DR. JULIE KAY WECKWORTH (Orcid ID : 0000-0002-2343-4181)

DR. NICHOLAS MICHAEL FOUNTAIN-JONES (Orcid ID : 0000-0001-9248-8493)

Article type : Original Article

Cross-species transmission and evolutionary dynamics of canine distemper virus during a spillover in African lions of Serengeti National Park

Running Title: Dynamics of CDV spillover in Serengeti lions

Julie K. Weckworth[1†], Brian W. Davis[2], Edward Dubovi[3], Nicholas Fountain-Jones[4], Craig Packer[5], Sarah Cleaveland[6,7], Meggan E. Craft[4], Ernest Eblate[8], Michael Schwartz[1,9], L. Scott Mills [10,‡], Melody Roelke-Parker[11]

[1]Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, W. A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT

[2]Department of Veterinary Integrative Biosciences, Texas A&M University College of Veterinary Medicine

[3]Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York

[4]Department of Veterinary Population Medicine, University of Minnesota, St. Paul, MN

[5]Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN

[6]The Boyd Orr Centre for Population and Ecosystem Health, Institute of Biodiversity, Animal Heal and Comparative Medicine, University of Glasgow, Glasgow, Scotland

[7]Nelson Mandela African Institution for Science and Technology, Arusha, Tanzania

[8]Tanzania Wildlife Research Institute, Arusha, Tanzania

[9]United States Department of Agriculture, Forest Service, National Genomics Center for Wildlife and Fish Conservation, Rocky Mountain Research Station, Missoula, MT

[10]Fisheries, Wildlife, and Conservation Biology Program, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA.

[11]Leidos Biomedical Research, Inc., National Cancer Institute, Frederick, MD

† Corresponding author: julie.weckworth@gmail.com

‡Current affiliation, Wildlife Biology Program and Office of the Vice President for Research and Creative Scholarship, University of Montana, Missoula, MT 59812, USA.

Abstract

The outcome of pathogen spillover from a reservoir to a novel host population can range from a "dead-end" when there is no onward transmission in the recipient population, to epidemic spread and even establishment in new hosts. Understanding the evolutionary epidemiology of spillover events leading to discrete outcomes in novel hosts is key to predicting risk and can lead to a better understanding of mechanisms of emergence. Here we use a Bayesian phylodynamic approach to examine cross-species transmission and evolutionary dynamics during a canine distemper virus spillover event causing clinical disease and population decline in an African lion population (*Panthera leo*) in the Serengeti Ecological Region between 1993 and 1994. Using 21 near-complete viral genomes from four species we found that this large-scale outbreak was likely ignited by a single cross-species spillover event from a canid reservoir to non-canid hosts less than one year before disease detection and explosive spread of CDV in lions. Cross-species transmission from other non-canid species likely fueled the high prevalence of CDV across spatially structured lion prides. Multiple lines of evidence suggest that spotted hyenas (*Crocuta crocuta*) could have acted as the proximate source of CDV exposure in lions. We report thirteen nucleotide substitutions segregating CDV strains found in canids and non-canids. Our results are consistent with the hypothesis that virus evolution played a role in CDV emergence in non-canid hosts following spillover during the outbreak, and suggests that host barriers to clinical infection can limit outcomes of CDV spillover in novel host species.

Keywords

Pathogen spillover, evolutionary epidemiology, Serengeti, canine distemper virus, carnivore conservation, African lion

Introduction

Spillover occurs when a pathogen maintained in a reservoir host or community is transmitted to another host population, and has been the cause of several disease outbreaks affecting endangered species (Daszak, Cunningham, & Hyatt, 2000). The outcome of pathogen spillover in novel hosts may vary from single or independent infections without onward transmission, (i.e. an epidemiological dead-end), to epidemic spread and even establishment in the new host (Lloyd-Smith et al., 2009; Parrish et al., 2008). Most pathogen

spillover events lead to a dead-end due to a physiological barrier to clinical infection and/or onward transmission in the recipient (Parrish et al., 2008; Plowright et al., 2017). However pathogen evolution may overcome species barriers, facilitating infection and/or onward transmission (Antia, Regoes, Koella, & Bergstrom, 2003; Pepin, Lass, C Pulliam, Read, & Lloyd-Smith, 2010). Understanding the epidemiological and evolutionary history of an emerging pathogen, e.g. its origin and spread and whether and when pathogen evolution occurs, is key in predicting the outcome of spillover events and developing control measures in target populations of conservation concern.

Canine distemper virus (CDV) is a highly contagious RNA virus associated with domestic dogs. Canine distemper disease is characterized by strong immunosuppression, and lesions in the gastrointestinal, respiratory, and neurological systems leading to morbidity and death in individuals (Greene & Appel, 1990). The frequency of reports of CDV spillover to novel wildlife hosts is increasing in recent years (Panda bear (*Ailuropoda melanoleuca*, Feng et al., 2016), rhesus monkey (*Macaca mulatta*, Qiu, 2011), two-toed sloth (*Choloepus didactylus*, Sheldon, Cushing, Wilkes, Anis, & Dubovi, 2017)), including emergence in some critically endangered populations of wild felids (Iberian lynx, *Lynx pardinus,* Meli et al., 2010), Amur leopard (*Panthera pardus orientalis,* Sulikhan et al., 2018), Amur tiger (*Panthera tigris altaica,* Seimon et al., 2013), and Asiatic lion (*Panthera leo persica,* Siddiqui, 2018)). In fact twenty-two families across five orders of taxa have been documented with CDV infection, as indicated by at least seroconversion. Susceptibility of different hosts is not well characterized however. Of 217 scientific articles on CDV in non-dog hosts, only 51.8% report clinical signs (Martinez-Gutierrez & Ruiz-Saenz, 2016). Clinicopathology in alternative hosts can vary from no clinical signs given infection to nearly 100% mortality as demonstrated by the domestic cat (*Felis catus*) and the domestic ferret (*Mustela putorius furo*), respectively (Beineke, Baumgärtner, & Wohlsein, 2015). While mutation and recombination of the CDV genome have been associated with distemper disease in novel hosts (da Fontoura Budaszewski et al., 2016; McCarthy, Shaw, & Goodman, 2007; Nikolin et al., 2017), the evolutionary dynamics of strains involved in extensive CDV outbreaks following spillover in complex host communities are not well understood.

In 1994, CDV spillover into the African lion population in the Serengeti Ecological Region (SER) resulted in an estimated population decline of over 30% (Roelke-Parker et al., 1996). Clinical signs characteristic of CDV infection were observed in African lions, domestic dogs (*Canis lupus familiaris*), spotted hyenas, and bat-eared foxes (*Otocyon megalotis*) (Roelke-Parker et al., 1996). Of 19 African lions necropsied,

14 had histopathological lesions consistent with CDV, viral inclusions, and CDV nucleocapsid proteins in their tissues (Roelke-Parker et al., 1996).

Serological and epidemiological data suggested that a domestic dog population in the northwest SER seeded the CDV outbreak in lions and hyenas in 1993-1994. Domestic dog pups in villages northwest of Serengeti National Park had CDV antibodies in all of 3 years that were monitored prior to the outbreak (1992-1994), indicating prolonged CDV exposure in this high-density population. In villages east of Serengeti National Park, in the Ngorongoro District where dogs occurred at lower density, CDV exposure was documented only sporadically before the outbreak in 1991 and 1994 (Cleaveland et al., 2000). African lions however had not been exposed to CDV for 13 years prior to the outbreak, suggesting that a wild reservoir was not present in SER during that period (Packer et al., 1999; Roelke-Parker et al., 1996).

