MOLECULAR ECOLOGY

Molecular Ecology (2017) **26**, 2111–2130 doi: 10.1111/mec.13902

MICROBIAL LOCAL ADAPTATION

Canine distemper virus in the Serengeti ecosystem: molecular adaptation to different carnivore species

VELJKO M. NIKOLIN,*†[1,3] XIMENA A. OLARTE-CASTILLO,*[3] NIKOLAUS OSTERRIEDER,† HERIBERT HOFER,* EDWARD DUBOVI,‡ CAMILA J. MAZZONI,*§ EDGAR BRUNNER,¶ KATJA V. GOLLER,*[2] ROBERT D. FYUMAGWA,** PATRICIA D. MOEHLMAN,†† DAGMAR THIERER* and MARION L. EAST*

*Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany, †Institut für Virologie, Freie Universität Berlin, Robert-von-Ostertag-Str. 7-13, 14163 Berlin, Germany, ‡Animal Health Diagnostic Centre, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA, §Berlin Center for Genomics in Biodiversity Research, Königin-Luise-Str. 6-8, 14195 Berlin, Germany, ¶Department of Medical Statistics, Faculty of Medicine, University of Göttingen, Humboldtallee 32, 37073 Göttingen, Germany, **Tanzania Wildlife Research Institute, P.O. Box 661, Arusha, Tanzania, ††EcoHealth Alliance, 460 West 34th St, New York, NY 10001-2320, USA*

Abstract

Was the 1993/1994 fatal canine distemper virus (CDV) epidemic in lions and spotted hyaenas in the Serengeti ecosystem caused by the recent spillover of a virulent domestic dog strain or one well adapted to these noncanids? We examine this question using sequence data from 13 'Serengeti' strains including five complete genomes obtained between 1993 and 2011. Phylogenetic and haplotype network analyses reveal that strains from noncanids during the epidemic were more closely related to each other than to those from domestic or wild canids. All noncanid 'Serengeti' strains during the epidemic encoded: (1) one novel substitution G134S in the CDV-V protein; and (2) the rare amino acid combination 519I/549H at two sites under positive selection in the region of the CDV-H protein that binds to SLAM (CD 150) host cell receptors. Worldwide, only a few noncanid strains in the America II lineage encode CDV-H 519I/549H. All canid 'Serengeti' strains during the epidemic coded CDV-V 134G, and CDV-H 519R/549Y, or 519R/549H. A functional assay of cell entry revealed the highest performance by CDV-H proteins encoding 519I/549H in cells expressing lion SLAM receptors, and the highest performance by proteins encoding 519R/549Y, typical of dog strains worldwide, in cells expressing dog SLAM receptors. Our findings are consistent with an epidemic in lions and hyaenas caused by CDV variants better adapted to noncanids than canids, but not with the recent spillover of a dog strain. Our study reveals a greater complexity of CDV molecular epidemiology in multihost environments than previously thought.

Keywords: canine distemper virus, lion, Serengeti, SLAM (CD150), spotted hyaena, virus–host adaptation

Received 24 June 2015; revision received 25 August 2016; accepted 2 September 2016

Correspondence: Marion L. East, Fax: +49 30 5126 104; E-mail: east@izw-berlin.de

[1]Present address: Boehringer Ingelheim Veterinary Research Center, Bemeroder Str. 31, 30559 Hannover, Germany

[2]Present address: Friedrich-Loeffler-Insitut, Bundesforschungsinstitut für Tiergesundheit, Südufer 10, 17493 Greifswald-Insel Riems, Germany

[3]Contributed equally to this work.

Introduction

Molecular adaptations of viruses to host species can cause host-range shifts and explain the distribution of genetically distinct strains across species (Crill *et al.* 2000; Moya *et al.* 2000; Turner & Elena 2000; Antia *et al.* 2003). Current evidence suggests that closely related species are more susceptible to infection by a given

virus or strain than more distantly related hosts (Read & Taylor 2001; Faria *et al.* 2013; Longdon *et al.* 2014), because they provide a similar cellular environment to which the pathogen is adapted. However, there are viruses where this does not apply (e.g. Greenwood *et al.* 2012; Shi *et al.* 2014). Adaptation of virus attachment proteins to the cell receptor used by viruses to gain entry to host cells can alter virus spread across host species and has resulted in epidemics in novel hosts (Moya *et al.* 2004; Allison *et al.* 2014; Shi *et al.* 2014).

Canine distemper virus (CDV) belongs to the genus *Morbillivirus* (family *Paramyxoviridae*). CDV has a nonsegmented, single-stranded, negative-sense RNA genome with six genes that produce six structural proteins. The nucleocapsid protein (CDV-N) encapsidates the viral RNA, and the phosphoprotein (CDV-P) together with the large protein (CDV-L) form the transcriptase/replicase complex. The CDV-P gene also expresses two nonstructural proteins, the CDV-C and CDV-V proteins. One function of the CDV-V protein is to suppress the innate immune response of hosts (von Messling *et al.* 2006; Röthlisberger *et al.* 2010). The haemagglutinin protein (CDV-H) and the fusion protein (CDV-F) facilitate virus entry and the matrix protein (CDV-M) is important for the budding of virus particles (Lamb & Kolakofsky 2001). As the CDV-H gene is the most diverse gene in the *Morbillivirus* genus (Pomeroy *et al.* 2008), it is used to segregate strains into lineages (Bolt *et al.*1997; Martella *et al.* 2006), whereas the CDV-L and CDV-M genes are more conserved than other genes in the *Paramyxoviridae* (McCarthy & Goodman 2010). The long untranslated region (UTR) between the CDV-M and F genes (M-F UTR) modulates virulence through transcriptional control of CDV-F gene expression (Anderson & von Messling 2008).

CDV causes a highly contagious disease in domestic dogs (*Canis lupus familiaris*, family Canidae) worldwide and infects a taxonomically broad range of carnivores (Appel & Summers 1995; Pomeroy *et al.* 2008), including endangered species (e.g. Williams *et al.* 1988; Goller *et al.* 2010; Meli *et al.* 2010; Seimon *et al.* 2014). The outcome of CDV infection ranges from subclinical to lethal; survival of infection provides lifelong immunity (Appel & Summers 1995). Clinical disease and virus shedding occurs once CDV has completed two stages of infection (Sawatsky *et al.* 2012). During the first stage, CDV targets the signalling lymphocytic activation molecule (SLAM, CD150) receptors expressed on the surface of T and B lymphocytes, macrophages and dendritic cells (Tatsuo *et al.* 2001; von Messling *et al.* 2006). Virus entry occurs when the CDV-H protein attaches to SLAM receptors and together with the CDV-F protein accomplishes fusion with host cell membranes (von Messling *et al.* 2001; Tatsuo *et al.* 2001). Virus amplification in lymphatic tissues and circulating lymphocytes induces immunosuppression and the spread of CDV through the bloodstream (Lamb & Kolakofsky 2001). In the second stage of infection, CDV enters epithelial cells using the nectin-4 receptor (Noyce *et al.* 2013), leading to the manifestation of disease and virus shedding (Sawatsky *et al.* 2012).

The mammalian order Carnivora contains two suborders, the Caniformia and Feliformia (Flynn *et al.* 2005). Carnivore SLAM proteins reflect this ancient bifurcation (Nikolin *et al.* 2012b), as SLAM proteins from caniform species within family Canidae have high amino acid sequence identity and a far lower identity to those in feliform species in families Felidae and Hyaenidae. McCarthy *et al.* (2007) proposed that efficient virus binding to SLAM receptors in carnivores other than the domestic dog (hereafter termed dog) requires adaptations in the SLAM-binding region of the CDV-H protein. They identified that sites 530 and 549 in this region were under positive selection and that most CDV-H proteins from dog strains encode tyrosine (Y) at site 549 whereas those in strains from other carnivores mostly specified histidine (H). This finding is supported by several empirical studies (Monne *et al.* 2011; Kameo *et al.* 2012; Nikolin *et al.* 2012a; Origgi *et al.* 2012) and the occurrence of the Y549H substitution during experimental adaptation of a dog CDV strain to ferrets (*Mustela putorius furo*, family Mustelidae) (von Messling *et al.* 2003). A quantitative assessment of the functional importance of the residue at site 549 revealed that CDV-H proteins encoding 549Y performed significantly better in cells expressing dog SLAM receptors than those encoding 549H, whereas proteins encoding 549H performed better in cells expressing lion (*Panthera leo*, family Felidae) SLAM than those specifying 549Y, thereby indicating antagonistic pleiotropic effects of the residue at site 549 on CDV-H-mediated attachment to SLAM receptors in distantly related carnivore species (Nikolin *et al.* 2012b). Even so, variants specifying 549Y can cause clinical disease in noncanids (e.g. Meli *et al.* 2010; Keller *et al.* 2012; Nagao *et al.* 2012; Seimon *et al.* 2014) and those encoding 549H can cause disease in dogs (e.g. Fischer *et al.* 2016).

The worldwide spread of CDV from the USA reached South Africa in the 1920s (Panzera *et al.* 2015), but it is unclear when it arrived in East Africa. A CDV epidemic emerged in the Serengeti National Park (NP) in northeast Tanzania in November 1993 (Haas *et al.* 1996) and then spread through the Serengeti ecosystem (Roelke-Parker *et al.* 1996), including the Maasai Mara National Reserve (NR) in Kenya (Harrison *et al.* 2004). Unexpectedly, it caused high morbidity and mortality in juvenile and adult lions (Roelke-Parker *et al.* 1996; Munson *et al.* 2008) and predominantly juvenile (<12 months of age)

spotted hyaenas (*Crocuta crocuta*, family Hyaenidae, hereafter termed hyaena) (Haas *et al.* 1996). In canids, infection was confirmed in two bat-eared foxes (*Otocyon megalotis*), and neurological signs were observed in two silver-backed jackals (*Canis mesomelas*) and one golden jackal (*Canis aureus*) (Roelke-Parker *et al.* 1996). Since this epidemic, CDV has not caused clinical disease in either the lion (Munson *et al.* 2008) or hyaena populations (M. L. East and H. Hofer unpublished data) in the Serengeti NP, despite serological data revealing three periods of increased CDV seroprevalence (1999–2001, 2006/2007 and 2010/2011) in lions in the Serengeti NP (Munson *et al.* 2008; Viana *et al.* 2015), and hyaenas in the Maasai Mara NR in 2000 (Harrison *et al.* 2004).

Although CDV spillover from dogs can cause epidemics in wildlife, molecular and ecological studies indicate that this virus can be maintained in wildlife populations (Almberg *et al.* 2010; Riley & Wilkes 2015). To explain high CDV-mediated mortality in lions during the 1993/1994 epidemic, Roelke-Parker *et al.* (1996) proposed that a virulent CDV 'biotype' spread from dogs outside the Serengeti NP to wild carnivores inside the park. More recently, a stochastic model (Guiserix *et al.* 2007) suggested that low herd immunity among lions, rather than CDV virulence, better explained the epidemic. Neither a stochastic model (Craft *et al.* 2008) or network model (Craft *et al.* 2009) could explain the observed spread of CDV in prides during the epidemic purely by lion-to-lion transmission, and hence, both posited that a single variant circulating in dogs and wild carnivores was repeatedly introduced by jackals and hyaenas to prides.

Our study examines the distribution and outcome of CDV infection in carnivores in the Serengeti ecosystem to identify the strain(s) associated with: (i) mortality among lions and hyaenas in 1993/1994, (ii) high levels of benign exposure in these noncanids since the epidemic, despite (iii) ongoing lethal disease in canids. If CDV in 1993/1994 spread from dogs to lions and hyaenas, as proposed by Roelke-Parker *et al.* (1996), we expect genetic sequence data to reveal a close relationship between strains from dog and noncanid hosts during this epidemic. The theory of virus–host adaptation suggests that virus transmission chains within one host species could lead to beneficial adaptations to this host which might be detrimental to the virus in distantly related hosts. We therefore predicted that (1) CDV strains adapted to dogs should outcompete non-canid-adapted strains in dogs and cause clinical disease in closely related wild canids but not in noncanids; (2) if several strains circulate, then those adapted to noncanids should outcompete canid-adapted strains in noncanids, but not in canids; and (3) strains causing clinical disease in wild canids are less likely to do so in noncanids. We tested these ideas by examining the relationship between strains from different carnivores and the infection outcome from 1993 to 2011. We also assess the functional importance of the amino acid combination at two sites under positive selection in the CDV-H SLAM-binding region in cellular environments expressing SLAM receptors from canids or noncanids.

