicus*) in Italy. *Journal of Wildlife Diseases*, 55, 504–508. <https://doi.org/10.7589/2018-07-182>
- Allison, A. B., Harbison, C. E., Pagan, I., Stucker, K. M., Kaelber, J. T., Brown, J. D., Ruder, M. G., Keel, M. K., Dubovi, E. J., Holmes, E. C., & Parrish, C. R. (2012). The role of multiple hosts in the cross-species transmission and emergence of a pandemic parvovirus. *Journal of Virology*, 86, 865–872. <https://doi.org/10.1128/JVI.06187-11>
- Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37–48.
- Baric, R. S., Yount, B., Hensley, L., Peel, S. A., & Chen, W. (1997). Episodic evolution mediates interspecies transfer of a murine coronavirus. *Journal of Virology*, 71, 1946–1955.
- Benbaccer, L., Kut, E., Besnardeau, L., Laude, H., & Delmas, B. (1997). Interspecies aminopeptidase-N chimeras reveal species-specific receptor recognition by canine coronavirus, feline infectious peritonitis virus, and transmissible gastroenteritis virus. *Journal of Virology*, 71, 734–737.
- Chang, H.-W., de Groot, R. J., Egberink, H. F., & Rottier, P. J. (2010). Feline infectious peritonitis: Insights into feline coronavirus pathobiogenesis and epidemiology based on genetic analysis of the viral 3c gene. *Journal of General Virology*, 91, 415–420.
- Charley, B., & Laude, H. (1988). Induction of alpha interferon by transmissible gastroenteritis coronavirus: Role of a transmembrane glycoprotein E1. *Journal of Virology*, 62, 8–11.
- Cui, J., Eden, J. S., Holmes, E. C., & Wang, L. F. (2013). Adaptive evolution of bat dipeptidyl peptidase 4 (DPP4): Implications for the origin and emergence of Middle East respiratory syndrome coronavirus. *Virology Journal*, 10, 304. <https://doi.org/10.1186/1743-422X-10-304>
- Damas, J., Hughes, G. M., Keough, K. C., Painter, C. A., Persky, N. S., Corbo, M., Hiller, M., Koepfli, K.-P., Pfenning, A. R., Zhao, H., Genereux, D. P., Swofford, R., Pollard, K. S., Ryder, O. A., Nweeia, M. T., Lindblad-Toh, K., Teeling, E. C., Karlsson, E. K., & Lewin, H. A. (2020). Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proceedings of the National Academy of Sciences of the USA*, 117, 22311–22322. <https://doi.org/10.1073/pnas.2010146117>
- Daszak, P., Cunningham, A. A., & Hyatt, A. D. (2000). Emerging infectious diseases of wildlife – threats to biodiversity and human health. *Science*, 287, 443–449. <https://doi.org/10.1126/science.287.5452.443>
- de Groot, R. J., Cowley, J. A., Enjuanes, L., Faaberg, K. S., Perlman, S., Rottier, P. J., & Gorbalenya, A. E. (2012). Order of *Nidovirales*. In A. King, M. Adams, E. Carstens, & E. J. Lefkowitz (Eds.), *Virus taxonomy, the 9th report of the International Committee on Taxonomy of Viruses* (pp. 785–795). Academic Press.
- Decaro, N., Mari, V., Campolo, M., Lorusso, A., Camero, M., Elia, G., Martella, V., Cordioli, P., Enjuanes, L., & Buonavoglia, C. (2009). Recombinant canine coronaviruses related to transmissible gastroenteritis virus of swine are circulating in dogs. *Journal of Virology*, 83, 1532–1537.
- Decaro, N., Martella, V., Elia, G., Campolo, M., Desario, C., Cirone, F., Tempesta, M., & Buonavoglia, C. (2007). Molecular characterisation of the virulent canine coronavirus CB/05 strain. *Virus Research*, 125, 54–60.
- Delmas, B., Gelfi, J., L'Haridon, R., Vogel, L. K., Sjöström, H., Norén, O., & Laude, H. (1992). Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature*, 357, 417–420.