A phylogenetic analysis of partial gene sequences (389 bp of the phosphoprotein gene, 257 bp hemagglutinin gene) generated from the outbreak suggested that a single CDV strain circulated among all affected carnivores, wild and domestic, indicating no species boundaries to CDV infection (Carpenter et al., 1998). Recently however, a phylogenetic analysis using whole genome sequences (15,690 bp) revealed that the strain of CDV in circulation in canid species (domestic dogs and bat-eared fox) in fact was distinct from the CDV strain found in non-canid species (African lion and spotted hyena) during the period of the outbreak in SER (Nikolin et al., 2017). A genetic distinction between CDV sequences obtained from canids and non-canids during the same outbreak led these authors to hypothesize and test for the presence of a genetic barrier to cross-species transmission between these groups. They concluded that due to a putative species barrier, it was unlikely that domestic dogs caused the 1993-1994 outbreak in lions (Nikolin et al., 2017). However, the timing of divergence of the two strains with respect to the observed outbreak is unknown. Thus, the role of domestic dogs and the role of viral evolution in driving the epidemic spread of CDV in novel host species during the 1993-1994 outbreak is still not well understood.

In this study, we integrated epidemiological and genomic data to reconstruct cross-species transmission and evolutionary dynamics of the 1993-1994 CDV outbreak in Serengeti lions at an unprecedented resolution. Near-complete CDV genome sequences were generated from tissues of clinically infected individuals to ensure that any observed genetic diversity originated in the host and was not introduced by selective pressures in viral isolation on cultivated cells in-vitro. Data were analyzed to assess: 1) the evolutionary history of the CDV strain causing widespread mortality in Serengeti lions in 1993-1994, 2) the role domestic dogs played in the outbreak, and 3) the role of other carnivore species in driving infection in

lions. Additionally, we report all mutations segregating CDV strains associated with the putative reservoir and the spillover population during this event. Our results reveal how the 1993-1994 CDV outbreak unfolded in time and space across a complex multi-host community and provide insights into how CDV can emerge in vulnerable populations of conservation concern.

Materials and Methods

Study system and disease surveillance

Wild and domestic carnivores were sampled in the Serengeti Ecological Region (SER) between November 1992 and March 1996, spanning a period before, during, and after a large-scale, multi-host CDV outbreak (Figure 1). The SER includes Serengeti National Park and adjoining conservation areas, including the Ngorongoro District, and the Serengeti District, an unprotected area of dense human settlement along the northwest border of the national park. The area over which these samples were collected is spatially contiguous, without physical barriers to host movement. Domestic dogs rarely occur inside the national park boundaries, however direct contact between domestic dogs and wild carnivores occurs in the villages adjacent to the boundaries of the park (Cleaveland et al., 2000; Craft et al., 2017) and wild canid and non-canid species co-occur in the park.

Throughout the period of sampling, blood and tissue samples were collected routinely from wild carnivores that were captured for research, snare removal, or when encountered in a moribund or dead state. Approximately 350 lions were monitored regularly in the Serengeti lion project study area (dashed polygon, Figure 1), of these 106 were handled one or more times in conducting regular study activities. Additionally, spotted hyenas were handled occasionally for snare removal and for radio-collaring for a separate research project. Between January and February of 1994, sampling was conducted on 8 dead or dying lions. In response to the increased mortality observed in the African lion population a disease investigation was launched, during which ~5 apparently healthy lions from each of 11 prides were targeted for sampling. All other carnivores were sampled opportunistically when encountered moribund or dead during the study period. Serum and tissue samples were originally stored at -20 °C and then archived in liquid nitrogen or kept at -80°C at the National Cancer Institute, National Institutes of Health (Bethesda, Maryland). Ethical animal sampling protocols were approved by the Serengeti Wildlife Research Institute and the University of Minnesota Institutional Animal Care and use Committee (#0801A24001).

Screening clinical specimens for CDV

To screen clinical specimens for the presence of CDV RNA for sequencing in subsequent procedures, a reverse transcription quantitative PCR (RT-qPCR) assay (Path-ID™ Multiplex One-Step RTPCR Kit) was used at the Animal Health and Diagnostic Center (Ithaca, New York). Total RNA was extracted from specimens using commercial kits (MagMAX™-96 Viral RNA Isolation Kit or the QIAamp Viral RNA Mini Kit) according to manufacturer's instructions. The RT-qPCR reaction conditions were as follows: 10 min at 95°C, followed by 40 cycles of 15 sec at 95°C, and 60 sec at 60°C. Each reaction used 2-µl total RNA extract in a 25-µl volume reaction. Oligonucleotide concentrations were used at concentrations according to manufacturer's instructions.

Assessing spotted hyena CDV exposure

To assess if CDV circulated in spotted hyenas prior to the onset of CDV-related morbidity and mortality in lions, virus neutralization (VN) assays were performed on hyena sera (Table S1, Supporting Information) following standard procedures for VN assays in microtiter plates. Two-fold serum dilutions (50-µl) in duplicate were mixed with 100-300 $TCID_{50}$ of CDV (Onderstepoort strain – Baker Institute) in a 50-µl volume. Mixtures were incubated for at least 1 hour at room temperature. A 100-µl volume of indicator cells (Vero - ATCC) was added to each well and the plates were placed in a CO_2 incubator at 37°C for 4 days. Wells were scored for the presence or absence of typical CDV cytopathology. Antibody values were given as titers (reciprocal of end-point dilution). Our calculation used serum dilutions with a 50% end-point determination. In instances where the test samples exhibited toxicity to the indicator cells, the medium in the microtiter plates was changed after 12-18 hours on test. This change of medium does not affect the Ab titer of the sample, but does reduce the toxic properties of some of the test samples.

Generating genome sequence data

Two approaches were used to generate near-complete CDV genome sequences. First, we explored the feasibility of deep sequencing all nucleic acids extracted from a clinical sample (i.e. without specific PCR amplification of the CDV sequence target). A single library was prepared directly from total RNA according to manufacturer's protocol (NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina®). The remaining samples were sequenced using an amplicon-based deep sequencing protocol in which total RNA extracts were reverse transcribed and PCR amplified in 13-15 overlapping segments using published primers (Riley & Wilkes, 2015) to obtain whole CDV genomes minus the extreme 3' and 5' non-coding ends. Amplicons were generated using either a one-step (SuperScript™ III One-Step RT-PCR System with Platinum™ *Taq* DNA Polymerase) or two-step (SuperScript™ IV First-Strand Synthesis System with Q5® High-Fidelity DNA

Polymerase or QIAGEN Multiplex PCR Kit) protocol. Total RNA extracts were DNase-treated prior to RT-PCR (OPTIZYME™ DNase I (RNase-Free), Fisher BioReagents™). Further details of amplicon generation by one or two-step RT-PCR can be found in Appendix 1 and Table S2, Supporting Information. Libraries were prepared from 13-15 standardized and pooled PCR amplicons. Briefly, pools of 500 ng were sheared on a Covaris E220 Focused-ultrasonicator following the microTUBE protocol for 500 bp in 130- µl. Sonication products were then purified using Agencourt® AMPure® XP PCR purification beads at a 1:1.8 ratio and eluted in 50µl of 10mM Tris-HCl. Libraries were prepared using the TruSeq® Nano DNA Library Prep kit. Paired-end libraries were sequenced with an Illumina MiSeq producing 2 x 150 bp reads.

Paired end Illumina data generated for the single library prepared from total RNA (i.e. no CDV-specific amplification) was trimmed for Illumina adapters and barcodes using Trimmomatic 0.36 (Bolger, Lohse, & Usadel, 2014) and bases below Phred-scaled quality score (Q) 25 were removed. *De novo* genome assembly was performed using Velvet for k-mer sizes 25, 29, 31 and 35 (Zerbino & Birney, 2008), and using SPAdes 3.9 with variable k-mers (Bankevich et al., 2012). Resulting contigs were BLASTed to the non-redundant GenBank database (Morgulis et al., 2008) to filter host or exogenous DNA from *de novo* assembly. Contigs were aligned to the CDV genome most similar to remaining contigs (Canine morbillivirus virus isolate 164071, accession EU716337.1) using BWA (Li & Durbin, 2009). A consensus sequence for the resequenced haploid genome was generated using 90% identity threshold. This consensus sequence has been deposited into GenBank under accession number MT136710.