Materials and methods

Study area and sampling

We investigated 13 CDV variants (Table S2, Supporting information) obtained between 1993 and 2011 from five wild carnivore species and the dog. Ten variants came from the Serengeti NP, a protected area of 14 763 km^2 in northern Tanzania, where dogs are prohibited (East *et al.* 2012), and three from the Tanzanian sector of the Serengeti ecosystem that covers approximately 25 000 km^2 and straddles the Tanzania–Kenya border in East Africa.

CDV variants were found in tissue samples from three dead hyaenas killed by lions, or by vehicles on roads in the Serengeti NP, in blood samples from three hyaenas anaesthetized for the removal of wire snares set by illegal bushmeat hunters (Hofer *et al.* 1993) and tissues from two cubs that succumbed to canine distemper during the epidemic in 1993 and 1994. One variant from an African wild dog (*Lycaon pictus*, family Canidae) came from a pack that died of CDV in 2007 (Goller *et al.* 2010) and another from a golden jackal was obtained from bloody faeces from a pup that succumbed to canine distemper in 2011. Samples were stored either at −80 °C until processed or placed in RNAlater and then stored frozen at −10 °C in Tanzania and then at −80 °C at the Leibniz Institute for Zoo and Wildlife Research until processed. Sequences were also obtained from four CDV isolates collected in the Serengeti NP in 1994 and stored at the Virology Section Diagnostic Center, Cornell University (i.e. *SNP/1994/African lion, SNP/1994/spotted hyaena, SNP/1994/bat-eared fox, SE/1994/domestic dog*). RNA isolation from samples was performed using RNeasy mini kit (Qiagen, Germany) according to the manufacturer's instructions. In total, we obtained five complete CDV genome sequences, six complete CDV-H gene sequences plus an additional seven partial CDV-H gene sequences, all of which spanned the SLAM-binding region of the CDV-H gene, and eight complete CDV-P gene sequences (Table S2, Supporting information).

Genetic sequencing

We obtained complete CDV genome sequences from five variants (*SNP/1994/African lion, SNP/1994/spotted*

hyaena, *SE/1994/domestic dog*, *SE/2007/African wild dog*, *SE/2011/golden jackal*) using a hybrid-capture enrichment protocol (Maricic *et al.* 2010; Tsangaras *et al.* 2014) in which PCR-generated 'baits' are used to capture target sequences from pooled cDNA libraries for sequencing using next-generation technologies. Baits were generated from PCR products covering the complete genome of variant *SE/1994/domestic dog*. Libraries were indexed by sample and were built as previously described (Meyer & Kircher 2010). Baits and cDNA libraries were constructed in separate laboratories to avoid contamination. Enriched libraries from the hybrid capture were deep-sequenced on the Illumina MiSeq™ platform (Illumina Inc., San Diego, CA, USA). Sequence quality was assessed as previously described (Olarte-Castillo *et al.* 2015). Trimmed reads from each sample were assembled against reference strain 171391–513 (Accession no. KJ123771), using the medium sensitivity method with up to 10 iterations in GENEIOUS 9.0.2 (Biomatters Ltd, Auckland, New Zealand). Genome organization and predicted gene products of the sequenced 'Serengeti' variants (Table S2, Supporting information) were identical to those of other canine distemper viruses. The genomes were 15 690 nucleotides long and contained six open reading frames (ORF) coding for six structural proteins (N, P, M, F, H and L). Within the CDV-P gene sequence was an additional ORF coding for the nonstructural C protein. Reverse transcription–PCR (RT–PCR) was performed to amplify the complete CDV-H gene (1824 nucleotides, position within viral genome: 7079–8902) and the complete P gene (1524 nucleotides, position within the viral genome: 1801–3324). When we could not obtain the entire CDV-H gene, we amplified a 514-nucleotide segment of the 3′-end of this gene (position within viral genome: 8479–8993) which included the SLAM-binding region. PCR products were cloned into the PCR 2.1 TOPO vector and sequenced. Nucleic acid sequences were edited and aligned using the multiple alignment method (ClustalW) incorporated within the computer program MEGA 6.0 (www.megasoftware.net) (Tamura *et al.* 2013). Accession numbers and details for CDV-P and CDV-H gene sequences generated by this study are listed in Table S2 (Supporting information).

Genetic analyses

Sequences obtained in this study, together with those retrieved from GenBank, were aligned using the MUSCLE algorithm (Edgar 2004) in GENEIOUS 7.1.2 (Biomatters Ltd). We determined the phylogenetic relationship between strains based on complete genome sequences, and the complete sequences of the H and P gene using the maximum-likelihood method as implemented in MEGA 6.0 (Tamura *et al.* 2013). We determined the best fitted model using the Bayesian information criterion (BIC). The best substitution models identified were the general time-reversible plus invariant sites plus gamma (γ) distribution (GTR+I+G) model (e.g. Strimmer & von Haeseler 2009), the Tamura 3-parameter model (Tamura 1992) with a γ distribution (Wiens & Servedio 1998; Gadagkar & Kumar 2005) and the Hasegawa–Kishino–Yano model (Hasegawa *et al.* 1985) for the complete genome, the CDV-H and the CDV-P genes, respectively. For the complete M-F UTR and the CDV-N, CDV-M, CDV-F and CDV-L genes, the Tamura 3-parameter model was used. For amino acid sequences of the four proteins, the Jones–Taylor–Thornton model (Jones *et al.* 1992) was used. Statistical support was provided by 1000 bootstrap replicates, and the respective percentages are quoted on branches. The haplotype networks for 'Serengeti' variants were built using the median-joining algorithm (Bandelt *et al.* 1999) by the computer program NETWORK 4.6.1.2. (fluxus-engineering.com). Similarities along the genome (Fig. 1a) were visualized with SIMPLOT 5.1 (Lole *et al.* 1999) using the Kimura 2-parameter distance model with a sliding frame of 400 nucleotides and a step distance of 20 bases.

Sites under positive selection

To identify amino acid sites in the CDV-H SLAM-binding region under positive selection, we first used nested maximum-likelihood site models in CODEML as implemented in PAML version 4 (Yang 1997, 2007). These models allow the ratio (ω) between nonsynonymous substitutions (dN) and synonymous substitutions (dS) to vary from site to site. Three pairs of nested models compared the results of null models (M0, M1a and M7, respectively) in which no sites were permitted to evolve ($\omega \leq 1$), with those from models (M3, M2a and M8, respectively) that allowed positive selection of some site ($\omega > 1$, see Table S3, Supporting information for details). Likelihood ratio tests (LRTs) were used to assess which of the two nested models best fitted the data set. When results of the LRTs significantly supported the existence of positive selection pressure, the naïve empirical Bayes (NEB) (Nielsen & Yang 1998; Yang *et al.* 2000) and the Bayes empirical Bayes (BEB) (Yang *et al.* 2005) approaches were used to identify the sites evolving under positive selection.

Nested maximum-likelihood site models assume that diversifying selection pressure at each site remains constant over time and affects most lineages in a phylogenetic tree (Murrell *et al.* 2012). This assumption does not fit CDV-H gene sequences from 'Serengeti' strains and noncanid strains in the America II lineage, which encode a rare amino acid at site 519 absent from other

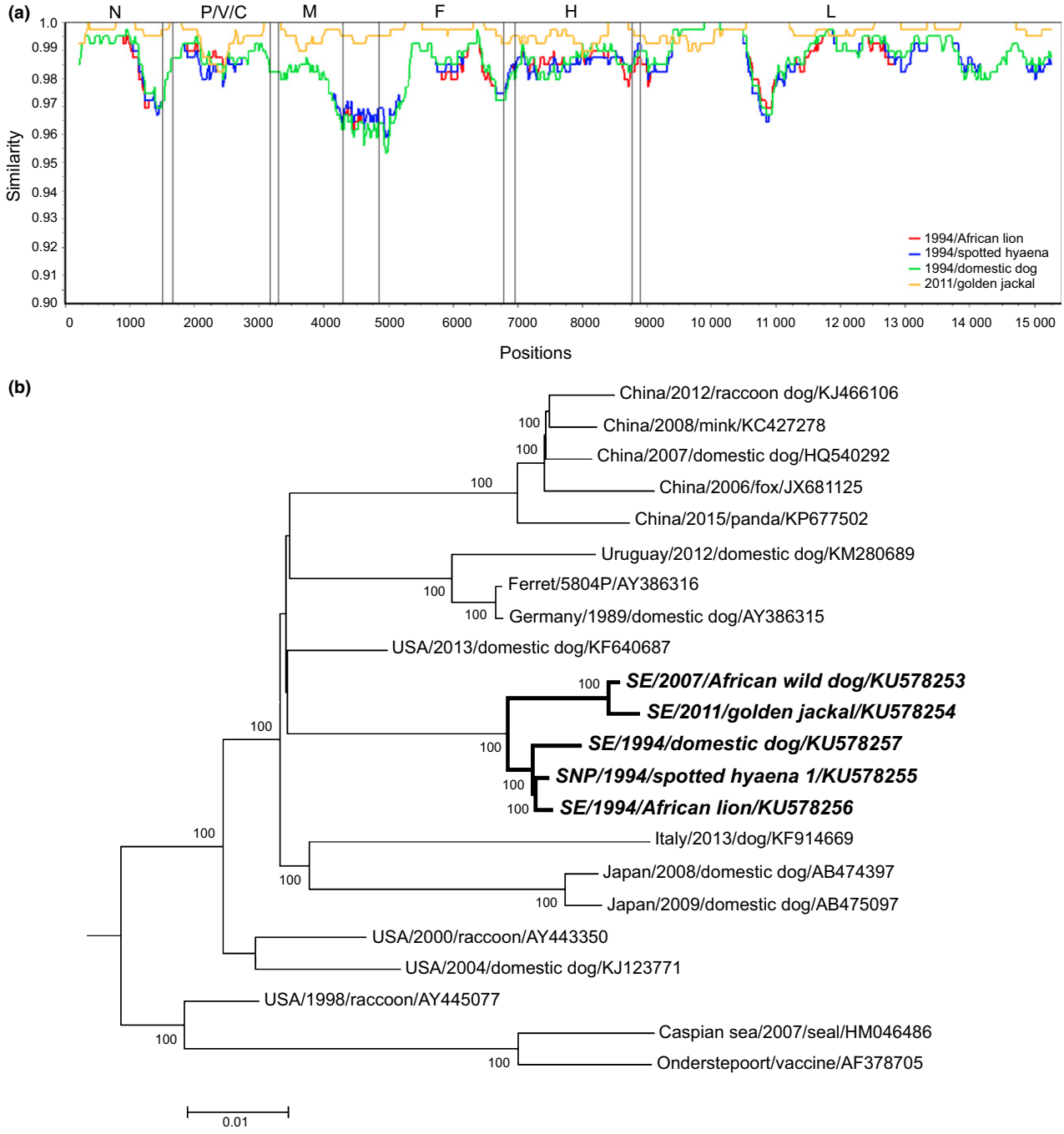


Fig. 1 Comparison of the complete genome of CDV strains from the Serengeti ecosystem. (a) Similarity plot analysis: The percentage similarity of complete genome sequences of four 'Serengeti' variants from different hosts (plotted in different colours: lion – red; spotted hyaena – blue; domestic dog – green; golden jackal – yellow) to one 'Serengeti' strain from an African wild dog (Table S2, Supporting information). The positions of gene regions (N, P, M, F, H, L) are indicated above the plot. (b) The phylogenetic relationship of 'Serengeti' strains to those from other areas, based on the maximum-likelihood (ML) method and nucleotide sequences of complete genomes (15 476 nucleotides long). Branch lengths are scaled to nucleotide substitutions per site. Statistical support for nodes was provided by 1000 replicates. Numbers at nodes indicate bootstrap values >70%. 'Serengeti' variants are presented on bold branches and in bold italics. For each strain, the origin, year and host species are quoted. Strains from the Serengeti National Park (SNP) and wider Serengeti ecosystem (SE) are indicated.

lineages (Table S1, Supporting information). Mixed-effects models of evolution (MEME) permit the identification of pervasive positive selection and episodic positive selection that affects specific sites along particular lineages by allowing the distribution of ω to vary between sites and between branches at a site (Murrell

et al. 2012). We applied MEME as implemented in the Datamonkey server (http://www.datamonkey.org) using model HKY85 and the default cut-off of 0.1. The empirical Bayes procedure was used to identify branches subject to positive selection at identified sites. Events were considered to be under positive selection if the empirical Bayes factor (EBF) exceeded 50.

We ran all models using the CDV-H SLAM-binding region sequences from 13 'Serengeti' strains (Table S2, Supporting information) and 33 strains worldwide (all GenBank accession numbers are provided in Fig. S12, Supporting information). Three 'Serengeti' strains had identical sequences in this region (strains *SE/2007/ African wild dog*, *SNP/1996/spotted hyaena*, *SNP/1997/spotted hyaena*); hence, 11 unique 'Serengeti' sequences were submitted to the MEME model.