- Delmas, B., Gelfi, J., Sjöström, H., Norén, O., & Laude, H. (1994). Further characterisation of aminopeptidase-N as a receptor for coronaviruses. *Advances in Experimental Medicine and Biology*, 342, 293–298.
- Demogines, A., Farzan, M., & Sawyer, S. L. (2012). Evidence for ACE2-utilizing coronaviruses (CoV) related to severe acute respiratory syndrome CoV in bats. *Journal of Virology*, 86, 6350–6353. <https://doi.org/10.1128/JVI.00311-12>
- Dye, C., Temperton, M., & Siddell, S. G. (2007). Type I feline coronavirus spike glycoprotein fails to recognize aminopeptidase N as a functional receptor on feline cell lines. *Journal of General Virology*, 88, 1753–1760.
- East, M. L., Moestl, K., Benetka, V., Pitra, C., Höner, O. P., Wachter, B., & Hofer, H. (2004). Coronavirus infection of spotted hyenas in the Serengeti ecosystem. *Veterinary Microbiology*, 102, 1–9.
- Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5, 113.
- Enjuanes, L. (2000). Coronaviridae. In M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carsten, M. K. Estes, S. M. Lemon, H. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, & R. B. Wickner (Eds.), *Virus taxonomy* (pp. 835–849). Academic Press.
- Gallagher, T. M., & Buchmeier, M. J. (2001). Coronavirus spike protein in viral entry and pathogenesis. *Virology*, 279, 371–374. <https://doi.org/10.1006/viro.2000.0757>
- Gao, F. S., Hu, G. X., Xia, X. Z., Gao, Y. W., Bai, Y. D., & Zou, X. H. (2009). Isolation and identification of a canine coronavirus strain from giant pandas (*Ailuropoda melanoleuca*). *Journal of Veterinary Science*, 10, 261–263.
- Gibbs, M. J., Armstrong, J. S., & Gibbs, A. J. (2000). Sister-scanning: A Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, 16, 573–582.
- Goller, K. V., Fickel, J., Hofer, H., Beier, S., & East, M. L. (2013). Coronavirus genotype diversity and prevalence of infection in the Serengeti National Park, Tanzania. *Archives of Virology*, 158, 729–734.
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, 38, W695–W699.
- Govindasamy, L., Hueffer, K., Parrish, C. R., & Agbandje-McKenna, M. (2003). Structures of host range-controlling regions of the capsids of canine and feline parvoviruses and mutants. *Journal of Virology*, 77, 12211–12221.
- Graham, R. L., & Baric, R. S. (2010). Recombination, reservoirs, and the modular spike: Mechanisms of coronavirus cross-species transmission. *Journal of Virology*, 84, 3134–3146.
- Greenwood, A. D., Tsangaras, K., Ho, S. Y. W., Szentiks, C. A., Nikolin, V. M., Ma, G., Damiani, A., East, M. L., Lawrenz, A., Hofer, H., & Osterrieder, N. (2012). A potentially fatal mix of herpes in zoos. *Current Biology*, 22, 1–5. <https://doi.org/10.1016/j.cub.2012.07.035>
- Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modelling. *Electrophoresis*, 18, 2714–2723.
- Hahn, C., Bachmann, L., & Chevreaux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – A baiting and iterative mapping approach. *Nucleic Acids Research*, 41, e129. <https://doi.org/10.1093/nar/gkt371>



- Harrison, T. M., Mazet, J. K., Holekamp, K. E., Dubovi, E., Engh, A. L., Nelson, K., Van Horn, R. C., & Munson, L. (2004). Antibodies to canine and feline viruses in spotted hyenas (*Crocuta crocuta*) in the Masai Mara National Reserve. *Journal of Wildlife Diseases*, *40*, 1–10.
- Hassell, J. M., Begon, M., Ward, M. J., & Fèvre, E. M. (2017). Urbanization and disease emergence: Dynamics at the wildlife–livestock–human interface. *Trends in Ecology and Evolution*, *32*, 55–67.