Paired-end Illumina data generated from amplicons across the remaining 20 samples was trimmed for Illumina adapters and barcodes using Trimmomatic, with bases below Q25 trimmed and bases below Q30 masked aligned to EU716337.1 using BWA. Total coverage and nucleotide counts were calculated on a per-base basis for every position from base 77 to 15,623, corresponding to a missingness of <0.1. Consensus sequences were derived from strict, 90%, nucleotide frequency. These sequences have been deposited in GenBank (accession numbers MT136705-MT136709 and MT136711-MT136725).

Genetic data analysis

Phylogenetic analysis of CDV strains involved in 1993-1994 Serengeti outbreak

The near whole genomes generated were used to construct maximum likelihood phylogenetic trees. Additional whole CDV genome sequence data from an African lion, a spotted hyena, a bat-eared fox, and a domestic dog from this outbreak are publically available (Nikolin et al., 2017). We did not include this data in our analysis because these sequences were generated from CDV which was isolated in vitro on animal cell

lines other than the original hosts. Instead, we generated sequence data from clinical specimens from the same individuals that were used for viral isolation by Nikolin et. al. Sequence alignments were performed using the MUSCLE algorithm in MEGA7 (Edgar, 2004; Kumar, Stecher, & Tamura, 2016). The Hasegawa-Kishino-Yano model (HKY, (Hasegawa, Kishino, & Yano, 1985)) model of nucleotide substitution with uniform rates among sites was the best fit to the data as determined by the Bayesian Information Criterion (BIC) in jModelTest (Posada, 2008). A phylogenetic tree was constructed using the Maximum Likelihood method in MEGA7 (Kumar et al., 2016) with statistical support derived from derived from 1000 bootstrap replicates.

Reconstructing evolutionary history of diverging CDV strains

To estimate the evolutionary rate, divergence times, and phylogenetic relationships between CDV sequences we generated in this study, we used a Bayesian Markov Chain Monte Carlo approach implemented in BEAST (Drummond, Suchard, Xie, & Rambaut, 2012) with the BEAGLE library (Ayres et al., 2012). Prior to this analysis, we explored the sequence data for evidence of recombination using RDP4 and SimPlot (Lole et al., 1999; Martin & Rybicki, 2000). Additionally, TempEst was used to qualitatively explore whether the sequence data contained sufficient temporal signal for this analysis which relies on the accumulation of mutations between heterochronously sampled sequences (Rambaut, Lam, Carvalho, & Pybus, 2016). A positive correlation was observed between genetic distance and time of sampling including all canid and non-canid sequences ($R^2 = 0.3326$, TempEst plot not shown) indicating suitability for molecular clock analysis. MLE stepping stone sampling and path sampling model selection were used to determine the best-fit clock and demographic models. Strict and relaxed lognormal clock models were evaluated in combination with four demographic clock models, including constant, exponential, skyline, and skyride. The best-fit clock model of sequence evolution was the strict clock (BF=50.99) with the coalescent constant population demographic model (Table S3, Supporting Information).

Three independent MCMC chains were run for 100M steps, of which 10M were discarded as burn-in. Tree and parameter files were logged every 10,000 steps. Traces were checked in Tracer for convergence, i.e. ESS>200 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018). TreeAnnotator from the BEAST package was used to combine and annotate trees and the resulting maximum clade credibility tree was visualized in FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Ancestral Host State Reconstruction

We estimated the host state at all branches and internal nodes across the time-measured MCC tree using a discrete phylogenetic diffusion model with 4 host states (African lion, spotted hyena, bat-eared fox,

and domestic dog) in BEAST version 1.8.4 (Lemey, Rambaut, Drummond, Suchard, & Ali, 2009). A non-reversible continuous-time Markov chain was specified to allow the estimation of asymmetric host transition rates and the Bayesian Stochastic Search Variable procedure (BSSVS) was enabled to constrain the number of rates allowed for explaining the history of cross-species spread.

Haplotype Network Analysis

A haplotype network was reconstructed using the TCS algorithm in PopARt to visualize the relationships among genotypes in our sample and count the number of mutational steps between them (Clement, M., Snell, Q., Walker, P., Posada, D., & Crandall, 2002; Leigh & Bryant, 2015).

Results

CDV antigen detection

RT-qPCR screening of 210 samples from 13 species detected CDV antigen in 43 individuals (Table 1). Of the individuals for which we have health observations, all had clinical signs and/or histopathology indicative of CDV infection except for two spotted hyena adults and one African lion whom appeared healthy but may have had enlarged popliteal lymph nodes. One adult pregnant hyena killed by lions was in good physical condition at time of death. Her intrauterine offspring had histopathology consistent with CDV infection.

Timing of spotted hyena exposure to CDV

All of twelve apparently healthy adult and sub-adult hyenas sampled between December 29, 1992 and September 23, 1993 were CDV seronegative, i.e. no antibodies were detected at the minimum readable dilution (Table 2, Table S1, Supporting information). Three hyena cubs sampled post-mortem between December 20, 1993 and July 3, 1994 were qPCR positive for CDV RNA and had histological signs consistent with CDV pathology (Linda Munson, unpublished data). These cubs had detectable CDV titer values, distinguishing this sample set from individuals sampled pre-outbreak, however two of the three titers were too low to reliably distinguish from cross-reactivity with other pathogens (with titer values 8 and 12). The apparent seronegative or low status of clinically sick hyenas suggests that these individuals did not mount an adaptive immune response to CDV antigen. Of four hyenas sampled between January 6, 1995 and June 20, 1995, all were seropositive with titers ranging from 96 to 256. Four serum neutralization tests were inconclusive due to toxicity and not included in these results.

Phylogenetic relationships of CDV strains

Near-complete CDV genomes (11,926 bp-15,547 bp in length, of 15,690 bp) were sequenced from 21 individuals representing four carnivore hosts infected during the 1994 Serengeti outbreak. These included African lions, spotted hyenas, bat-eared fox, and domestic dogs (Table S4, Supporting Information). Seven of the 10 African lions sampled were part of a long-term ecological study on Serengeti lions whose condition and location (and those of their pride mates) were ascertained regularly (Packer et al., 2005). Three lions were unknown to researchers and have presumed origins outside of the study area.

Phylogenetic analysis by the Maximum Likelihood method of 21 CDV sequences from multiple hosts sampled during the 1993-1994 outbreak in SER, indicates that two unique but closely related CDV strains were in circulation (similarity 99.82%) (Fig. 2). The strains are separated by taxa such that one monophyletic clade contains only species from Canidae (domestic dogs and bat-eared foxes in our sample; Bootstrap value, BV=0.95), and the other contains only species that are not from Canidae (African lion and spotted hyena, BV=0.93). Hereafter these strains are referred to as the "canid" and "non-canid" strains following the distinction made in previous studies on CDV host tropism (McCarthy et al., 2007; Nikolin et al., 2017).

Of 18 lion prides under surveillance, 17 were exposed to CDV, the majority of which suffered associated morbidity and mortality (Craft, Hawthorne, Packer, & Dobson, 2008). CDV sequences generated from lions in 4 prides fall into 3 distinct clusters, or transmission groups, in the phylogenetic tree (Fig. 2). Hyena sequences are interspersed among all transmission groups and are basal in 2 of 3 groups that also contain lions, supporting regular cross-species transmission of the non-canid strain among hyenas and lions.

The non-canid clade is characterized by a large polytomy (with the exclusion of one sequence, PLE-658) consistent with a rapid radiation from a common ancestor of these viral genotypes (Fig. 2). PLE-658 was an unknown lioness found dead within the territory of a well-known study pride in May 1994, after most mortality in the lion study population had occurred. That she was unknown to researchers suggests that she was from outside the study area, possibly a nomad or displaced by social upheaval in the wake of widespread mortality. All other non-canid sequences clustered together regardless of time or location of sampling.