Whole 'Serengeti' CDV genomes were checked for potential recombinant sequences and recombination break points using the Recombination Detection Program (RDP) (Martin & Rybicki 2000), GENECONV (Sawyer 1989), BOOTSCAN (Martin *et al.* 2005b), MaxChi (Maynard Smith 1992), CHIMAERA (Posada & Crandall 2001) and SISCAN (Gibbs *et al.* 2000) methods implanted in RDP3 (Martin *et al.* 2005a). No recombination events were found.

Construction of expression plasmids and assay

We constructed expression vectors (pCG) containing the CDV-H gene (that encodes the protein attaching to the SLAM receptor) and CDV-F genes (that encodes the fusion protein which together with the SLAM receptor facilitates virus fusion with the host cell membrane) from strains *A75/17*, *Lion94SNP* and *Dog94SE* and the mutant *A75/17-549H* as previously described (Nikolin *et al.* 2012b). Site-directed mutagenesis of CDV-H genes at amino acid site 519 was performed on pCG plasmids coding a wild-type CDV-H gene from strain *A75/17* and its mutant CDV-H *A75/17-549H* (encoding H at position 549), yielding expression plasmids encoding additional two recombinant haemagglutinin genes: *A75/17 –519I* and *A75/17–519I/549H*. The QuikChange Site-Directed Mutagenesis kit (Stratagene, USA) was used to introduce specific point mutations according to the manufacturers' instructions, using primers *A75/17-R519I*-S (5′-GTTGCC TACACA GAATTT TATATA TGTCAT AGCAAC ATATG -3′) and *A75/17-R519I*-AS (5′- CATATG TTGCTA TGACAT ATATAA AATTCT GTGTAG GCAAC -3′). Vero cell lines (ATCC CCL-81) stably expressing lion, domestic cat and dog SLAM were obtained as previously described (Nikolin *et al.* 2012b).

CDV infection in cultured cells leads to the formation of multinucleated cells termed syncytia. The proficiency of syncytium formation is chiefly determined by the CDV-H protein (von Messling *et al.* 2001). The assay quantified the mean number of nuclei (MNN) per syncytium as a measure of the performance of CDV-H proteins at cell entry in Vero cell lines that stably expressed SLAM from two known CDV hosts (dog and lion) and one species, the domestic cat (*Felis catus*, family Felidae), in which CDV replicates in lymphatic tissues and induces seroconversion but does not induce clinical disease (Appel *et al.* 1974; Ikeda *et al.* 2001). The assay methods are detailed in full by Nikolin *et al.* (2012b). Briefly, Vero cell lines that stably expressed dog SLAM, lion SLAM or domestic cat SLAM were seeded in 12-well plates and cotransfected using 1 μL of Fugene transfection reagent (Roche, Germany) with 0.10 μg of each CDV-H and CDV-F expression plasmids in six different combinations (1) pCG-*domestic dog94SE*-CDV-H + pCG-*African lion94SNP*-CDV-F (*domestic dog94SE*), (2) pCG-*Lion94SNP*-CDV-H + pCG-*African lion94SNP*-CDV-F (*African lion94SNP*), (3) pCG-*A75/17*-CDV-H + pCG-A75/17-CDV-F (*A75/17*), (4) pCG-*A75/17*-519I-CDV-H + pCG-*A75/17*-CDV-F (the experimentally mutated *A75/17-519I*), (5) pCG-*A75/17*-549H-CDV-H + pCG-*A75/17*-CDV-F (the experimentally mutated *A75/17-549H*) and (6) pCG-*A75/17*-519I/549H-CDV-H + pCG-*A75/17*-CDV-F (the experimentally double mutated *A75/17-519I/549H*). Twelve hours after transfection, cells were fixed with 2% paraformaldehyde in PBS and permeabilized with 0.2% Tween 20 in PBS, and nuclei were visualized with Hoechst 33342 trihydrochloride, trihydrate (Invitrogen, USA). Pictures were taken using the Zeiss AxioVert S100 microscope (Carl-Zeiss, Germany) and recorded with a CCD camera (Axiocam, Germany). For each cell line, six independent transfection repeats were performed. For each experiment, five random images were taken of each well, syncytia were marked and nuclei were counted in a blinded fashion. A total of 60 syncytia for each cell type and plasmid combination were counted from which we calculated an MNN per syncytium per experiment.

Statistical analysis

The experiments performed for the *in vitro* assay corresponded to a split-plot design of independent experiments. We used the 'nonparametric marginal model' (NMM) (Brunner *et al.* 1999; Brunner & Puri 2001) to assess the statistical significance of the assay results (six replicates of the NMM per experiment, Table S4, Supporting information) produced by Vero cell lines with different SLAM receptors and CDV-H proteins from different strains. The NMM has the advantage over general linear model of not making any assumptions about data distribution, only uses the empirical distribution of the data and can be used to analyse factorial

experiments of arbitrary complexity. Testing for significant differences between the effects of different factors such as type of receptor or type of strain was performed by employing an ANOVA-type statistic (ATS) which approximately follows an F-distribution (Brunner *et al.* 1999; Brunner & Puri 2001).

Results

The similarity and genetic relationship of 'Serengeti' strains

We present the first complete CDV genomes from Africa (five strains from five host species; for accession numbers, see Table S2, Supporting information). A similarity plot (Fig. 1a) based on nucleotide sequences of these genomes, in which *SE/2007/African wild dog/KU578253* was set as the reference, revealed that one strain from a golden jackal in 2011 (plotted in yellow) was more similar to the reference sequence than strains from 1994 in lion, hyaena and dog (plotted in red, blue and green, respectively, Fig. 1a). Low similarity, in the C-terminus of CDV-N and CDV-M proteins, M-F UTR and N-terminus of CDV-F protein, is apparent between strains from 1994 and those from 2007 and 2011.

A phylogenetic analysis using complete CDV genome nucleotide sequences (for accession numbers, see Fig. 1b) placed all 'Serengeti' strains together in one distinct, well-supported clade. Within this 'Serengeti' clade, strains from 1994 were grouped together, as were those from 2007 and 2011. The strains from lion and hyaena in 1994 were more closely related to each other than to the strain from a dog in that year. The same analysis using deduced amino acid sequences for the complete genome yielded similar results for 'Serengeti' strains (Fig. S1a, Supporting information). 'Serengeti' strains were most closely related to one domestic dog strain from the USA in both analyses (Figs 1b and S1a, Supporting information).

Analysis of complete CDV-H gene nucleotide sequences from six 'Serengeti' strains and 57 strains from CDV lineages worldwide (Table S1, Supporting information, strains marked in blue) placed all 'Serengeti' sequences in one statistically well-supported clade, indicating a lineage evolving independently from previously described geographical lineages, including the South African lineage (Fig. 2). We term this new lineage the East African lineage. Sequence identity between complete CDV-H gene sequences from the East African lineage and the South African lineage was relatively low (92.7–93.9%) as was the identity between the East African lineage and the European wildlife lineage (93.7–95.6%). 'Serengeti' strains from 1994 formed a distinct cluster in which strains from lion and hyaena were more closely related to each other than to strains from a dog and a bat-eared fox (Fig. 2). Strains from 2007 and 2011 were placed separately from those from 1994.

Analysis of 13 partial 'Serengeti' CDV-H gene nucleotide sequences (Table S2, Supporting information) placed strains from 1993, 1994 and 1997 together (Fig. S1, Supporting information) and those collected between 1996 and 2011 in a separate cluster. The identity between sequences in these two clusters varied from 98.2 to 98.4%. In the 1993 to 1997 cluster (but not in the 1996–2011 cluster), strains from canids were placed separately from noncanid strains (Fig. S1b, Supporting information).

Analysis of complete CDV-P gene nucleotide sequences from eight 'Serengeti' strains (Table S2, Supporting information) and 14 strains from elsewhere (for GenBank accession numbers, see Fig. 3) placed 'Serengeti' strains together in one distinct clade. Within this clade, strains from hyaena and lion in 1993 and 1994 were more closely related to each other than to strains from a dog and bat-eared fox in 1994. Strains from wild canids in 2007 and 2011 were placed separately from those from 1994.

Analyses of nucleotide sequences from the CDV-N, CDV-M, CDV-F and CDV-L genes (Figs S2a–S5a, Supporting information, respectively) and the M-F UTR (Fig. S6a, Supporting information) all revealed that strains from 1994 were more closely related to each other than to those from 2007 and 2011. Results for the CDV-N, CDV-F and CDV-L genes revealed that strains from hyaena and lion in 1994 were more closely related to each other than to the 1994 dog strain (Figs S2a, S4a and S5a, Supporting information, respectively), but the bootstrap value for the CDV-F gene was 69% (Fig. S4a, Supporting information). Results from the CDV-M gene (Fig. S3, Supporting information) and the M-F UTR (Fig. S6a, Supporting information) placed canid and noncanid strains from 1994 together. Analyses using amino acid sequences for CDV-N, CDV-M, CDV-F and CDV-L proteins (Figs S2b–S5b, Supporting information, respectively) revealed similar results to those obtained from nucleotide sequences, except that CDV-N amino acid sequences from canid and noncanid strains in 1994 clustered together (Fig. S2b, Supporting information).

Haplotype networks

A haplotype network analysis using nucleotide sequences of the SLAM-binding region (Fig. 4a) in 13 'Serengeti' strains (Table S2, Supporting information) revealed (Fig. 4b) that the 1994 dog strain was separated from the 1993 hyaena strain by two nodes, both involving strains from wild canids plus two

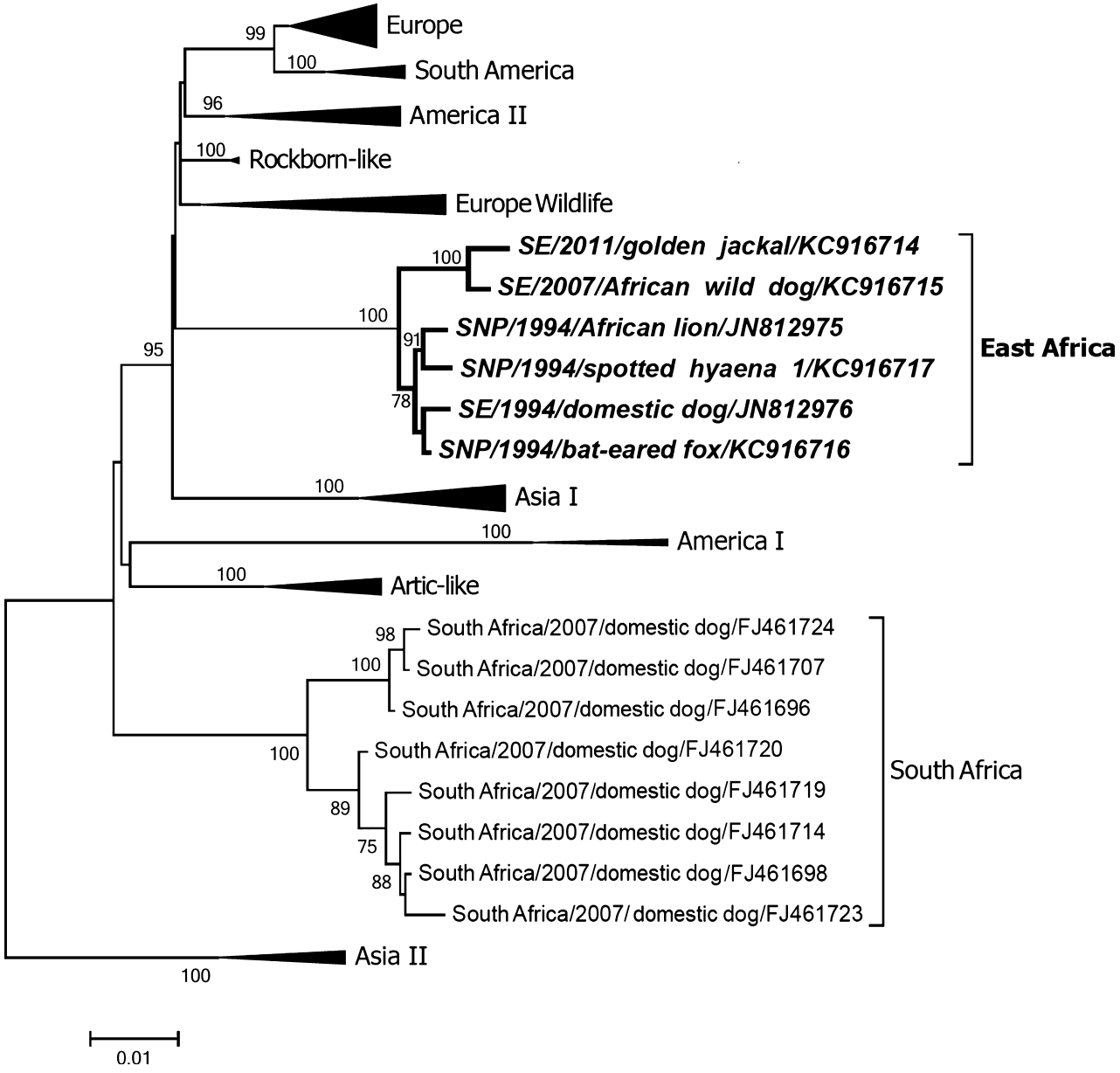


Fig. 2 The phylogenetic relationship of CDV strains from the Serengeti ecosystem to known CDV geographical lineages worldwide based on the maximum-likelihood (ML) method and nucleotide sequences of the complete CDV-H gene (1824 nt). Branch lengths are scaled to nucleotide substitutions per site and 'Serengeti' variants are presented on bold branches and in bold italics. Statistical support for nodes was provided by 1000 replicates. Numbers at nodes indicate bootstrap values >70%. Lineages from outside Africa are collapsed. Brackets encompass strains within lineages. For each strain, the origin, year and host species are quoted. Strains from the Serengeti National Park (SNP) and wider Serengeti ecosystem (SE) are indicated.

nonsynonymous amino acid substitutions: (1) Y549H (Fig. 4b) present in a bat-eared fox strain; and (2) R519I, arginine (R) to isoleucine (I) at site 519 present in the 1993 hyaena strain (Fig. 4a,b), which encoded 519I/549H. Only noncanid strains radiate from this 1993 hyaena haplotype (Fig. 4b), including all those from lion and hyaenas during the 1993/1994 epidemic, plus strains from hyaenas in 1997 and 1999 encoding 519I/549H (Fig. 4a). The network loop (Fig. 4b) indicates that one amino acid substitution in strain type 519I/549H after the epidemic would result in strains encoding 519R/549H.