- Heene, J. L., Evermann, J. F., McKeirnan, A. J., Marker-Kraus, L., Roelke, M. E., Bush, M., Wildt, D. E., Meltzer, D. G., Colly, L., & Lukas, J. (1990). Prevalence and implications of feline coronavirus infections of captive and free-ranging cheetahs (*Acinonyx jubatus*). *Journal of Virology*, *64*, 1964–1972.
- Hegyi, A., & Kolb, A. F. (1998). Characterisation of determinants involved in the feline infectious peritonitis virus receptor function of feline aminopeptidase N. *Journal of General Virology*, *79*, 1387–1391.
- Herrewegh, A. A., Smeenk, I., Horzinek, M. C., Rottier, P. J., & de Groot, R. J. (1998). Feline coronavirus type II variants 79–1683 and 79–1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. *Journal of Virology*, *5*, 4508–4514.
- Herrewegh, A. A., Vennema, H., Horzinek, M. C., Rottier, P. J., & de Groot, R. J. (1995). The molecular genetics of feline coronaviruses: Comparative sequence analysis of the ORF7a/7b transcription unit of different biotypes. *Virology*, *212*, 622–631.
- Hofer, H., & East, M. L. (2003). Behavioral processes and costs of coexistence in female spotted hyenas: A life history perspective. *Evolutionary Ecology*, *17*, 315–331.
- Hoffmann, M., Müller, M. A., Drexler, J. F., Glende, J., Erdt, M., Gützkow, T., Losemann, C., Binger, T., Deng, H., Schwegmann-Weßels, C., Esser, K.-H., Drosten, C., & Herrler, G. (2013). Differential sensitivity of bat cells to infection by enveloped RNA viruses: Coronaviruses, paramyxoviruses, filoviruses, and influenza viruses. *PLoS One*, *8*, e72942.
- Hohdatsu, T., Izumiya, Y., Yokoyama, Y., Kida, K., & Koyama, H. (1998). Differences in virus receptor for type I and type II feline infectious peritonitis virus. *Archives of Virology*, *143*, 839–850.
- Hohdatsu, T., Okada, S., Ishizuka, Y., Yamada, H., & Koyama, H. (1992). The prevalence of types I and II feline coronavirus infections in cats. *Journal of Veterinary Medicine Science*, *54*, 557–562.
- Johnson, C. K., Hitchens, P. L., Pandit, P. S., Rushmore, J., Evans, T. S., Young, C. C. W., & Doyle, M. M. (2020). Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proceedings of the Royal Society of London*, *287*, 20192736.
- Kolb, A. F., Hegyi, A., Maile, J., Hesiter, A., Hagemann, M., & Siddell, A. G. (1998). Molecular analysis of the coronavirus-receptor function of aminopeptidase N. In L. Enjuanes, S. Siddell, & W. Spaan (Eds.), *Coronaviruses and arteriviruses* (pp. 61–67). Springer.
- Kosakovskiy-Pond, S. L., & Frost, S. D. W. (2005). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*, 1208–1222.
- Kosakovskiy-Pond, S. L., Poon, A. F. Y., Leigh Brown, A. J., & Frost, S. D. W. (2008). A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus. *Molecular Biology and Evolution*, *25*, 1809–1824.
- Kuiken, T., Holmes, E. C., McCauley, J., Rimmelzwaan, G. F., Williams, C. S., & Grenfell, B. T. (2006). Host species barriers to influenza virus infections. *Science*, *312*, 394–397.
- Kuo, L., Godeke, G.-J., Raamsman, M. J. B., Masters, P. S., & Rottier, P. J. M. (2000). Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: Crossing the host cell species barrier. *Journal of Virology*, *74*, 1393–1406. <https://doi.org/10.1128/JVI.74.3.1393-1406.2000>
- Lai, M. M., Baric, R. S., Makino, S., Keck, J. G., Egbert, J., Leibowitz, J. L., & Stohlman, S. A. (1985). Recombination between nonsegmented RNA genomes of murine coronaviruses. *Journal of Virology*, *56*, 449–456.
- Lau, S. K. P., Woo, P. C. Y., Li, K. S. M., Huang, Y., Tsoi, H.-W., Wong, B. H. L., Wong, S. S. Y., Leung, S.-Y., Chan, K.-H., & Yuen, K.-Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences USA*, *102*, 14040–14045.