Evolutionary history of diverging CDV strains

Our time-calibrated maximum clade credibility tree dated the most recent common ancestor (MRCA) of the canid and non-canid associated strains as approximately 1.76 years before the last specimen in the analysis was collected (i.e., February 1993) (95% HPD: 1.12-2.38 y) (Fig. 3). This timing is roughly 10.2 months before the first observed case in the non-canid strain, a moribund hyena cub found on December 20, 1993 with extensive hemorrhaging into the small intestine (CCR-6). Lesions in the nervous, lymphoid, and

respiratory systems of the cub were consistent with CDV infection as determined by immunohistochemistry (Linda Munson, unpublished data). The time to the most recent common ancestor (tMRCA) between PLE-658 and all other non-canid sequences was approximately 1.46 years before the last specimen (95% HPD: 1.13-1.83 y), or around 6.6 months before the first observed non-canid case. The approximate timing of the radiation of all non-canid viral genotypes excluding PLE-658 was around 1.23 years before the last specimen (95% HPD: 1.02-1.47 y) or 3.8 months before the first non-canid case observed. The median evolutionary rate of the CDV genome during this outbreak was 6.525×10^{-4} substitutions per site per year, i.e. 10.2 substitutions per year.

Cross-species transmission dynamics

Ancestral host-state reconstruction predicted that cross-species transmission from hyenas to lions occurred regularly, with hyenas being the highest probable predicted host state for all transmission groups containing lions (PP=0.67-0.79, Fig.3).The estimated host state transition rate from spotted hyena to African lion was well supported (Bayes Factor, BF=21.77, Posterior probability, PP=0.91; Table 3). Moderate evidence supports the estimated host transition rates between domestic dog and bat-eared fox (BFs >6, PPs > 0.73), and from African lion to spotted hyena (BF>6, PP = 0.74; Table 3), while transitions between any canid/non-canid pair have low support (BFs <2.11, PP <0.48). Intra-specific transmission was predicted to have occurred in both hyenas and lions (PP≥0.9). Similarly, intraspecific transmission was predicted in both domestic dogs and bat-eared foxes with high probability (P=0.99). Host state at internal nodes in the canid clade could not accurately be predicted from the data (PP=0.51 and 0.5). Similarly, host state at the root was unresolved. Spotted hyena was predicted to be the host origin of the CDV strain causing this outbreak, however the majority of sampled trees did not support this hypothesis due likely to the small sample size. In particular, PPs for the host set were 0.33, 0.25, 0.17, and 0.25 for spotted hyena, bat-eared fox, African lion, and domestic dog, respectively.

Mutational steps between sampled CDV sequences

A haplotype network analysis supported the genetic distinction between strains associated with canid and non-canid species and reveals 13 nucleotide substitutions separating the two (Fig. 4a.). Of these, 7 substitutions are nonsynonymous and occur in the P/V, F, and H regions of the genome (Fig. 4b.), genes whose functions facilitate efficient replication, host immune evasion, and host cell receptor binding and membrane fusion. The position and character of amino acid changes in the genome are listed in Table 4. PLE-658 shares 1 synonymous and 2 nonsynonymous nucleotide substitutions with the canid strain in the L gene

region (asterisks Fig. 4b.), a protein that catalyzes the replication of viral genomic RNA. The residues on the L gene that are affected by these substitutions are conserved across all other published CDV sequences. Despite sharing nucleotide substitutions with both the non-canid strain and the canid strain, PLE-658 does not show evidence of recombination (data not shown).

Discussion

In this study, we found evidence that a large-scale outbreak of canine distemper in Serengeti lions was precipitated by a recent spillover from a canid reservoir into the non-canid community, infecting at least African lions and spotted hyenas. Our data suggests that a CDV strain associated with non-canid hosts diverged from a canid-associated strain less than 11 months before the detection of clinical signs and extensive spread of CDV in non-canid hosts. Thirteen single nucleotide polymorphisms across the whole genome, coding seven amino acid differences in biologically relevant regions of the CDV genome, separate these strains. Following this strain divergence, we did not observe transmission of strains between canid and non-canid hosts despite apparent co-circulation. These results suggest the presence of a host barrier to CDV infection between canids (e.g. domestic dogs and bat-eared foxes) and non-canids (i.e. lions and hyenas). Genetic sequences within each group, canid and non-canid, were highly similar suggesting no or a low barrier to CDV infection between more closely related host species in our sample, i.e. between domestic dog and bat-eared fox, and between African lion and spotted hyena. Ancestral host-state reconstruction suggests that hyenas may have been an important source of CDV infecting lions and affirms that onward transmission in the lion population may have also occurred.

CDV spillover from a canid reservoir

Our data support the hypothesis that CDV spilled over from a canid reservoir to precipitate the widespread distemper outbreak in non-canids in the Serengeti National Park in 1993/1994. Our Bayesian analysis of time-stamped genomic data revealed that the two strains of CDV circulating during the outbreak diverged from a common ancestor just months before the detection of clinical signs in Serengeti lions and hyenas. Multiple lines of evidence suggest that this common ancestor was a canid, i.e. the direction of transmission at the root of this outbreak was from canid to non-canid hosts.

If a non-canid host had maintained CDV in the Serengeti then we would expect to see evidence of infection in hyenas and lions prior to the estimated time of most recent common ancestor, February 1993. In fact, all of the hyenas sampled before the onset of observed morbidity and mortality beginning in December 1993 were seronegative, hyenas exhibiting clinical signs during the outbreak had low CDV-neutralizing titers

(a common finding in individuals that are not mounting a successful response to infection), and all of the hyenas sampled after the outbreak had high titers suggesting recovery. Similarly the African lion population had not been exposed to CDV for thirteen years prior to December 1993 (Packer et al., 1999), when the first seropositive lion was detected. Although the spotted hyena serum dataset is small and does not represent a random sample, taken together the pattern suggests that CDV was not in circulation in the wild Serengeti carnivore community prior to onset of clinical signs in these hosts.

By contrast, evidence of canine distemper virus infection and/or previous exposure was reported in domestic dog populations northwest and east of Serengeti National Park from at least 1992-1994 (Cleaveland et al., 2000) and in the Masai Mara National Reserve in Kenya north of Serengeti from at least 1989-1991 (Alexander & Appel, 1994). Anecdotally, jackal and bat-eared foxes in the MMNR declined during this period and African wild dogs were locally extirpated, suggesting that wild canid species were clinically infected with CDV there (Alexander & Appel, 1994).

A single or limited spillover, likely explains CDV infection in the non-canid clade, as indicated by phylogenetic monophyly. Multiple incursions of CDV from a canid reservoir were unlikely to have fueled the outbreak in non-canid species given the high genetic similarity of all non-canid species sequences to the exclusion of the canid sequences. That all sequences from non-canid hosts can be traced to a single ancestor regardless of time or location of sampling, despite more than four years of documented CDV circulation in domestic dogs in the ecosystem (Alexander & Appel, 1994; Cleaveland et al., 2000), suggests that host barriers to infection may have limited CDV spillover in non-canid hosts prior to the outbreak.

Epidemiological evidence alone would suggest that CDV “spilled back” into dogs on the east side of the SER (in the Ngorogngoro Conservation Area) from wild non-canids, as suggested earlier (Cleaveland et al., 2007). However, molecular dating in this analysis shows that the canid and non-canid strains are genetically independent and share an ancestor nearly 1.5 years prior to the time at which the first canid was sampled. Thus, our data suggests that CDV spilled over to this low-density domestic dog population from the same source that infected non-canids.