A haplotype network based on 14 partial CDV-P gene nucleotide sequences (eight 'Serengeti' strains, five published 'Serengeti' lion strains (Roelke-Parker *et al*. 1996) and one African wild dog strain from Mkomazi NP (van de Bildt *et al*. 2002) in northeast Tanzania; for accession numbers, see Fig. 5), revealed that all

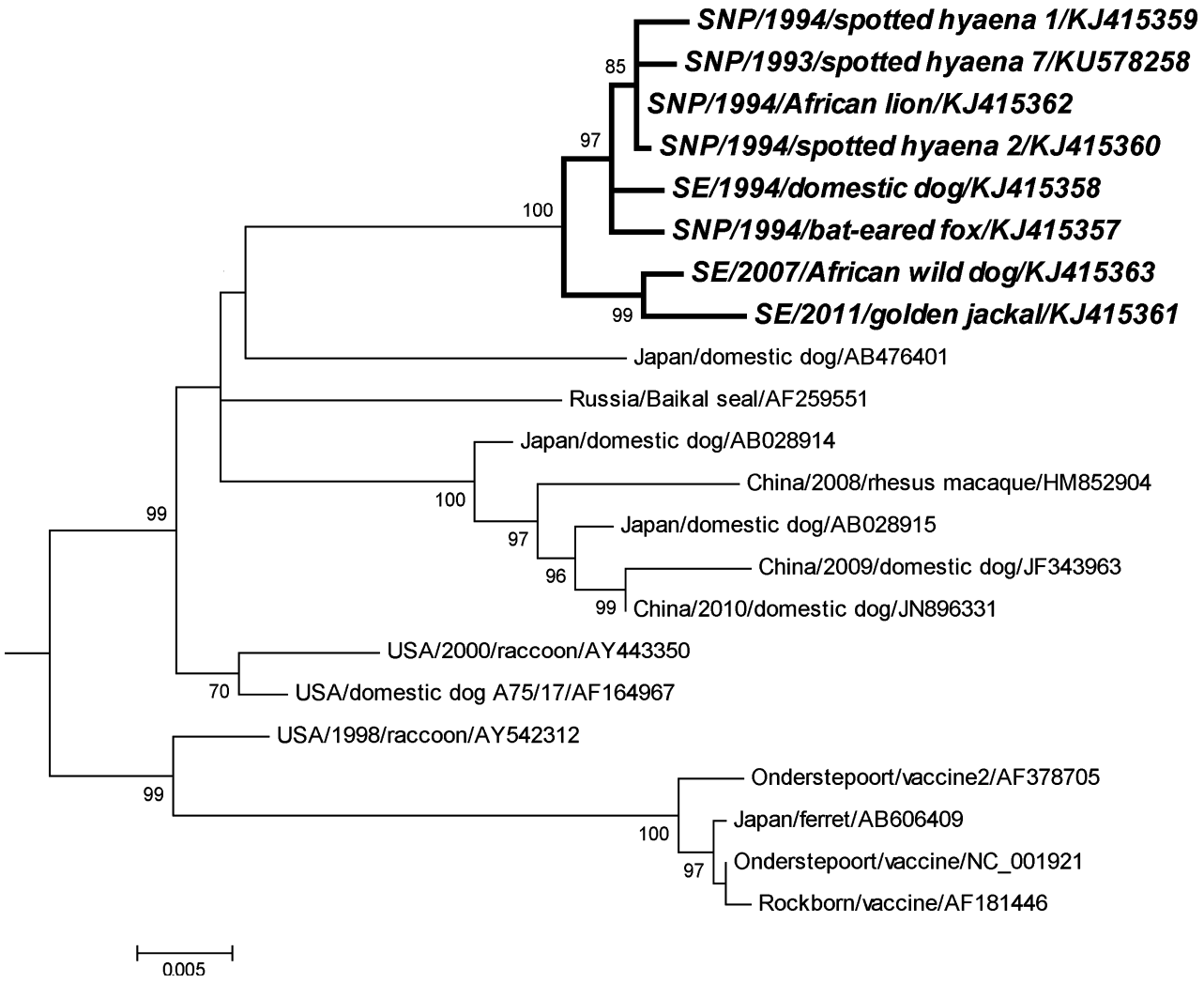


Fig. 3 The phylogenetic relationship of CDV variants from the Serengeti ecosystem to known CDV geographical lineages worldwide based on nucleotide sequences of the complete CDV-P genes (1524 nt). 'Serengeti' variants are presented on bold branches and in bold italics. Branch lengths are scaled to nucleotide substitutions per site. Statistical support for nodes was provided by 1000 replicates. Numbers at nodes indicate bootstrap values >70%. For each strain, the origin, year and host species are quoted. Strains from the Serengeti National Park (SNP) and wider Serengeti ecosystem (SE) are indicated.

haplotypes from canid hosts (plotted in red, Fig. 5a) branched from one hypothetical node as did one large noncanid (plotted in green, Fig. 5a) node representing the haplotype of four identical sequences, from which other noncanid haplotypes branched. The large noncanid node was separated from the hypothetical node by a single nucleotide substitution at site 134 in both the CDV-P and CDV-V proteins which resulted in a nonsynonymous amino acid substitution. All 'Serengeti' canid variants (Fig. 5b) encoded glycine (G) and all noncanid variants from 1993/1994 encoded serine (S). BlastX searches revealed that only noncanid 'Serengeti' variants encode S, whereas CDV variants worldwide encode G, regardless of host species (Fig. 5b). In the CDV-C protein, this nucleotide difference is a synonymous substitution (glutamine at site 126 in both canid and noncanid strains).

Haplotype networks using nucleotide sequences from the five 'Serengeti' strains were constructed for the remaining structural genes (CDV-N, CDV-M, CDV-F, CDV-L, Figs S7-S10, Supporting information, respectively) and the MF-UTR (Fig. S6b, Supporting information). Only in the conserved CDV-M gene network were all three strains from 1994 (Fig. S8, Supporting information) identical. Networks for the CDV-N, CDV-F and CDV-L genes (Figs S7, S9, S10, Supporting information, respectively) all revealed separate nodes for the 1994 lion and hyaena strains, and that these noncanid strains were separated from the 1994 dog strain by at least one hypothetical node. All

(a)

```
Strains                          500       510       519       530       540       549       560       570
SNP/1993/spotted hyaena 7/KJ415364 SQIMDKDVLTESNLVVLPTQNFIYVIATYDVSRDDHAIVYYVYDPIRTISYTHPFRLTTKGRPDFLRIECFVWDD
SNP/1994/spotted hyaena 1/KC916717 ..........................I.....................................H..........
SNP/1994/spotted hyaena 2/KC916718 ..........................I.....................................H..........
SNP/1994/African lion/JN812975     ..........................I.....................................H..........
SNP/1994/bat-eared fox/KC916716    ..........................R.....................................H..........
SE/1994/domestic dog/JN812976      ..................................G.....................A....Y..........
SNP/1996/spotted hyaena  8/KJ415365 ..................................................F..H..................
SNP/1997/spotted hyaena 3/KC916719 ..........................I.....I.....................H.............T.....
SNP/1997/spotted hyaena 4/KC916720 ..................................................F..H..................
SNP/1999/spotted hyaena 5/KC916722 ..........................I.....................F..H..................
SNP/2004/spotted hyaena 6/KC916722 ...V..................................................Y...P..............
SE/2007/African wild dog/KC916715  ..................................................F..H..................
SE/2011/golden jackal/KC916714     ..................................................Y..................
```

(b)