- Le Poder, S. (2011). Feline and canine coronaviruses: Common genetic and pathobiological features. *Advances in Virology*, *2011*, 609465.
- Levis, R., Cardellachio, C. B., Scanga, C. A., Compton, S. R., & Holmes, K. V. (1995). Multiple receptor-dependent steps determine the species specificity of HCV-229E infection. *Advances in Experimental Medicine and Biology*, *380*, 337–343.
- Li, F. (2012). Evidence for a common evolutionary origin of coronavirus spike protein receptor binding subunits. *Journal of Virology*, *86*, 2856–2858. <https://doi.org/10.1128/JVI.06882-11>
- Li, F. (2013). Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Research*, *100*, 246–254.
- Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annual Reviews of Virology*, *3*, 237–261.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, *25*, 1754–1760.
- Lin, C. N., Su, B. L., Huang, H. P., Lee, J. J., Hsieh, M. W., & Chueh, L. L. (2009). Field strain feline coronaviruses with small deletions in ORF7b associated with both enteric infection and feline infectious peritonitis. *Journal of Feline Medical Surgery*, *11*, 413–419.
- Locker, J. K., Griffiths, G., Horzinek, M. C., & Rottier, P. J. (1992). O-glycosylation of the coronavirus M protein. Differential localization of sialyltransferases in N- and O-linked glycosylation. *Journal of Biological Chemistry*, *267*, 14094–14101.
- Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W., & Ray, S. C. (1999). Full length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of inter-subtype recombination. *Journal of Virology*, *73*, 152–160.
- Longdon, B., Bockhurst, M. A., Russell, C. A., Welch, J. L., & Jiggins, F. M. (2014). The evolution and genetics of virus host shifts. *PLoS Path*, *10*, e1004395.
- Lu, G., Wang, Q., & Gao, G. F. (2015). Bat-to-human: Spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends in Microbiology*, *23*, 468–478. <https://doi.org/10.1016/j.tim.2015.06.003>
- Luan, Y., & Xu, W. (2007). The structure and main functions of aminopeptidase N. *Current Medical Chemistry*, *14*, 639–647.
- Ma, G., Wang, Y., & Lu, C. (2008). Molecular characterisation of the 9.36 kb C-terminal region of canine coronavirus 1-71 strain. *Virus Genes*, *36*, 491–497.
- Marescot, L., Benhaiem, S., Gimenez, O., Hofer, H., Lebreton, J.-D., Olarte-Castillo, X. A., Kramer-Schadt, S., & East, M. L. (2018). Social status mediates the fitness costs of infection with canine distemper virus in Serengeti spotted hyenas. *Functional Ecology*, *32*, 1237–1250. <https://doi.org/10.1111/1365-2435.13059>
- Marčić, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, *5*, e14004.
- Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*, vev003.
- Martin, D. P., Posada, D., Crandall, K. A., & Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombinant breakpoints. *Aids Research Human Retroviruses*, *12*, 98–102.
- Martin, D., & Rybicki, E. (2000). RDP: Detection of recombination among aligned sequences. *Bioinformatics*, *16*, 562–563.
- Maynard Smith, J. (1992). Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, *34*, 126–129.

- McCarthy, A. J., Shaw, M.-A., & Goodman, S. J. (2007). Pathogen evolution and disease emergence in carnivores. *Proceedings of the Royal Society of London*, *274*, 3165–3174.
- McGoldrick, A., Lowings, J. P., & Paton, D. J. (1999). Characterisation of a recent virulent transmissible gastroenteritis virus from Britain with a deleted ORF 3a. *Archives of Virology*, *144*, 763–770.
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., & Lopez, R. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic Acids Research*, *41*, W597–W600.
- Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5448.
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky-Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*, e1002764.
- Nikolin, V. N., Olarte-Castillo, X. A., Osterrieder, N., Hofer, H., Dubovi, E., Mazzoni, C. J., Brunner, E., Goller, K. V., Fyumagwa, R. D., Moehlan, P. D., Thierer, D., & East, M. L. (2017). Canine distemper virus in the Serengeti ecosystem: Molecular adaptation to different carnivore species. *Molecular Ecology*, *26*, 2111–2130.