Our results shed new light on the origins of the 1993-1994 outbreak. The phylogenetic discovery of Nikolin et al. (2017) that two strains circulated in canid and non-canid species in the Serengeti in 1993-1994, challenged the previous model of CDV emergence in Serengeti lions that had specified a domestic dog origin (Cleaveland et al., 2000). Specifically, their study concluded that genetic differences separating the two strains signify a genetic barrier to cross-species transmission between canids and non-canids, and therefore

CDV in non-canids must have come from a non-canid origin. Importantly, this conclusion implies the long-term maintenance of a non-canid adapted strain and suggests that the reservoir of CDV which threatens Serengeti lions and other non-canid carnivores, is the non-canid community itself. Our analysis of a more comprehensive, whole CDV genome dataset unbiased by the effects of viral isolation on sequence variation, upholds the finding that two strains circulated in canids and non-canids in 1993/1994. However, the integration of epidemiological and genomic data in our analysis offered the unique discovery that the CDV strain causing mortality in non-canid species had emerged in a few months, most likely from the domestic dog population. This is not surprising given that the strains share 99.82% similarity. Nevertheless, this result is illuminating because it reveals the process of emergence of a pathogenic CDV genotype in spotted hyenas and African lions from a canid origin, thus refocusing CDV surveillance and management on canid reservoirs.

Transmission Dynamics within strains

Our CDV phylogeny and haplotype network topologies reveal that hyenas may have been an important source of CDV in Serengeti lions at the population scale during the 1994 outbreak. If lion-to-lion transmission alone were responsible for CDV spread, then we would expect all lions to cluster together in these analyses, irrespective of pride affiliation. However, sequences from 4 sampled prides fall into 3 distinct clusters signifying that cross-species transmission is likely responsible for exposure in most prides sampled. Furthermore, hyenas are interspersed in every transmission chain that reached lions and occupy a basal position in two of these, supporting the role of hyenas as a CDV source to lions. Ancestral host state reconstruction supports this interpretation, implicating hyenas as the most probable source of CDV (of species that were sampled) in all transmission groups of the non-canid CDV strain. It is not surprising that the direction of transmission is from hyenas to lions, as most lethal encounters between hyenas and lions result in hyena mortality.

The sequence of clinical observations in wild carnivores suggests that CDV circulated in hyenas before causing disease in lions. Histopathology supports a CDV diagnosis in a lion-killed hyena as early as April 1993 (unpublished data, Munson), although CDV qPCR of this individual in this study was negative. Haas et al. (1996) report clinical signs in hyena cubs as early as November 1993. The first African lion mortalities related to the outbreak were not reported until January 1994 and clinical signs would not be observed in a living lion until February 1994 (Roelke-Parker et al., 1996). If only a single spillover can be inferred from the data, and clinical signs in hyenas were observed months ahead of clinical signs in lions, it follows that hyenas could be

the source of CDV for lions. We did not find evidence of CDV circulation in hyenas prior to the outbreak from serological data, however the sample size was small and limited the power to detect low-level circulation.

While spotted hyenas may have played a key role in driving the spread of the non-canid CDV strain in Serengeti lions at the population-scale, we found evidence for lion-to-lion transmission at the local scale. Specifically, 5 lion sequences representing two neighboring lion prides and one individual lion of unknown origin fall in a single cluster suggesting that lion-to-lion transmission occurred. This interpretation is supported by ancestral host state reconstruction that predicts lion-to-lion transmission with high probability in this cluster (PP=0.97-0.99). Thus, lion-to-lion transmission may partially explain exposure patterns within and between lion prides.

It is important to note that assigning transmission roles in sparsely sampled communities during a multi-host outbreak is complicated when unsampled susceptible host species are unaccounted for in the data. If a host species that was not sampled, e.g. mongoose, was responsible for CDV spread to both hyenas and lions, then our data might wrongly assign transmission roles to one of the sampled species. Although we cannot rule out the possibility that an unsampled host drove the observed patterns in our data, we believe that our interpretation that hyenas drove the spread of CDV in the lion population is reasonable given the sequence of case reports and what is known about hyena movement ecology.

Hyena movement ecology is consistent with the inference that hyenas spread CDV at the population level in lions. CDV emerged in 6 lion prides almost simultaneously despite that these prides were mostly spatially discontinuous and separated by up to 40 km (Craft, Hawthorne, Packer, & Dobson, 2008). The four most isolated of these prides were located on the short grass plains where, at the time of the outbreak, hundreds of thousands of wildebeest were congregating to calve. Hyenas commute long distances (40-80km) from den sites in the center of Serengeti National Park to the short grass plains in the southeast to take advantage of the abundant prey (Hofer & East, 1993). Spotted hyena commuters travel singly and in any one day several individuals from a single clan, whose base is in the center of Serengeti National Park, may be dispersed over hundreds of square kilometers, having traversed multiple lion pride and hyena clan territories to get there (Hofer & East, 1993). No other Serengeti carnivore to our knowledge makes such extensive, long-range movements. A previous study reports that clinical infection was limited to hyena cubs in 1994, which would argue against hyena involvement in long-distance spread (Nikolin et al., 2017). In this study, we observed CDV infection in the intrauterine offspring of an infected, pregnant adult spotted hyena, suggesting that adult hyenas were competent hosts and capable of spreading CDV.

CDV evolution during the 1994 outbreak

Viral evolution can surmount host barriers determining cross-species transmission outcome and/or clinical severity in hosts. In this study, we identified 13 nucleotide substitutions coding 7 amino acid residue substitutions in biologically relevant regions of the CDV genome, separating the canid and non-canid strains. The divergence of CDV genotypes associated with distantly related host species, driven by accumulation of mutations in a short period of time, suggests that CDV evolution may have lowered the host barrier leading to clinical infection in lions and hyenas.

We identified nonsynonymous nucleotide substitutions between the canid and non-canid strains in the Phosphoprotein (P)/V protein (V), Fusion (F), & Hemagglutinin (H) open reading frames. Testing the functional relevance of these substitutions is outside of the scope of our study. However, a well-annotated CDV genome exists because this virus is used as a model to understand Measles virus pathology. Thus, the position of substitutions in the genome reported here between canid and non-canid associated strains may indicate potential mechanisms for lowering the host barrier to clinical infection.

The CDV P gene encodes the phosphoprotein, part of the RNA polymerase complex, and a nonstructural protein V, which is critical for host immune evasion. V-deficient viruses replicate and spread in the lymphatic system of a host, but are unable to interfere with the immune response of the host and are cleared before epithelial cell infection – a condition that must be met for progression to disease and onward transmission (Sawatsky, Wong, Hinkelmann, Cattaneo, & von Messling, 2012; von Messling, Svitek, & Cattaneo, 2006). We found the unique G134S substitution in the non-canid strain consistent with Nikolin (2017) and an additional substitution K280E in the P/V region. To our knowledge, a mutation at site 280 has not been observed in any other published CDV genome. This mutation occurs in the highly-conserved zinc-binding domain of the V protein which is a critical component of Paramyxovirus strategy for evading host innate immunity (Röthlisberger et al., 2010; Svitek et al., 2014; von Messling et al., 2006).

The H and F genes code for transmembrane proteins that interact with susceptible host cell receptors and mediate viral and host cell membrane fusion, respectively. As the first point of CDV infection in a host, these proteins are considered critical determinants of host cell tropism. We found four amino acid differences between the canid and non-canid strains at amino acid positions 160, 178, 519, and 549 on the H gene and 1 difference at position 315 on the F gene (Table 4).

The specific mutation Y549H on H has been shown to be under selection and associated with host tropism (McCarthy 2007), though the exact relationship between variation at this site and clinical infection in

various host species is not clear (Gilbert et al., 2015; Terio & Craft, 2013). 549Y is associated with CDV strains from canid species and 549H is more commonly found in non-canid species (McCarthy et al., 2007; Nikolin, Wibbelt, Michler, Wolf, & East, 2012). Nikolin et al. (2012, 2017) demonstrate that 549H, especially when in combination with 519I, improves the ability of CDV to enter felid immune cells in-vitro. This finding suggests that the non-canid associated strain had mutations that optimized its ability to enter lion immune cells, i.e. these mutations may have made onward transmission between non-canid hosts more efficient.