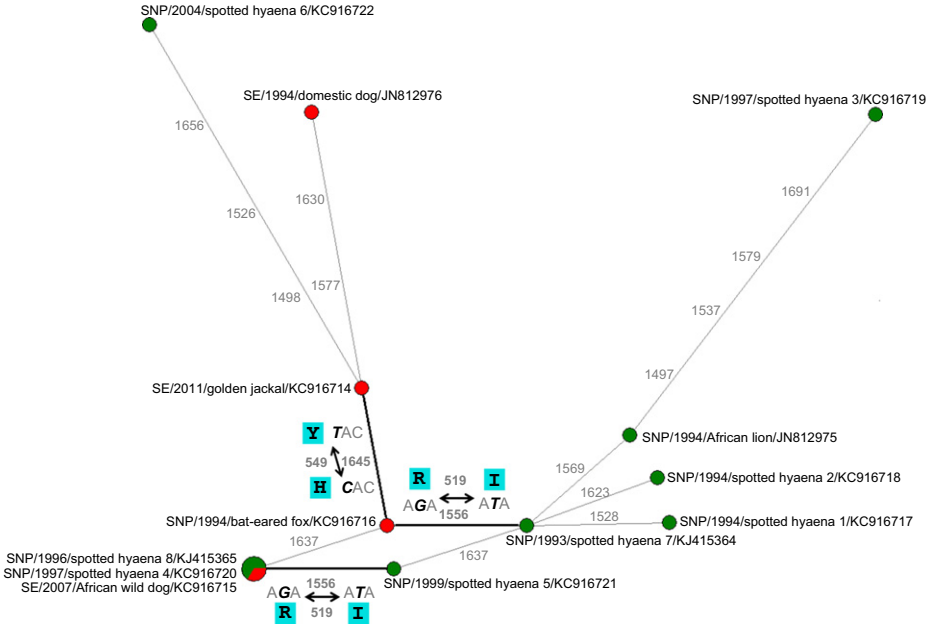


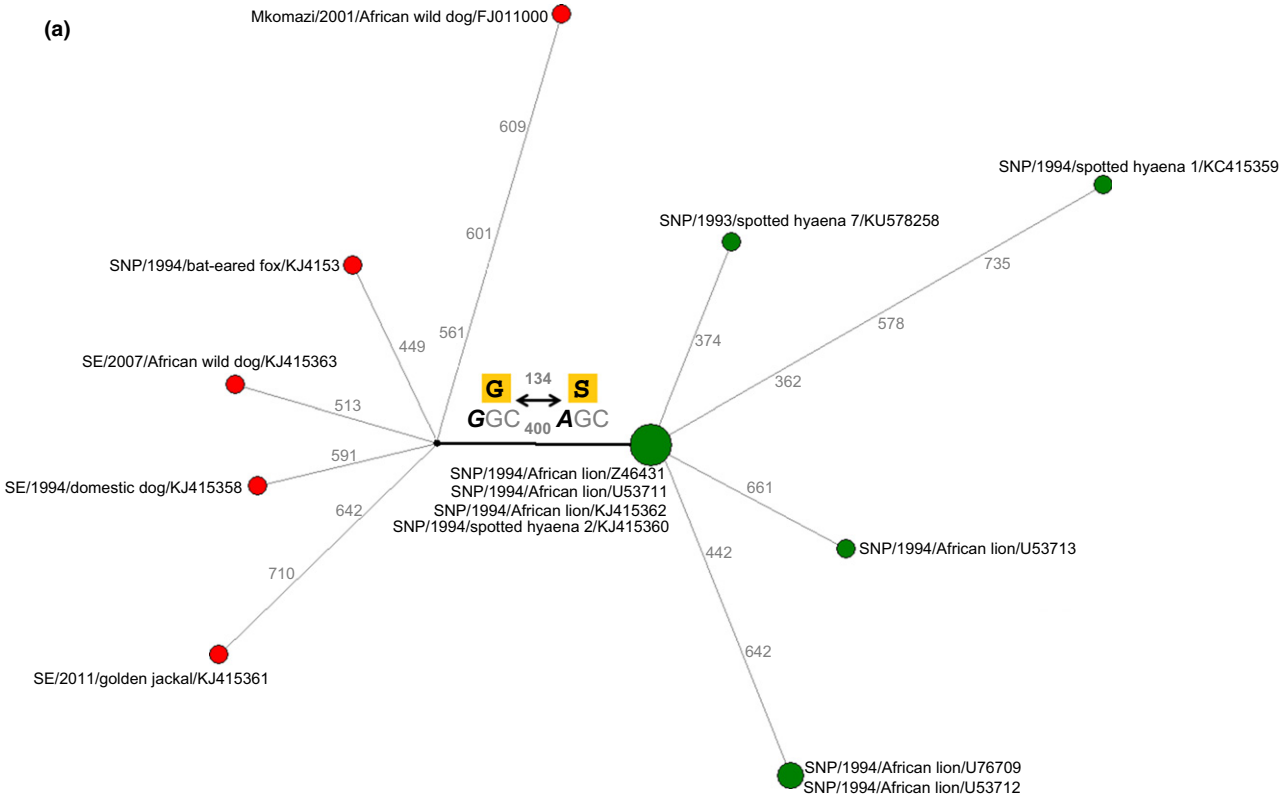
Fig. 4 The SLAM-binding region of CDV-H proteins from 'Serengeti' variants. (a) The amino acid sequence of the SLAM-binding region in CDV-H proteins from 'Serengeti' variants. Amino acid sites 519 and 549 are highlighted. The reference sequence is the variant from a spotted hyaena in 1993. (b) A haplotype network based on the SLAM-binding region from 'Serengeti' CDV variants. Circles represent haplotypes. The sizes of circles are proportional to the number of strains present at each haplotype. The proportion of strains within each haplotype from canid (red) and noncanid (green) hosts is illustrated. The branch length is proportional to the number of nucleotide changes, which are indicated along the branches. The nucleotide changes (bold italics) at position 1556 in codon 519 resulting in the substitution R519I, and at position 1645 in codon 549 resulting in substitution Y549H are indicated by arrows.

networks (Figs S7, S8, S9, S10, Supporting information) and particularly the CDV-F gene network (Fig. S9, Supporting information) revealed considerable nucleotide changes between strains from 1994 and those from 2007 and 2011.

A haplotype network using five complete 'Serengeti' CDV genome sequences (Fig. S11, Supporting information) revealed two hypothetical nodes and more than 40 nucleotide differences between the 1994 dog strain and either the hyaena or the lion strains from 1994, and

Fig. 5 (a) Haplotype network based on partial CDV-P gene fragments from strains in northern Tanzania. The network includes sequence data from 'Serengeti' variants and one African wild dog (FJ011000) variant from Mkomazi National Park in northeast Tanzania. Circles represent haplotypes; the sizes of circles are proportional to the number of strains within each node. The proportion of strains within each haplotype from canid (red) and noncanid (green) hosts is illustrated. Small black circles represent hypothetical nodes. The thicker branch indicates the nucleotide change at position 400 in codon 134 (in bold italics) that resulted in the amino acid substitution at site G134S. (b) Alignment of the amino acid sequence of the CDV-P (CDV-V) protein including residues 109 to 135 from strains recovered from 14 carnivore species, the rhesus macaque and vaccine strains. Highlighted in yellow is the unique residue substitution at position 134 observed exclusively in 'Serengeti' noncanid strains. Only amino acids that differ from those in the reference sequence at the top of the alignment are shown. Dots represent identical amino acids to the reference sequence and dashes indicate missing sequence information.

(a)



(b)

Strains	110 120 130 **134**
SE/1994/domestic dog/KJ415358	CYHVYDHSGEEVKGIADADSLVVPA**G**A
SNP/1994/bat-eared fox/KJ415357	CYYVYDHSGE.....**G**.
SE/2007/African wild dog/KJ415363	CYHVYDHSGE.....**G**.
SE/2011/golden jackal/KJ415361	CYHVYDHSGE.....**G**.
Mkomazi/2001/African wild dog/FJ011000	----------.....**G**.
SNP/1993/spotted hyaena 7/KU578258	CYHVYDHSGE.....**S**.
SNP/1994/spotted hyaena 1/KJ415359	CYHVYDHSGE..R.....**S**.
SNP/1994/spotted hyaena 2/KJ415360	CYHVYDHSGE.....G.....**S**.
SNP/1994/African lion/KJ415362	CYHVYDHSGE.....**S**.
SNP/1994/African lion/U53712	----------.....**S**.
SNP/1994/African lion/U76709	----------.....**S**.
SNP/1994/African lion/U53713	----------.....**S**.
SNP/1994/African lion/Z46431	----------.....**S**.
SNP/1994/African lion/U53711	----------.....**S**.
SNP/1994/African lion/U76708	----------.....**S**.
Russia/2004/tiger/KC579361	---------E.....E.....**G**.
Costa Rica/2006/jaquarundi/KP711845	CDHVYDHSGE.....E.....**G**T
Costa Rica/2007/puma/KP711846	CDHVYDHSGE.....E.....**G**T
Costa Rica/2007/puma/KP711847	CDHVYDHSGE.....E.....**G**T
Costa Rica/2008/ocelot/KP711848	CDHVYDHSGE.....E.....**G**T
Costa Rica/2008/ocelot/KP711849	CDHVYDHSGE.....E.....**G**T
Costa Rica/2008/ocelot/KP711850	CDHVYDHSGE.....E.....**G**T
China/2008/rhesus macaque/HM852904	CYHVYDHSGE.....E.....**G**.
China/2009/domestic dog/JF343963	CYHVYDHSGE.....E.....**G**.
China/2010/domesticdog/JN896331	CYHVYDHSGE.....E.....**G**.
Russia/Baikal seal/AF259551	CYHVYDHSGE.....E.....**G**.
USA/2000/raccoon/AY443350	CYHVYDHSGE.....E.....**G**.
USA/domestic dog A75 17/AF164967	CYHVYDHSGE.....E.....**G**.
USA/1998/raccoon/AY542312	CYHVYDHSGE.....E.....**G**T
Onderstepoort/vaccine2/AF378705	CDHVYDHSGE.....E.....**G**T
Japan/ferret/AB606409	CDHVYDHSGE.....E.....**G**T
Rockborn vaccine/AF181446	CDHVYDHSGE.....E.....**G**T

more than 200 nucleotide changes between the 1994 strains and those from 2007 and 2011.

Positive selection in the SLAM-binding region

The nested maximum-likelihood site models (Table S3, Supporting information) revealed positive selection at sites 530 and 549 in the CDV-H SLAM-binding region, as previously identified by McCarthy *et al.* (2007). The MEME model investigating site 519 in the SLAM-binding region also revealed evidence of episodic positive selection for some branches in the phylogeny (Fig. S12a, Supporting information), including those supporting the cluster of 'Serengeti' strains from noncanid hosts (EBF = 543) and those supporting strains in noncanids, that is American mink (*Neovison vison*, family Mustelidae) in Denmark, and leopard (*Panthera pardus*, family Felidae), raccoon (*Procyon lotor*, family Procyonidae) and javelina (*Pecari tajacu*, family Tayassuidae, order Artiodactyla) in the USA (all with EBF > 50). All these noncanid strains encode 519I. Results of the MEME model for site 549 also revealed positive selection at several branches in the phylogeny, including the branch supporting the cluster of 'Serengeti' strains from noncanid hosts (EBF > 1000) (Fig. S12b, Supporting information).

Together, these analyses provided evidence of positive selection at sites 519, 530 and 549 in the SLAM-binding region. The residue at site 530 was identical in all 'Serengeti' strains (Fig. 4a); hence, we focus on sites 519 and 549. Six 'Serengeti' strains (Fig. 4a), all from noncanids, encoded 519I, including all those from the 1993/1994 epidemic (Fig. 4a), and all strains from canids specified 519R (Fig. 4a), including those from the 1993/1994 epidemic. To our knowledge, only five strains encoding 519I were known before our study: all from noncanids manifesting clinical disease (Table S1, Supporting information).

The amino acid combinations at positions 519 and 549 in the SLAM-binding region of 'Serengeti' strains (Fig. 4a) were as follows: (i) 519I/549H six strains from noncanids; (ii) 519R/549Y two strains from canids and one strain from a noncanid; and (iii) 519R/549H two strains from canids and two strains from noncanids. Strains encoding combination 519R/549Y or 519R/549H caused fatal disease in three canid species, but three hyaenas were asymptomatic (Table S5, Supporting information), including one adult that was blood sampled in 1997 which survived until 2012. Strains encoding 519I/549H were found in three noncanids with clinical disease during the 1993/1994 epidemic (one lion and two juvenile hyaenas), and also in asymptomatic hyaenas, including one adult in 1994, one juvenile in 1997 killed by a lion and an adult in 1999 (Table S5, Supporting information).

Before our study, only four strains encoded 519I/549H were described (Table S1, Supporting information) and our phylogenetic analysis (Fig. 2) placed these strains in the America II lineage, including two strains from Denmark. Our study reports six 519I/549H strains in the East African lineage. Hence, CDV-H proteins encoding 519I/549H are currently only known from noncanid hosts, including lion and hyaena (this study), leopard, raccoon and American mink (Table S1, Supporting information). No 'Serengeti' strain encoded 519I/549R: this combination is only known from the javelina in the USA (Table S1, Supporting information).

During our monitoring period (1993–2011), wild canids succumbed to strains encoding 519R/549H and 519R/549Y, and a dog to a strain encoding 519R/549Y. No canid species was found infected with a strain specifying 519I/549H (Fig. 4a).

Additive effect of amino acids at sites 519 and 549

We tested the functional importance of amino acid substitutions Y549H and R519I and their possible additive effects on the ability of the CDV-H protein to bind to canid and noncanid SLAM (CD150) receptors using an established assay (Nikolin *et al.* 2012b). This assay quantified the MNN per syncytium as a proxy measure of the binding ability of CDV-H proteins to cells expressing SLAM receptors of a canid (dog) or two noncanid species (lion, domestic cat) and the consequent fusion activity of CDV-F protein. For this purpose, we used three wild-type CDV-H proteins, two from dog isolates encoding 519R/549Y (*SE/1994/domestic dog* and *A75/17)* and one from the 'Serengeti' lion isolate (*SNP/1994/African lion*) encoding 519I/549H. We also engineered three CDV-H mutants of the strain *A75/17*: (1) residue 519R was substituted for 519I to produce mutant *A75/17-519I*; (2) residue 549Y was substituted for 549H to produce mutant *A75/17-549H*; and (3) the residue at both sites were substituted to produce mutant *A75/17-519I/549H*.

The MNN per syncytium significantly varied between the six CDV-H proteins tested (ANOVA-type statistic, ATS $F = 96.18$, d.f. = 1.66 and 62.3, $P < 0.0001$, Fig. 6) and between cells expressing SLAM receptors from the three different carnivores tested (ATS $F = 20.29$, d.f. = 2.31 and 18, $P < 0.0001$, Fig. 6). The extent of the variation in MNN between CDV-H proteins significantly depended on the origin of the SLAM receptor (ATS $F = 15.69$, d.f. = 4.04 and 62.3, $P < 0.0001$, Fig. 6).