- Olarte-Castillo, X. A., Heeger, F., Mazzoni, C. J., Greenwood, A. D., Fyumagwa, R., Moehlan, P. D., Hofer, H., & East, M. L. (2015). Molecular characterization of canine kobuvirus in wild carnivores and the domestic dog in Africa. *Virology*, *477*, 89–97. <https://doi.org/10.1016/j.virol.2015.01.010>
- Olarte-Castillo, X. A., Hofer, H., Goller, K. V., Martella, V., Moehlan, P. D., & East, M. L. (2016). Divergent sapovirus variants and infection prevalence in wild carnivores in the Serengeti ecosystem: A long-term study. *PLoS One*, *11*, e0163548.
- Packer, C., Altizer, S., Appel, M., Brown, E., Martenson, J., O'Brien, S. J., Roelke-Parker, M., Hofmann-Lehmann, R., & Lutz, H. (1999). Viruses of the Serengeti: Patterns of infection and mortality in African lions. *Journal of Animal Ecology*, *68*, 1161–1178.
- Padidam, M., Sawyer, S., & Fauquet, C. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology*, *265*, 218–225.
- Pearks Wilkerson, A. J., Teeling, E. C., Troyer, J. L., Bar-Gal, G. K., Roelke, M., Marker, L., Pecon-Slattey, J., & O'Brien, S. J. (2004). Coronavirus outbreak in cheetahs: Lessons for SARS. *Current Biology*, *14*, R227–R228.
- Pedersen, N. C., Liu, H., Dodd, K. A., & Pesavento, P. A. (2009). Significance of coronavirus mutants in faeces and diseased tissues of cats suffering from feline infectious peritonitis. *Viruses*, *1*, 166–184.
- Pedersen, N. C., Liu, H., Scarlett, J., Leutenegger, C. M., Golovko, L., Kennedy, H., & Kamal, F. M. (2012). Feline infectious peritonitis: Role of the feline coronavirus 3c gene in intestinal tropism and pathogenicity based upon isolates from resident and adopted shelter cats. *Virus Research*, *165*, 17–28.
- Posada, D., & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences USA*, *98*, 13757–13763.
- Pratelli, A., Decaro, N., Tinelli, A., Martella, V., Elia, G., Tempesta, M., & Buonavoglia, C. (2004). Two genotypes of canine coronavirus simultaneously detected in the faecal samples of dogs with diarrhea. *Journal of Clinical Microbiology*, *42*, 1797–1799.
- Pratelli, A., Elia, G., Martella, V., Tinelli, A., Decaro, N., Marsilio, F., & Buonavoglia, C. (2002). M gene evolution of canine coronavirus in naturally infected dogs. *Veterinary Research*, *151*, 758–761.
- Pratelli, A., Tempesta, M., Greco, G., Martella, V., & Buonavoglia, C. (1999). Development of a nested PCR for the detection of canine coronavirus. *Journal of Virology Methods*, *80*, 11–15.
- Reguera, J., Ordoño, D., Santiago, C., Enjuanes, L., & Casasnovas, J. M. (2011). Antigenic modules in the N-terminal S1 region of the transmissible gastroenteritis virus spike protein. *Journal of General Virology*, *92*, 1117–1126.
- Reguera, J., Santiago, C., Mudgal, G., Ordoño, D., Enjuanes, L., & Casasnovas, J. M. (2012). Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Path*, *8*, e1002859.
- Risco, C., Antón, I. M., Suñé, C., Pedregosa, A. M., Martín-Alonso, J. M., Parra, F., Carrascosa, J. L., & Enjuanes, L. (1995). Membrane protein molecules of transmissible gastroenteritis coronavirus also expose the carboxy-terminal region on the external surface of the virion. *Journal of Virology*, *69*, 5269–5277.
- Rottier, P. J. M. (1995). The coronavirus membrane protein. In S. G. Siddell (Ed.), *The coronaviridae* (pp. 115–139). Plenum Press.