We found two additional mutations on the large (L) protein that occurred in all but one non-canid sequence, PLE-658. Mutations that separate PLE-658 and all other non-canid sequences may be particularly relevant to virus phenotype because the two chains of transmissions had dramatically different fates. The chain leading to PLE-658 apparently died out, whereas the other chain was ostensibly explosive. The L gene is involved in efficient replication of the viral genome, which is critical for onward transmission. A study of the closely related Measles virus demonstrated that mutations on the L protein can affect host tropism (Tahara, Takeda, & Yanagi, 2005).

Our data do not address if the observed mutations in the non-canid strain were necessary for clinical infection in lions and hyenas. However, a CDV sequence recently published from a hyena in South Africa (Loots, Du Plessis, Dalton, Mitchell, & Venter, 2017) bears 4 of the 7 amino acid substitutions that were observed in non-canids in SER in 1993-1994 (V- G134S, H-D178G, H- R519I and Y549H). This is notable because 1) this is the first case of CDV in a hyena ever reported outside of the 1993-1994 outbreak in SNP, and 2) the hyena sequence from South Africa belongs to a distinct and different lineage (Africa-1) than the Serengeti sequences (Africa-2) (Loots et al., 2017). That the South African and East African sequences share rare mutations despite having very different genetic backgrounds overall, supports that these mutations have functional significance in spotted hyena, and possibly African lion, clinical infection.

Conclusion/ Management implications

Though host species differences may limit CDV clinical infection and thus spillover outcome, our data supports the hypothesis that one or few evolutionary steps overcame these barriers in a short time frame in Serengeti lions in 1993-1994. Nevertheless, the occurrence of spillover in lions causing distemper disease and mortality in this system is apparently rare. This improved understanding of CDV spillover in African lions allows managers to weigh the risks of CDV exposure. Specifically, CDV management in the Serengeti has been focused on preventing lion exposure to CDV (Viana et al., 2015). However, long-term serological data from this population and from other wild felid populations suggest that regular exposure to CDV does not lead to

regular population impacts (Munson et al., 2008; Viana et al., 2015), and can actually increase herd-immunity. Thus, natural exposure to CDV may buffer Serengeti lion and hyena populations from extirpation during the rare occurrence of a pathogenic CDV spillover event in non-canids.

Our results highlight the role of spotted hyenas as a potential source of CDV infection in African lions and driver of long-distance spread of a pathogenic CDV strain during this outbreak. The appearance of CDV clinical signs in spotted hyenas months in advance of clinical signs in lions highlights the potential for hyenas as a sentinel species.

We identified four novel specific mutations separating a pathogenic non-canid CDV strain from a canid-associated strain in this outbreak, in addition to three that have been previously discussed (Nikolin et al., 2017). Although it is unclear if any or some of these mutations predict clinical infection in lions and hyenas, this study identifies candidate sites or regions to motivate future experiments. Until a better understanding of the functional significance can be reached, surveillance for these mutations in circulating CDV strains may inform CDV risk assessment in lion and hyena populations of conservation concern.

Acknowledgements

We thank Linda Munson for critical contributions made in documenting and diagnosing CDV as the cause of the 1993-1994 epidemic. The Messerli Foundation in Zurich, Switzerland provided funding for sample collection and field research. We are grateful to Tamara Max, Amy Glaser, Melissa Laverack, Nancy Cavender Zylich, and Christian Leutenegger for providing technical assistance and Vanessa Ezenwa, Gordon Luikart, Angie Luis, Mark Hebblewhite, Hugh Robinson, Byron Weckworth, and Gretchen Grimes for helpful comments on study design and/or earlier drafts of this manuscript. Funding for this study was provided by the Morris Animal Foundation, grant ID D14ZO-013, “Understanding spillover of a deadly virus in a well-studied African lion population”. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant #DGE-0809127 and the National Science Foundation grant #DGE-0504628.

References

Alexander, K. A., & Appel, M. J. G. (1994). African wild dogs (Lycaon pictus) endangered by a canine distemper epizootic among domestic dogs near the Masai Mara National Reserve, Kenya. *Journal of Wildlife Diseases*, *30*(4), 481–485. https://doi.org/10.7589/0090-3558-30.4.481

Antia, R., Regoes, R. R., Koella, J. C., & Bergstrom, C. T. (2003). The role of evolution in the emergence of infectious diseases. *Nature*, *426*(December), 8–11. https://doi.org/10.1038/nature02177.1.

Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., … Suchard, M. A. (2012). BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. *Systematic Biology*, *61*(1), 170–173. https://doi.org/10.1093/sysbio/syr100

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *19*(5), 455–77. https://doi.org/10.1089/cmb.2012.0021

Beineke, A., Baumgärtner, W., & Wohlsein, P. (2015). Cross-species transmission of canine distemper virus—an update. *One Health*, *1*, 49–59. https://doi.org/10.1016/J.ONEHLT.2015.09.002

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Carpenter, M. A., Appel, M. J. G., Roelke-Parker, M. E., Munson, L., Hofer, H., East, M., & O 'brien, S. J. (1998). Genetic characterization of canine distemper virus in Serengeti carnivores. *Veterinary Immunology and Immunopathology*, *65*, 259–266.

Cleaveland, S., Appel, M. G. J., Chalmers, W. S. K., Chillingworth, C., Kaare, M., & Dye, C. (2000). Serological and demographic evidence for domestic dogs as a source of canine distemper virus infection for Serengeti wildlife. *Veterinary Microbiology*, *72*, 217–227.

Cleaveland, S., Mlengeya, T., Kaare, M., Haydon, D., Lembo, T., Laurenson, M. K., & Packer, C. (2007). The conservation relevance of epidemiological research into carnivore viral diseases in the serengeti. *Conservation Biology*, *21*(3), 612–622. https://doi.org/10.1111/j.1523-1739.2007.00701.x

Clement, M., Snell, Q., Walker, P., Posada, D., & Crandall, K. (2002). TCS: Estimating gene genealogies. In *Parallel and Distributed Processing Symposium, International Proceedings, 2* (p. 184).

Craft, M. E., Hawthorne, P. L., Packer, C., & Dobson, A. P. (2008). Dynamics of a multihost pathogen in a carnivore community. *Journal of Animal Ecology*. https://doi.org/10.1111/j.1365-2656.2008.01410.x

Craft, M. E., Vial, F., Miguel, E., Cleaveland, S., Ferdinands, A., & Packer, C. (2017). Interactions between domestic and wild carnivores around the greater Serengeti ecosystem. *Animal Conservation*, *20*(2), 193–204. https://doi.org/10.1111/acv.12305

da Fontoura Budaszewski, R., Streck, A. F., Nunes Weber, M., Maboni Siqueira, F., Muniz Guedes, R. L., & Wageck Canal, C. (2016). Influence of vaccine strains on the evolution of canine distemper virus. *Infection, Genetics and Evolution*, *41*, 262–269. https://doi.org/10.1016/J.MEEGID.2016.04.014

Daszak, P., Cunningham, A. A., & Hyatt, A. D. (2000). Emerging infectious diseases of wildlife--threats to biodiversity and human health. *Science (New York, N.Y.)*, *287*(5452), 443–9. https://doi.org/10.1126/SCIENCE.287.5452.443

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian hylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–7. https://doi.org/10.1093/nar/gkh340

Feng, N., Yu, Y., Wang, T., Wilker, P., Wang, J., Li, Y., … Xia, X. (2016). Fatal canine distemper virus infection of giant pandas in China. *Scientific Reports*, *6*, 27518. https://doi.org/10.1038/srep27518

Gilbert, M., Soutyrina, S. V., Seryodkin, I. V., Sulikhan, N., Uphyrkina, O. V., Goncharuk, M., … Miquelle, D. G. (2015). Canine distemper virus as a threat to wild tigers in Russia and across their range. *Integrative Zoology*, *10*(4), 329–343. https://doi.org/10.1111/1749-4877.12137

Greene, C. E., & Appel, M. J. (1990). Canine Distemper. In C. E. Greene (Ed.), *Infectious Diseases of the Dog and Cat* (pp. 226–241).