The post hoc comparison of the two CDV-H proteins with the 519I/549H combination (i.e. *A75/17-519I/549H* and *SNP/1994/African lion*) confirmed that there was no difference between these two proteins in the MNN produced (ATS $F = 0.93$, d.f. = 1 and 13.9, NS). Both

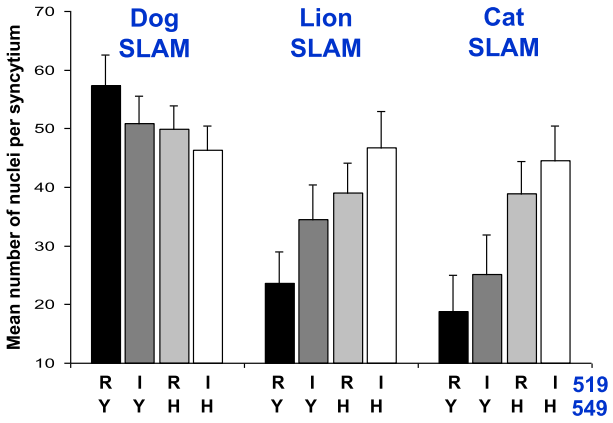


Fig. 6 The functional impact of different amino acid combinations at sites 519 and 549 in the SLAM-binding region of CDV-H proteins. The impact of combining amino acids isoleucine [I] or arginine [R] at site 519 with tyrosine [Y] or histidine [H] at site 549 on the performance of CDV-H proteins in terms of the mean number of nuclei (MNN) per syncytium formed, in cell cultures expressing SLAM receptors from three carnivore species. Black bars – proteins with combination 519R/549Y (from variants *SE/1994/domestic dog* and *A75/17*). Dark grey bars – protein with combination 519I/549Y (from mutant *A75/17-519I*). Light grey – protein with combination 519R/549H (from mutant A75/17-549H). White bars – proteins with combination 519I/549H (from variant *SNP/1994/African lion* and mutant *A75/17 519I-549H*).

proteins responded differently to the SLAM receptors of the three carnivores tested (ATS $F = 4.39$; d.f. = 1.67 and 10.8; $P = 0.045$, Fig. 6); the effect of the SLAM receptors did not depend on the origin (strain) of the CDV-H protein (ATS $F = 0.95$, d.f. = 1.93 and 13.9, NS).

Nectin-4 receptor binding region

The five amino acid sites in the CDV-H protein (those highlighted in Fig. S13, Supporting information) required for virus binding to the nectin-4 receptor were identical in all 'Serengeti' strains.

Discussion

We report differences in the distribution and infection outcome of genetically distinct CDV strains in canid and noncanid hosts in the Serengeti ecosystem. We found that lions and hyaenas during the 1993/1994 epidemic were infected by strains encoding the rare amino acid combination 519I/549H in the CDV-H SLAM-binding region (Fig. 4) and one novel G134S substitution (Fig. 5) in the nonstructural CDV-V protein. Canids were not infected by this strain type, and during the epidemic, a bat-eared fox was fatally infected with a strain encoding CDV-H 519R/549H and CDV-V 134G, and one dog was fatally infected with a strain encoding CDV-H 519R/549Y and CDV-V 134G. Although just two nucleotide substitutions (Fig. 4b) convert the SLAM-binding region from the most commonly described amino acid combination 519R/549Y to 519I/549H, evidence of substitution R519I is lacking from all CDV lineages except for the America II lineage and the East African lineage (Table S1, Supporting information). All strains encoding 519I/549H in both these lineages were from noncanid hosts. To our knowledge, we are the first to report strains encoding CDV-V 134S, which came exclusively from noncanids during the 1993/1994 epidemic.

Phylogenetic results from five structural CDV genes (Figs 2 and 3, Figs S4 and S5, Supporting information) and the complete CDV genomes from 'Serengeti' strains (Fig. 1a) revealed that strains from lions and hyaenas during the 1993/1994 epidemic were more closely related to each other than to strains from a dog and a bat-eared fox: only the conserved CDV-M gene (Fig. S3, Supporting information) did not show this topology. These findings are consistent with the idea that the 1993/1994 epidemic in lions and hyaenas was caused by the evolution, in the diverse wild carnivore guild in the Serengeti ecosystem, of a strain better adapted to noncanids than canids. Similarly, genetically distinct strains of canine kobuvirus in this ecosystem circulate in domestic dogs and wild carnivores (Olarte-Castillo *et al.* 2015), and distinct coronavirus strains circulate in wild canid and noncanid hosts (Goller *et al.* 2013).

Nested maximum-likelihood site substitution models applied to sequences of the CDV-H SLAM-binding region indicated positive selection at sites 530 and 549 (Table S3, Supporting information) as reported by McCarthy *et al.* (2007). The functional importance of site 549 has been discussed (McCarthy *et al.* 2007; Nikolin *et al.* 2012b). As the amino acid at site 530 was identical in all 'Serengeti' strains (Fig. 4a), we do not discuss this site further. In contrast to the previous models, MEME models (Murrell *et al.* 2012) search for both transient and episodic positive selection, and applying this method to amino acid sites 519 and 549 (Fig. 4a) revealed significant evidence of selection pressure at both sites (Fig. S12, Supporting information), and specifically episodic selection at site 519 (Fig. S12a, Supporting information) on phylogeny branches supporting strains encoding 519I from noncanids (Table S1, Supporting information) in both the East African lineage (hyaena and lion) and America II lineage (leopard, raccoon, javelina, American mink).

Our finding that strains encoding 519I/549H cause fatal outcomes only in noncanids (lion and hyaena)

during the 1993/1994 epidemic but not in canids is consistent with previous reports of the virulence of this strain type to noncanids (Table S1, Supporting information) worldwide (Appel *et al.* 1994; Harder *et al.* 1996; Bolt *et al.* 1997; Trebbien *et al.* 2014). No strain encoding 519I/549R was found in the Serengeti ecosystem, but this strain type caused a fatal CDV in javelina (Appel *et al.* 1991; Table S1, Supporting information). As we only screened carnivore species for CDV infection, we cannot exclude the possibility that strains encoding 519I/549R might occur in the Serengeti ecosystem in a noncarnivore such as the warthog (*Phacochoerus aethiopicus*). We only retrieved strains encoding CDV-H 519R/549Y or 519R/549H from canid species (Fig. 4a) with fatal disease (Table S5, Supporting information) and from asymptomatic hyaenas (Table S5, Supporting information). In line with the vast majority of dog strains worldwide (Table S1, Supporting information), the strain from a dog during the 1993/1994 epidemic specified CDV-H 519R/549Y (Fig. 4a).

Our functional test of the additive effect (Fig. 6) of the amino acid combinations at sites 519 and 549 in the CDV-H SLAM-binding region revealed that 519R/549Y proteins had the highest performance in cells expressing dog SLAM receptors and a particularly poor performance in cells expressing lion SLAM receptors and cat SLAM receptors (Fig. 6). In contrast, CDV-H proteins encoding 519I/549H performed relatively well in cells expressing lion SLAM receptors, cat SLAM receptors and dog SLAM receptors, but the performance of this combination in cells expressing dog SLAM was significantly below that of proteins specifying 519R/549Y, 519R/549H or 519I/549Y. These findings suggest that CDV-H proteins encoding 519I/549H exhibited the strongest 'generalist' traits, those specifying 519R/549Y displayed traits indicative of a dog 'specialist' strain (*sensu* Nikolin *et al.* 2012b) and those encoding 519R/549H or 519I/549Y showed traits between these two extremes. During the first infection stage, dogs would therefore be particularly prone to strains encoding 519R/549Y, whereas lions would be particularly prone to infection by those encoding 519I/549H and far less so to infection with strains encoding 519R/549Y (Fig. 6). Hence, our results indicate that strain type 519R/549Y would outcompete 519I/549H when infecting domestic dog immune cells and strain type 519I/549H would outcompete both 519R/549Y and 519R/549H when infecting lion immune cells.

During the 1993/1994 epidemic, the outcome of infection varied across species and with strain type. Adult hyaenas were asymptomatic when infected with variants encoding 519I/549H that were fatal to juvenile hyaenas (Haas *et al.* 1996). Asymptomatic infection most likely represents a 'dead end' for CDV because virus shedding probably does not occur (Sawatsky *et al.* 2012). This suggests that most adult hyaenas in the Serengeti ecosystem are 'dead-end' hosts for CDV, regardless of strain type, although lions might be infected if they killed an asymptomatic hyaena and hence contacted strain type 519I/549H from its blood. Juvenile hyaenas developed clinical disease during the 1993/1994 epidemic and were presumably the main shedders and transmitters of CDV within hyaena clans. By contrast, lions regardless of age developed clinical signs during this epidemic, hence shed and transmitted strain type 519I/549H within prides. We have no evidence that strain types 519R/549Y or 519R/549H in the East African lineage caused clinical disease in either lions or hyaenas thus suspect that these noncanids are 'dead-end' hosts for them (Table S5, Supporting information). Our results and those from the America II lineage indicate that strain type 519I/549H does not cause clinical CDV in canids. Hence, it is unlikely that canids transmitted this strain type to lions and hyaenas during the 1993/1994 epidemic.

The host range of CDV strains and the outcome of infection will depend on virus traits other than those affecting virus binding to SLAM receptors. Considering the five amino acid sites of the CDV-H protein thought important for binding measles virus and CDV to nectin-4 receptors (Langedijk *et al.* 2011; Sawatsky *et al.* 2012), we found no evidence for adaptations to noncanid hosts at these sites (Fig. S13, Supporting information).

We report a novel G134S amino acid substitution in CDV-V proteins (Fig. 5) only in noncanid strains from the 1993/1994 epidemic obtained by us and others (Fig. 5b). This substitution may influence strain virulence, because the N-terminal of the CDV-V protein disrupts interferon signalling in hosts (von Messling *et al.* 2006; Röthlisberger *et al.* 2010), thereby influencing the outcome of infection (Randall & Goodbourn 2008). No other publically available CDV sequences currently encode this substitution (Fig. 5b), including those from wild felids (e.g. Harder *et al.* 1996; Avendaño *et al.* 2016).

Since detailed carnivore research started in the Serengeti NP in the 1960s, clinical CDV has been reported in wild canids, including African wild dogs (Schaller 1972; Goller *et al.* 2010), silver-backed jackals (Moehlman 1983; Roelke-Parker *et al.* 1996), bat-eared foxes (Roelke-Parker *et al.* 1996) and golden jackals (Roelke-Parker *et al.* 1996; this study). In contrast, research on hyaenas and lions during a similar period (e.g. Kruuk 1972; Schaller 1972; Bertram 1973; Hanby & Bygott 1979; Hofer & East 1995; Munson *et al.* 2008) revealed no evidence of clinical CDV before 1993 (Haas *et al.* 1996), except for possibly two lions noted with nervous twitches (Schaller 1972). This suggests that apart from

the period spanning the 1993/1994 epidemic, wild canids were more likely to manifest clinical distemper than noncanids.

In the years following the 1993/1994 epidemic, serological evidence revealed periods when exposure to CDV increased in hyaenas and lions (Harrison *et al.* 2004; Munson *et al.* 2008; Viana *et al.* 2015), a phenomenon termed 'silent' epidemics (Harrison *et al.* 2004; Munson *et al.* 2008). Our results suggest that lions and hyaenas typically clear canid-adapted CDV strains before clinical sign and virus shedding occurs and that outbreaks of canid-adapted strains in wild canid populations cause asymptomatic exposure in noncanids. These benign outcomes most likely explain 'silent' outbreaks in lions and hyaenas. Fatal CDV in an African wild dog pack in 2007 (Goller *et al.* 2010) and a golden jackal pup in 2011 is consistent with this idea, as both occurred when serological data from lions indicated elevated asymptomatic exposure (Viana *et al.* 2015). Asymptomatic exposure to canid strains would increase herd immunity (Guiserix *et al.* 2007) and provide lifelong protection against future infection among exposed noncanids.

We are unaware of CDV disease epidemics in lions and hyaenas in other African ecosystems, although serological studies report exposure to CDV in lions, hyaenas and other noncanid species such as cheetah (*Acinonyx jubatus*, family Felidae), leopard and caracal (*Felis caracal*, family Felidae), without evidence of clinical disease (e.g. Munson *et al.* 2004; Driciru *et al.* 2006; Alexander *et al.* 2010; Thalwitzer *et al.* 2010). A likely explanation for this phenomenon is the exposure of noncanids to canid-adapted CDV strains. Exposure of carnivores to other morbilliviruses such as rinderpest in the past (Dehay *et al.* 1965) and currently to peste des petits ruminants (PPR) (Swai *et al.* 2009) through the consumption of infected ungulates is also thought to induce antibody protection against CDV, and could be mistakenly interpreted as evidence of CDV exposure.

Models of the 1993/1994 epidemic in lions in the Serengeti ecosystem assume transmission of a single CDV genetic variant among wild carnivores (both canids and noncanids) and dogs (Craft *et al.* 2008, Craft *et al.* 2009; Viana *et al.* 2015). Our results question this assumption by revealing genetically distinct strains in canids and noncanids during the epidemic (Figs 2–5) and by experimentally showing that canids would be unlikely to propagate the CDV-H 519I/549H CDV-V 134S strain type during the 1993/1994 epidemic in lions (Fig. 6). In the hyaena population, virus shedders of strain type CDV-H 519I/549H CDV-V 134S during the disease epidemic were predominantly juveniles based at clan communal dens. In fission–fusion hyaena clans, communal dens are important social centres (Hofer & East 1995) that can facilitate pathogen transmission within clans (East *et al.* 2013). For this reason, exposure of adult clan members to virus shedding juveniles should relatively swiftly elevate herd immunity to CDV in clans and result in rapid 'burn-out' of naïve juveniles at communal dens. We consider that juvenile hyaenas with clinical CDV at communal dens would rarely propagate the epidemic in lion prides. We suspect other noncanids contributed more significantly to the spread of CDV in the lion population during the 1993/1994 disease epidemic.