- Sánchez-Morgado, J. M., Poynter, S., & Morris, T. H. (2004). Molecular characterisation of a virulent 580 canine coronavirus BGF strain. *Virus Research*, *104*, 27–31.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539.
- Sironi, M., Cagliani, R., Forni, D., & Clerici, M. (2015). Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nature Review Genetics*, *16*, 224–236.
- Sjöström, H., Norén, O., & Olsen, J. (2000). Structure and function of aminopeptidase N. *Advances in Experimental Medicine Biology*, *477*, 25–34.
- Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*, W609–W612.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729.
- Terada, Y., Matsui, N., Noguchi, K., Kuwata, R., Shimoda, H., Soma, T., Mochizuki, M., & Maeda, K. (2014). Emergence of pathogenic coronaviruses in cats by homologous recombination between feline and canine coronaviruses. *PLoS One*, *9*, e106534.
- Thalwitzer, S., Wachter, B., Robert, N., Wibbelt, G., Müller, T., Lonzer, J., Meli, M. L., Bay, G., Hofer, H., & Lutz, H. (2010). Seroprevalences to viral pathogens in free-ranging and captive cheetahs (*Acinonyx jubatus*) on Namibian farmland. *Clinical Vaccine Immunology*, *17*, 232–238.
- Tresnan, D. B., Levis, R., & Holmes, K. V. (1996). Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I. *Journal of Virology*, *12*, 8669–8674.
- Tsangaras, K., Siracusa, M. C., Nikolaidis, N., Ishida, Y., Cui, P., Vielgrader, H., Helgen, K. M., Roca, A. L., & Greenwood, A. D. (2014). Hybridization capture reveals evolution and conservation across the entire Koala retrovirus genome. *PLoS One*, *4*, e95633.
- Tusell, S. M., Schittone, S. A., & Holmes, K. V. (2007). Mutational analysis of aminopeptidase N, a receptor for several group 1 coronaviruses, identifies key determinants of viral host range. *Journal of Virology*, *81*, 1261–1273.
- Vijaykrishna, D., Smith, G. J., Zhang, J. X., Peiris, J. S., Chen, H., & Guan, Y. (2007). Evolutionary insights into the ecology of coronaviruses. *Journal of Virology*, *81*, 4012–4020.
- Wentworth, D. E., & Holmes, K. V. (2001). Molecular Determinants of Species Specificity in the Coronavirus Receptor Aminopeptidase N (CD13): Influence of N-Linked Glycosylation. *Journal of Virology*, *75*(20), 9741–9752. <https://doi.org/10.1128/JVI.75.20.9741-9752.2001>
- Whittaker, G. A., André, N., & Millet, J. K. (2018). Improving virus taxonomy by recontextualizing sequence-based classification with biologically relevant data: The case of the alphacoronavirus-1 species. *mSphere*, *3*, e00463–e517.
- Woodroffe, R., Prager, K. C., Munson, L., Conrad, P. A., Dubovi, E. J., & Mazet, J. A. K. (2012). Contact with domestic dogs increases pathogen exposure in endangered African wild dogs (*Lycan pictus*). *PLoS One*, *7*, e30099.

- Woolhouse, M. E. J., Haydon, D. T., & Antia, R. (2005). Emerging pathogens: The epidemiology and evolution of species jumps. *Trends in Ecology and Evolution*, 20, 238–244.
- Wozencraft, W. C. (2005). Order Carnivora. In D. E. Wilson, & D. M. Reeder (Eds.), *Mammal species of the world: A taxonomic and geographic reference*, 3rd ed. (pp. 532–628). Johns Hopkins University Press.
- Wu, K. L., Peng, G. Q., Wilken, M., Geraghty, R. J., & Li, F. (2012). Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *Journal of Biological Chemistry*, 287, 8904–8911.
- Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, 13, 555–556.
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24, 1586–1591.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Olarte-Castillo XA, dos Remédios JF, Heeger F, et al. The virus–host interface: Molecular interactions of *Alphacoronavirus-1* variants from wild and domestic hosts with mammalian aminopeptidase N. *Mol Ecol*. 2021;30:2607–2625. <https://doi.org/10.1111/mec.15910>