Haas, L., Hofer, H., East, M., Wohlsein ', P., Liess, B., & Barrett, T. (1996). Canine distemper virus infection spotted hyaenas in Serengeti. *Veterinary Microbiology*, *49*, 147–152.

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174. https://doi.org/10.1007/BF02101694

Hofer, H., & East, M. L. (1993). The commuting system of Serengeti spotted hyaenas:how a predator copes with migratory prey. II. Intrusion pressure and commuters' space use. *Anim. Behav.* , *46*, 559–574.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Molecular Biology and Evolution*, *33*(7), 1870–4.

Leigh, J. W., & Bryant, D. (2015). popart : full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Lemey, P., Rambaut, A., Drummond, A. J., Suchard, M. A., & Ali, Y. (2009). Bayesian Phylogeography Finds Its Roots. *PLoS Computational Biology*, *5*(9), e1000520. https://doi.org/10.1371/journal.pcbi.1000520

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lloyd-Smith, J. O., George, D., Pepin, K. M., Pitzer, V. E., Pulliam, J. R. C., Dobson, A. P., … Grenfell, B. T. (2009).

Epidemie dynamics at the human-animal interface. *Science*, *326*(5958), 1362–1367. https://doi.org/10.1126/science.1177345

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., … Ray, S. C. (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*(1), 152–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9847317

Loots, A. K., Du Plessis, M., Dalton, D. L., Mitchell, E., & Venter, E. H. (2017). Genome Sequences of Three Vaccine Strains and Two Wild-Type Canine Distemper Virus Strains from a Recent Disease Outbreak in South Africa. *Genome Announcements*, *5*(27). https://doi.org/10.1128/genomeA.00603-17

Martin, D., & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *BIOINFORMATICS APPLICATIONS NOTE*, *16*(6), 562–563. Retrieved from http://

Martinez-Gutierrez, M., & Ruiz-Saenz, J. (2016). Diversity of susceptible hosts in canine distemper virus infection: A systematic review and data synthesis. *BMC Veterinary Research*, *12*(1), 1–11. https://doi.org/10.1186/s12917-016-0702-z

McCarthy, A. J., Shaw, M.-A., & Goodman, S. J. (2007). Pathogen evolution and disease emergence in carnivores. *Proceedings. Biological Sciences / The Royal Society*, *274*(1629), 3165–74. https://doi.org/10.1098/rspb.2007.0884

Meli, M., Simmler, P., Cattori, V. M., Fernando Vargas, Astrid Palomares, Fransisco Lopez-Bao, … Lutz, H. (2010). Importance of canine distemper virus (CDV) infection in free-ranging Iberian lynxes (Lynx pardinus). *Veterinary Microbiology*, *146*(1–2), 132–137. https://doi.org/10.1016/J.VETMIC.2010.04.024

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., Schäffer, A. A., … An, P. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics*, *24*(16), 1757–1764. https://doi.org/10.1093/bioinformatics/btn322

Munson, L., Terio, K. A., Kock, R., Mlengeya, T., Roelke, M. E., Dubovi, E., … Packer, C. (2008). Climate Extremes Promote Fatal Co-Infections during Canine Distemper Epidemics in African Lions. *PLoS ONE*, *3*(6), e2545. https://doi.org/10.1371/journal.pone.0002545

Nikolin, V. M., Olarte-Castillo, X. A., Osterrieder, N., Hofer, H., Dubovi, E., Mazzoni, C. J., … East, M. L. (2017). Canine distemper virus in the Serengeti ecosystem: molecular adaptation to different carnivore species. *Molecular Ecology*. https://doi.org/10.1111/mec.13902

Nikolin, V. M., Wibbelt, G., Michler, F.-U. F., Wolf, P., & East, M. L. (2012). Susceptibility of carnivore hosts to

strains of canine distemper virus from distinct genetic lineages. *Veterinary Microbiology*, *156*(1–2), 45–53. https://doi.org/10.1016/j.vetmic.2011.10.009

Packer, C., Altizer, S., Appel, M. J. G., Brown, E., Martenson, J., O'Brien, S., … Lutz, H. (1999). Viruses of the Serengeti: Patterns of infection and mortality in African lions. *Journal of Animal Ecology*, *68*(6), 1161–1178. https://doi.org/10.1046/j.1365-2656.1999.00360.x

Packer, C., Hilborn, R., Mosser, A., Kissui, B., Borner, M., Wilmshurst, J., … Sinclair, A. R. E. (2005). timing EcoLogical Change , Group Territoriality , and Population, *307*(5708), 390–393.

Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E., Burke, D. S., Calisher, C. H., … Daszak, P. (2008). Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases. *Microbiology and Molecular Biology Reviews*, *72*(3), 457–470. https://doi.org/10.1128/MMBR.00004-08

Pepin, K. M., Lass, S., C Pulliam, J. R., Read, A. F., & Lloyd-Smith, J. O. (2010). Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nature Publishing Group*, *8*. https://doi.org/10.1038/nrmicro2440

Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. https://doi.org/10.1038/nrmicro.2017.45

Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, *25*(7), 1253–1256. https://doi.org/10.1093/molbev/msn083

Qiu, W. (2011). Canine Distemper Outbreak in Rhesus Monkeys, China. *Emerging Infectious Diseases*. https://doi.org/10.3201/eid1708.101153

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rambaut, A., Lam, T. T., Carvalho, L. M., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*, *2*(1). https://doi.org/10.1093/ve/vew007

Riley, M. C., & Wilkes, R. P. (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. https://doi.org/10.1186/s12985-015-0445-7

Roelke-Parker, M. E., Munson, L., Packer, C., Kockil, R., CIeaveland, S., Carpenter, M., … G AppeII, M. J. (1996).

A 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 STAT1 and STAT2 nuclear import. *Journal of Virology*, *84*(13), 6328–43. https://doi.org/10.1128/JVI.01878-09

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*(7), 3658–66. https://doi.org/10.1128/JVI.06414-11

Seimon, T. A., Miquelle, D. G., Chang, T. Y., Newton, A. L., Korotkova, I., Ivanchuk, G., … McAloose, D. (2013). Canine distemper virus: an emerging disease in wild endangered Amur tigers (Panthera tigris altaica). *mBio*, *4*(4), e00410-13. https://doi.org/10.1128/mBio.00410-13

Sheldon, J. D., Cushing, A. C., Wilkes, R. P., Anis, E., & Dubovi, E. J. (2017). Serologic response to canine distemper vaccination in captive Linnaeus's two-toed sloths (*Choloepus didactylus*) after a fatal canine distemper virus outbreak. *Journal of Zoo and Wildlife Medicine*, *48*(4), 1250–1253. https://doi.org/10.1638/1042-7260-48.4.1250

Siddiqui, Z. (2018). More Asiatic lions in India test positive for virus after 23 deaths - Reuters. Retrieved June 25, 2019, from https://www.reuters.com/article/us-india-wildlife-lions/more-asiatic-lions-in-india-test-positive-for-virus-after-23-deaths-idUSKCN1MK21M

Sulikhan, N. S., Gilbert, M., Yu Blidchenko, E., Naidenko, S. V, Ivanchuk, G. V, Yu Gorpenchenko, T., … Seimon, T. A. (2018). Canine Distemper Virus in a Wild Far Eastern Leopard Panthera pardus orientalis. *Journal of Wildlife Diseases Wildlife Disease Association*, *54*(1), 0–0. https://doi.org/10.7589/2017-03-065

Svitek, N., Gerhauser, I., Goncalves, C., Grabski, E., Döring, M., Kalinke, U., … von Messling, V. (2014). Morbillivirus control of the interferon response: relevance of STAT2 and mda5 but not STAT1 for canine distemper virus virulence in ferrets. *Journal of Virology*, *88*(5), 2941–50. https://doi.org/10.1128/JVI.03076-13

Tahara, M., Takeda, M., & Yanagi, Y. (2005). Contributions of matrix and large protein genes of the measles virus edmonston strain to growth in cultured cells as revealed by recombinant viruses. *Journal of Virology*, *79*(24), 15218–25. https://doi.org/10.1128/JVI.79.24.15218-15225.2005

Terio, K. A., & Craft, M. E. (2013). Canine Distemper Virus (CDV) in Another Big Cat: Should CDV Be Renamed Carnivore Distemper Virus? *mBio*, *4*(5), e00702-13-e00702-13. https://doi.org/10.1128/mBio.00702-13

Viana, M., Cleaveland, S., Matthiopoulos, J., Halliday, J., Packer, C., Craft, M. E., … Orr, B. (2015). Dynamics of

a morbillivirus at the domestic-wildlife interface: Canine Dis- temper Virus in domestic dogs and lions. *PNAS*, *112*(5), 1464–1469.