It was argued that stark differences in CDV virulence in lions during the 1993/1994 epidemic and later 'silent' outbreaks resulted from heavier burdens of *Babesia* infection in 1994 (Munson *et al.* 2008). This explanation assumes that variants infecting lions in 1994 and later years had approximately the same virulence in lions and that the difference in infection outcome was determined by the burden of *Babesia*. Our findings question this idea and rather suggest that the strain infecting lion during the 1993/1994 epidemic more severely compromised their immune responses, resulting in heavier burdens of *Babesia* infection.

Although the CDV epidemic in lions and hyaenas is thought to have ended by 1995, we found two hyaenas infected with strain type 519I/549H after this period (Fig. 4a), one juvenile in 1997 and one adult in 1999. Whether infection of the juvenile contributed to its predation by a lion is unknown (Table S5, Supporting information). Between 2000 and 2011, strains encoding 519I/549H were not found, suggesting their possible fade-out in the Serengeti NP. Even so, we cannot exclude the possibility that this strain type continued to circulate in other noncanid species, possibly in small felids and species from the family Viverridae, such as common genet (*Genetta genetta*), which is susceptible to CDV infection (López-Peña *et al.* 2001). Of the 26 wild carnivore species (Sinclair & Arcese 1995) known to occur in the Serengeti ecosystem, five belong to the Canidae and of these, four have been reported with fatal CDV infection. Besides lions and spotted hyaenas, currently little is known about CDV infection in the 19 other noncanid species present in the ecosystem.

Knowledge of the molecular ecology of CDV in areas containing many noncanid species is limited (Alexander *et al.* 2010; Almberg *et al.* 2010). Our study reveals a far greater complexity to the molecular epidemiology of CDV in the Serengeti ecosystem than is currently recognized, involving distinct variants that differ in their virulence to canid and noncanid species in a diverse carnivore guild. The challenge will be to incorporate our improving knowledge of molecular epidemiology into mathematical models to more accurately predict the probable impact of CDV in multihost communities and threatened species.

Acknowledgements

We thank the COSTECH, TAWIRI, TANAPA and the NCAA for their support. We thank Maximilian Driller, Annie Francis, Jörns Fickel, Bjorn Figenschou, Alex Greenwood and Stephan Karl for assistance, and François Balloux and three anonymous referees for their helpful comments. This study was funded by the Leibniz Institute for Zoo and Wildlife Research, the Deutsche Forschungsgemeinschaft through grant EA5/2-1 and the DFG-GRAKO 1121, within the ZIBI, the Freie Universität Berlin, and the EcoHealth Alliance.

References

Alexander KA, McNutt JW, Briggs MB *et al.* (2010) Multi-host pathogens and carnivore management in southern Africa. *Comparative Immunology, Microbiology and Infectious Diseases*, **33**, 249–265.

Allison AB, Kohler DJ, Ortega A *et al.* (2014) Host-specific parvovirus evolution in nature is recapitulated by *In Vitro* adaptation to different carnivore species. *PLoS Pathogens*, **10**, e1004475.

Almberg ES, Cross PC, Smith DW (2010) Persistence of canine distemper virus in the Greater Yellowstone Ecosystem's carnivore community. *Ecological Applications*, **20**, 2058–2074.

Anderson DE, von Messling V (2008) Region between the canine distemper virus M and F gene modulates virulence by controlling fusion gene expression. *Journal of Virology*, **82**, 10510–10518.

Antia R, Regoes RR, Koella JC, Bergstrom CT (2003) The role of evolution in the emergence of infectious diseases. *Nature*, **426**, 658–661.

Appel MJG, Summers BA (1995) Pathogenicity of morbilliviruses for terrestrial carnivores. *Veterinary Microbiology*, **44**, 187–191.

Appel MJG, Sheffy BE, Percy DH, Gaskin JM (1974) Canine distemper in domestic cats and pigs. *American Journal of Veterinary Research*, **35**, 803–806.

Appel MJG, Reggiardo C, Summers BA *et al.* (1991) Canine distemper virus infection and encephalitis in javelinas (collared peccaries). *Archives in Virology*, **119**, 147–152.

Appel MJG, Yates RA, Foley GL *et al.* (1994) Canine distemper epizootic in lions, tigers and leopards. *Journal of Veterinary Diagnostic Investigation*, **6**, 277–288.

Avendaño R, Barrueta F, Soto-Fournier S, Chavarría Monge O, Gutiérrez-Espeleta Chaves A (2016) Canine distemper virus in wild felids of Costa Rica. *Journal of Wildlife Diseases*, **52**, 373–377.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bertram BCR (1973) Lion population regulation. *African Journal of Ecology*, **11**, 215–225.

van de Bildt MWG, Kuiken T, Visee AM, Lema S, Fitzjohn TR, Osterhaus ADME (2002) Distemper outbreak and its effect on African wild dog conservation. *Emerging Infectious Diseases*, **8**, 211–221.

Bolt G, Jensen TD, Gottschalck E *et al.* (1997) Genetic diversity of the attachment (H) protein gene of current field isolates of canine distemper virus. *Journal of General Virology*, **78**, 367–372.

Brunner E, Puri ML (2001) Nonparametric methods in factorial designs. *Statistical Papers*, **42**, 1–52.

Brunner E, Munzel U, Puri ML (1999) Rank-score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, **70**, 286–317.

Craft ME, Hawthorne PL, Packer C, Dobson AP (2008) Dynamics of a multihost pathogen in a carnivore community. *Journal of Animal Ecology*, **77**, 1257–1264.

Craft ME, Volz E, Packer C, Meyers LA (2009) Distinguishing epidemic waves from disease spillover in a wildlife population. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **276**, 1777–1785.

Crill WD, Wichman HA, Bull JJ (2000) Evolutionary reversals during viral adaptation to alternating hosts. *Genetics*, **154**, 27–37.

Dehay PD, Stone SS, Karzon DT, Katz S, Enders J (1965) Clinical and immune responses of alien host to inoculation with measles, rinderpest, and canine distemper virus. *American Journal of Veterinary Research*, **26**, 1359–1373.

Driciru M, Siefer L, Prager KC *et al.* (2006) A serosurvey of viral infections in lions (*Panthera leo*), from Queen Elizabeth National Park, Uganda. *Journal of Wildlife Diseases*, **42**, 667–671.

East ML, Nyahongo JW, Goller KV, Hofer H (2012) Does the vastness of the Serengeti limit human-wildlife conflicts? In: *Fencing for Conservation: Restriction of Evolutionary Potential or a Riposte to Threatening Processes?* (eds Somers MJ, Hayward MW), pp. 125–151. Springer, New York, New York.

East ML, Kurze C, Wilhelm K, Benhaiem S, Hofer H (2013) Factors influencing *Dipylidium* sp. infection in a free-ranging social carnivore, the spotted hyaena (*Crocuta crocuta*). *International Journal for Parasitology: Parasites and Wildlife*, **2**, 257–265.

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.

Faria NR, Suchard MA, Rambaut A, Streicker DG, Lemey P (2013) Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, **368**, 20120196.

Fischer CDB, Gräf T, Ikuta N *et al.* (2016) Phylogenetic analysis of canine distemper virus in South America clade 1 reveals unique molecular signatures of the local epidemic. *Infection, Genetics and Evolution*, **41**, 135–141.

Flynn JJ, Finarelli JA, Zehr S, Hsu J, Nedbal MA (2005) Molecular phylogeny of the carnivore (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. *Systematic Biology*, **54**, 317–337.

Gadagkar SR, Kumar S (2005) Maximum likelihood outperforms maximum parsimony even when evolutionary rates are heterotachous. *Molecular Biology and Evolution*, **22**, 2139–2141.

Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, **16**, 573–582.

Goller KV, Fyumagwa RD, Nikolin VM *et al.* (2010) Fatal canine distemper infection in a pack of African wild dogs in the Serengeti ecosystem, Tanzania. *Veterinary Microbiology*, **146**, 245–252.

Goller KV, Fickel J, Hofer H, Beirer S, East ML (2013) Coronavirus genotype diversity and prevalence of infection in the

Serengeti National Park, Tanzania. *Archives of Virology*, **158**, 729–734.

Greenwood AD, Tsangaras K, Ho SYW *et al.* (2012) A potential fatal mix of herpes in zoos. *Current Biology*, **22**, 1727–1731.

Guiserix M, Bahi-Jaber N, Fouchet D, Sauvage F, Pontier D (2007) The canine distemper epidemic in the Serengeti: are lions victims of a new highly virulent canine distemper virus strain, or is pathogen circulation stochasticity to blame? *Journal of the Royal Society of London Interface*, **17**, 1127–1134.

Haas L, Hofer H, East M, Wohlsein P, Liess B, Barrett T (1996) Canine distemper virus infection in Serengeti spotted hyaenas. *Veterinary Microbiology*, **49**, 147–152.

Hanby JP, Bygott JD (1979) Population changes in lions and other predators. In: *Serengeti, Dynamics of an Ecosystem* (eds Sinclair ARE, Norton-Griffiths M), pp. 249–262. Chicago University Press, Chicago, Illinois.

Harder TC, Kenter M, Vos H *et al.* (1996) Canine distemper virus from diseased large felids: biological properties and phylogenetic relationships. *Journal of General Virology*, **77**, 397–405.

Harrison TM, Mazet JK, Holekamp KE *et al.* (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.

Hasegawa M, Kishino H, Yano T (1985) Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hofer H, East ML (1995) Population dynamics, population size, and the commuting system of Serengeti spotted hyaenas. In: *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem* (eds Sinclair ARE, Arcese P), pp. 332–363. University of Chicago Press, Chicago, Illinois.

Hofer H, East ML, Campbell KLI (1993) Snares, commuting hyaenas and migratory herbivores: humans as predators in the Serengeti. *Symposia of the Zoological Society London*, **65**, 347–366.

Ikeda Y, Nakamura K, Miyazawa T *et al.* (2001) Seroprevalence of canine distempervirus in cats. *Clinical and Diagnostic Laboratory Immunology*, **8**, 641–644.

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in Biosciences*, **8**, 275–282.

Kameo Y, Nagao Y, Nishio Y *et al.* (2012) Epizootic canine distemper virus infection among wild mammals. *Veterinary Microbiology*, **154**, 222–229.

Keller SM, Gabriel M, Terio KA *et al.* (2012) Canine distemper in an isolated population of fishers (*Martes pennanti*) from California. *Journal of Wildlife Diseases*, **48**, 1035–1041.

Kruuk H (1972) *The Spotted Hyena*. Chicago University Press, Chicago, Illinois.

Lamb RA, Kolakofsky D (2001) *Paramyxoviridae*: the viruses and their replication. In: *Fields Virology*, 4th edn (eds Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus ES), pp. 1305–1340. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.

Langedijk JPM, Janda J, Origgi FC *et al.* (2011) Canine distemper virus infects canine keratinocytes and immune cells by using overlapping and distinct regions located on one side of the attachment protein. *Journal of Virology*, **85**, 11242–11254.

Lole KS, Bollinger RC, Paranjape RS *et al.* (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*, **73**, 152–160.

Longdon B, Bockhurst MA, Russell CA, Welch JL, Jiggins FM (2014) The evolution and genetics of virus host shifts. *PLOS Pathogens*, **10**, e1004395.

López-Peña M, Vázquez S, Alemañ N *et al.* (2001) Canine distemper in a genet (*Gennetta gennetta*), associated with endogenous lipid pneumonia. *Journal of Comparative Pathology*, **124**, 207–211.

Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE*, **5**, e14004.

Martella V, Elia G, Lucente M *et al.* (2006) Heterogeneity within the hemagglutinin genes of canine distemper virus (CDV) strains detected in Italy. *Veterinary Microbiology*, **116**, 301–309.

Martin D, Rybicki E (2000) RDP: detection of recombination among aligned sequences. *Bioinformatics*, **16**, 562–563.

Martin DP, Williamson C, Posada D (2005a) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, **21**, 260–262.

Martin DP, Posada D, Crandall KA, Williamson C (2005b) A modified bootscan algorithm for automated identification of recombinant sequences and recombinant breakpoints. *Aids Research Human Retroviruses*, **12**, 98–102.

Maynard Smith J (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126–129.

McCarthy AJ, Goodman SJ (2010) Reassessing conflicting evolutionary histories of the Paramyxoviridae and the origins of respiroviruses with Bayesian multigene phylogenies. *Infection, Genetics and Evolution*, **10**, 97–107.

McCarthy AJ, Shaw M-A, Goodman SJ (2007) Pathogen evolution and disease emergence in carnivores. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **274**, 3165–3174.

Meli ML, Simmler P, Cattori V *et al.* (2010) Importance of canine distemper virus CDV infection in free-ranging Iberian lynxes (*Lynx pardinus*). *Veterinary Microbiology*, **146**, 132–137.

von Messling V, Zimmer G, Herrler G, Haas L, Cattaneo R (2001) The hemagglutinin of canine distemper virus determines tropism and cytopathogenicity. *Journal of Virology*, **75**, 6418–6427.

von Messling V, Springfeld C, Devaux P, Cattaneo R (2003) A ferret model of canine distemper virulence and immunosuppression. *Journal of Virology*, **77**, 12579–12591.

von Messling V, Svitek N, Cattaneo R (2006) Receptor (SLAM [CD150]) recognition and the V protein sustain swift lymphocyte-based invasion of mucosal tissue and lymphatic organs by a morbillivirus. *Journal of Virology*, **80**, 6084–6092.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, pdb.prot5448. doi: 10.1101/pdb.prot5448.

Moehlman PD (1983) Socioecology of silver-backed and golden jackals, *Canis mesomelas* and *C. aureus*. In: *Recent Advances in the Study of Mammalian Behavior* (eds Eisenberg JF, Kleiman DG), pp. 423–438. American Society of Mammalogists, Lawrence.

Monne I, Fusaro A, Valastro V *et al.* (2011) Distinct CDV genotype causing a major epidemic in Alpine wildlife. *Veterinary Microbiology*, **150**, 63–69.

Moya A, Elena SF, Bracho A, Miralles R, Barrio E (2000) The evolution of RNA viruses: a population genetics view. *Proceedings of the National Academy of Sciences USA*, **97**, 6967–6973.

Moya A, Holmes EC, Gonzalez-Candelas F (2004) Population genetics and evolutionary epidemiology of RNA viruses. *Nature Reviews Microbiology*, **2**, 279–288.

Munson L, Marker L, Dubovi E, Spencer JA, Everman JF, O'Brien SJ (2004) Serosurvey of viral infections in free-ranging Namibian cheetahs (*Acinonyx jubatus*). *Journal of Wildlife Diseases*, **40**, 23–31.

Munson L, Terio KA, Kock R *et al.* (2008) Climate extremes promote fatal co-infections during canine distemper epidemics in African lions. *PLoS ONE*, **3**, e2545.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, **8**, e1002764.

Nagao Y, Nishio Y, Shiomoda H *et al.* (2012) An outbreak of canine distemper virus in tigers (*Panthera tigris*): Possible transmission from wild animals to zoo animals. *Journal of Veterinary Medical Science*, **74**, 699–705.

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**, 929–936.

Nikolin VM, Wibbelt G, Michler F-UF, Wolf P, East ML (2012a) Susceptibility of carnivore hosts to strains of canine distemper virus from distinct genetic lineages. *Veterinary Microbiology*, **156**, 45–53.

Nikolin VM, Osterrieder K, von Messling V *et al.* (2012b) Antagonistic pleiotropy and fitness trade-offs reveal specialist and generalist traits in strains of canine distemper. *PLoS ONE*, **7**, e50955.

Noyce RS, Delpeut S, Richardson CD (2013) Dog nectin-4 is an epithelial cell receptor for canine distemper virus that facilitates virus entry and syncytia formation. *Virology*, **436**, 210–220.

Olarte-Castillo XA, Heeger F, Mazzoni C *et al.* (2015) Molecular characterization of canine kobuvirus in wild carnivores and the domestic dog in Africa. *Virology*, **477**, 89–97.

Origgi FC, Plattet P, Sattler U *et al.* (2012) Emergence of canine distemper strains with modified molecular signature and enhanced neuronal tropism leading to high mortality in wild carnivores. *Veterinary Pathology*, **49**, 913–929.

Panzera Y, Sarute N, Iraola G, Hernández M, Pérez R (2015) Molecular phylogeography of canine distemper virus: geographical origin and global spread. *Molecular Phylogenetics and Evolution*, **92**, 147–154.

Pomeroy LW, Bjørnstad ON, Holmes EC (2008) The evolutionary and epidemiological dynamics of the Paramyxoviridae. *Journal of Molecular Evolution*, **66**, 98–106.

Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences USA*, **98**, 13757–13763.

Randall RE, Goodbourn S (2008) Interferons and viruses: and interplay between induction, signalling, antiviral responses and countermeasures. *Journal of General Virology*, **89**, 1–47.

Read AF, Taylor LH (2001) The ecology of genetically diverse infections. *Science*, **292**, 1099–1102.

Riley MC, Wilkes RP (2015) Sequencing of emerging canine distemper virus strain reveals new distinct genetic lineage in the United States associated with disease in wildlife and domestic canine populations. *Virology Journal*, **12**, 219.

Roelke-Parker ME, Munson L, Packer C *et al.* (1996) Canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature*, **379**, 441–445.

Röthlisberger A, Wiener D, Schweizer M, Peterhans E, Zurbriggen A, Plattet P (2010) Two domains of the V protein of virulent canine distemper virus selectively inhibit STAT 1 and STAT 2 nuclear import. *Journal of Virology*, **84**, 6328–6343.

Sawatsky B, Wong X-X, Hinkelmann S, Cattaneo R, von Messling V (2012) Canine distemper virus epithelial cell infection is required for clinical disease but not for immunosuppression. *Journal of Virology*, **86**, 3658–3666.

Sawyer S (1989) Statistical test for detecting gene conversion. *Molecular Biology and Evolution*, **6**, 526–538.

Schaller G (1972) *The Serengeti Lion: A Study of Predator-Prey Relations*. Chicago University Press, Chicago, Illinois.

Seimon TA, Miquelle DG, Chang TY *et al.* (2014) Canine distemper virus: an emerging disease in wild endangered Amur tigers (*Panthera tigris altaica*). *MBio*, **4**, e00410–e00413.

Shi Y, Wu Y, Zhang W, Qi J, Gao GF (2014) Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nature Reviews Microbiology*, **12**, 822–831.

Sinclair ARE, Arcese P (1995) *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem*, pp. 639. University of Chicago Press, Chicago, Illinois.

Strimmer K, von Haeseler A (2009) Genetic distances and nucleotide substitution models: Theory. In: *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (eds Lemey P, Salemi M, Vandamme A-M), pp. 111–125. Cambridge University Press, Cambridge.

Swai ES, Kapaga A, Kivaria F, Tinuga D, Joshua G, Sanka P (2009) Prevalence and distribution of Peste des petits ruminants virus antibodies in various districts of Tanzania. *Veterinary Research Communications*, **33**, 927–936.

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Molecular Biology and Evolution*, **9**, 678–687.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Tatsuo H, Ono N, Yanagi Y (2001) Morbilliviruses use signaling lymphocyte activation molecules (CD150) as cellular receptors. *Journal of Virology*, **75**, 5842–5850.

Thalwitzer S, Wachter B, Robert N *et al.* (2010) Seroprevalences to viral pathogens in free-ranging and captive cheetahs (*Acinonyx jubatus*) on Namibian farmland. *Clinical and Vaccine Immunology*, **17**, 232–238.

Trebbien R, Chriel M, Struve T, Hjulsager CK, Larsen G, Larsen LE (2014) Wildlife reservoirs of canine distemper virus resulted in a major outbreak in Danish farmed mink (*Neovison vison*). *PLoS ONE*, **9**, e85598.

Tsangaras K, Siracusa MC, Nikolaidis N *et al.* (2014) Hybridization capture reveals evolution and conservation across the entire Koala retrovirus genome. *PLoS ONE*, **9**, e95633.

Turner PE, Elena SF (2000) Cost of host radiation in an RNA virus. *Genetics*, **156**, 1465–1470.

Viana M, Cleaveland S, Matthiopoulos J *et al.* (2015) Dynamics of a morbillivirus at the domestic-wildlife interface: canine distemper virus in domestic dogs and lions. *Proceedings of the National Academy of Sciences USA*, **112**, 1464–1469.

Wiens JJ, Servedio MR (1998) Phylogenetic analysis and intraspecific variation: performance of parsimony, likelihood, and distance methods. *Systematic Biology*, **47**, 228–253.

Williams ES, Thome ET, Appel MJG, Belitsky BW (1988) Canine distemper in black-footed ferrets (*Mustela nigripes*) from Wyoming. *Journal of Wildlife Diseases*, **24**, 385–398.

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS*, **13**, 555–556.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**, 431–449.

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107–1118.

The study was designed by V.N., N.O., H.H. and M.E. Field research was conducted by H.H., K.G., R.F., P.M., D.T. and M.E. Laboratory work was performed by V.N., D.T. and X.O-C. Data analyses were conducted by V.N., X.O-C., H.H., C.M., E.B. and M.E. The manuscript was written by V.N., X.O-C., H.H. and M.E. with input from all authors.

Data accessibility

Nucleotide sequence data for CDV sequences were deposited in GenBank: complete CDV genomes, complete CDV-H gene sequences, partial CDV-H gene sequences, complete CDV-P gene sequences, partial CDV-P gene sequences (accession numbers are quoted in Figs 1–3 and Table S2). The original data used to run statistical tests on the results from our functional assay are shown in Table S4.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 The phylogenetic relationship of 'Serengeti' strains: (a) to strains worldwide, based on complete genome sequences (4818 amino acids long), ML tree with branch length scaled to amino acid substitutions per site and 'Serengeti' variants are presented in bold italics; (b) among all 13 CDV-H gene sequences (Table S2), ML tree with branch length scaled to nucleotide substitutions per site. For both trees, statistical support for nodes was provided by 1000 replicates. Numbers at nodes indicate bootstrap values >70. For each strain, the origin, year and host species are quoted. Strains from the Serengeti National Park (SNP) and wider Serengeti ecosystem (SE) are indicated.

Fig. S2 Phylogenetic relationship of 'Serengeti' CDV strains based on sequences of the complete nucleocapsid gene.

Fig. S3 Phylogenetic relationship of 'Serengeti' CDV strains based on sequences of the complete matrix gene.

Fig. S4. Phylogenetic relationship of 'Serengeti' CDV strains based on sequences of the complete fusion gene.

Fig. S5 Phylogenetic relationship of 'Serengeti' CDV strains based on sequences of the complete large gene.

Fig. S6 Phylogenetic relationship between 'Serengeti' strains based on complete CDV M-F intergenic sequences.

Fig. S7 Haplotype network analysis based on nt sequences of the CDV nucleocapsid gene sequences of five 'Serengeti' strains: Red circles indicate CDV haplotypes from canid hosts, green circles indicate strains from non-canid hosts and black circles represent hypothetical nodes.

Fig. S8 Haplotype network analysis based on nt sequences of the CDV matrix gene sequences of five 'Serengeti' strains: Red circles indicate CDV haplotypes from canid hosts, green circles indicate strains from non-canid hosts and the size of circles are proportional to the number of strains presenting each haplotype.

Fig. S9 Haplotype network analysis based on nt sequences of the CDV fusion gene sequences of five 'Serengeti' strains: Red circles indicate CDV haplotypes from canid hosts, green circles indicate strains from non-canid hosts and black circles represent hypothetical nodes.

Fig. S10 Haplotype network based on the complete CDV-L gene from five host species.

Fig. S11 Haplotype network based on whole CDV genome sequences from five host species.

Fig. S12 Mixed-effects model of evolution (MEME) based on sequences encoding the SLAM binding region of the canine distemper virus haemagglutinin (CDV-H) gene investigating (a) site 519, and (b) site 549.

Fig. S13 The nectin-4 receptor binding region of the CDV-H gene. The highlighted sites are those currently considered important for virus binding.

Table S1 Canine distemper virus (CDV) strains in geographical lineages and categories of host species (domestic dogs, wild canids and non-canids).

Table S2 Canine distemper strains (CDV) from wild carnivores and a domestic dog in the Serengeti National Park (SNP) and Serengeti ecosystem (SE) between 1993 and 2011.

Table S3 Results of nested maximum likelihood codon substitution models based on CDV-H amino acid sequences

spanning the complete SLAM binding region (180 nt long), from 33 strains worldwide plus 13 'Serengeti' strains.

Table S4 Original data used to test the effect of the amino acid combination encoded by haemagglutinin proteins (CDV-H) at sites 519 and 549 in six CDV strains on the mean number of nuclei per syncytium formed in Vero cells with stable expression of SLAM receptors of the domestic dog, African lion and domestic cat.

Table S5 The manifestation of signs of infection with canine distemper virus (CDV) in carnivore hosts from the Serengeti ecosystem.