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*(12), 6084–92. https://doi.org/10.1128/JVI.00357-06

[dataset] Weckworth, J., Davis, B., Dubovi, E., Fountain-Jones, N., Packer, C., Cleaveland, S., Craft, M., Eblate, E., Schwartz, M., Mills, L.S.. Roelke-Parker, M. Near whole genome canine distemper virus sequences from Serengeti, 1993-1994; GenBank; https://www.ncbi.nlm.nih.gov/nuccore/MT136705... https://www.ncbi.nlm.nih.gov/nuccore/MT136725

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–9. https://doi.org/10.1101/gr.074492.107

Data Accessibility Statement

Canine distemper virus near whole genome sequence and partial genome sequence data are available in GenBank using accession numbers MT136705-MT136725. Data associated with specimens can be found in tables S1 and S4 in the Supporting information accompanying this manuscript.

Author Contributions

The study was designed by J.W., C.P., M.P., S.C., S.M., M.S. Field investigation was conducted and samples were contributed by M.P., C.P., and S.C. Lab work was conducted by J.W. and E.D. Data analysis was conducted by J.W., B.D., N.F.J. The manuscript was written by J.W. with contributions from all authors.

Tables and Figures

Table 1. RT-qPCR results by species.

Species	CDV qPCR positive	Number tested	Proportion positive
Cheetah	0	2	0
Spotted hyena	9	15	0.6
Domestic dog	6	80	0.075

Black-backed jackal	1	2	0.5
Common genet	0	1	0
African wild dog	0	1	0
Serval	0	2	0
Banded mongoose	0	2	0
Bat-eared fox	3	5	0.6
Aardwolf	0	1	0
African lion	24	96	0.25
Common leopard	0	2	0
Meller's mongoose	0	1	0
Total	43	210	0.2047619

Table 2. Canine distemper virus (CDV) seroprevalence in spotted hyenas captured before, during, and after the 1993-1994 CDV outbreak in the Serengeti Ecological Region.

Capture period	**Pre-outbreak** **Dec 92 – Sept 93**	**Outbreak period** **Dec 93 – July 94**	**Post-outbreak** **Jan 95 – Jun 95**
Seroprevalence (%)	0†	33‡	100
Evidence for CDV infection	None recorded	Lymphoid and CNS lesions consistent with CDV, and/or RT-PCR positive tissue	None recorded
Sample size	12	3	4

† No antibodies were detected at the minimum readable dilution 1:8 (n=11), 1:16 (n=1)

‡ Three of three samples had detectable antibodies, though two were under the cutoff of 1:16 (at 1:8 and 1:12)

Table 3. Bayes support for pairwise diffusion rates between host species during the 1993-1994 canine distemper virus (CDV) outbreak in the Serengeti Ecological Region.

From	To	Bayes Factor	Posterior Probability
Spotted hyena	Domestic dog	0.66	0.23
	Bat-eared fox	0.69	0.23
	African lion	21.77	0.91
Domestic dog	Bat-eared fox	6.10	0.73
	African lion	1.15	0.34
	Spotted hyena	1.64	0.42
Bat-eared fox	African lion	1.45	0.39
	Spotted hyena	2.11	0.48
	Domestic dog	6.46	0.74
African lion	Spotted hyena	6.55	0.74
	Domestic dog	0.98	0.30
	Bat-eared fox	1.06	0.32

Table 4. Amino acid differences between canid and non-canid canine distemper virus (CDV) strains in circulation in Serengeti Ecological Region during a large-scale 1993-1994 CDV outbreak.

Gene	Canid aa	Position on gene	Non-canid aa
P/V	Glycine (G)	134	Serine (S)
P/V	Lysine (K)	280	Glutamic acid (E)
F	Arginine (R)	315	Glutamine (Q)
H	Arginine (R)	160	Lysine (K)
H	Aspartic acid (D)	178	Glycine (G)
H	Arginine (R)	519	Isoleucine (I)
H	Tyrosine (Y)	549	Histidine (H)
L	Leucine (L)	93†	Phenylalanine (F)
L	Asparagine (N)	1402†	Histidine (H)

† Sequence PLE-658 does not bear this mutation

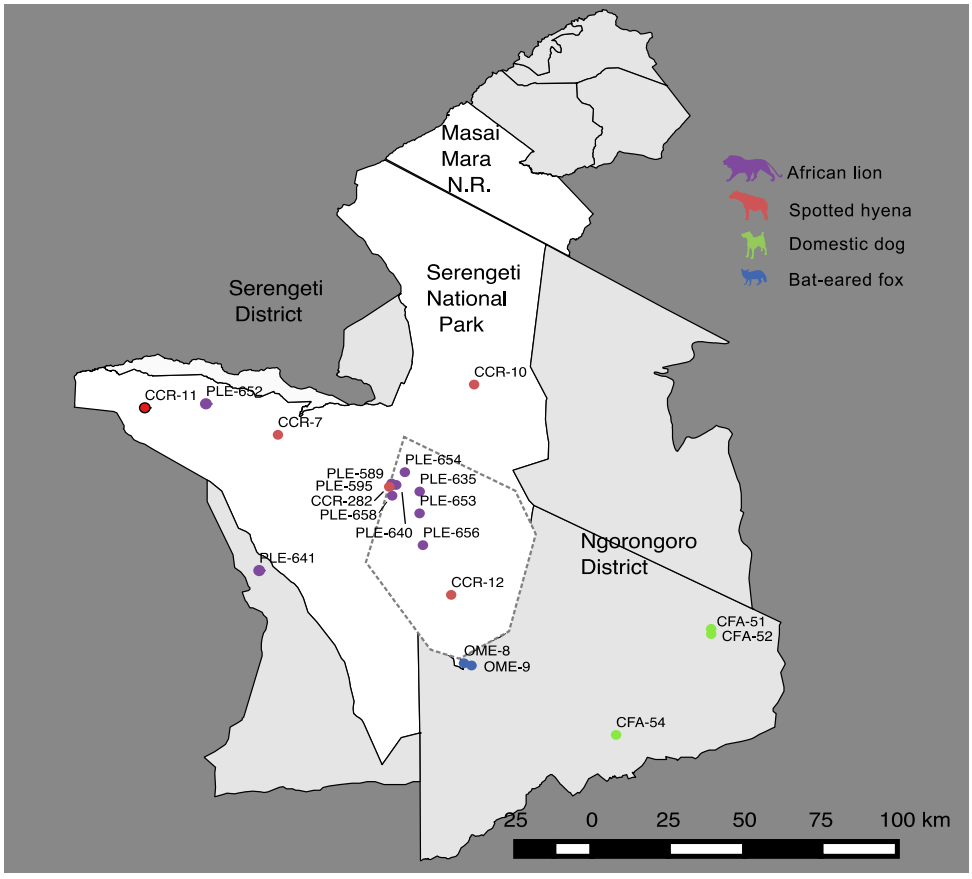

Figure 1. Map showing the location where individual sequences generated were collected in the Serengeti ecological region during the 1993-1994 canine distemper virus outbreak affecting African lions and spotted hyenas. White polygons indicate national parks and light grey polygons represent conservation areas.

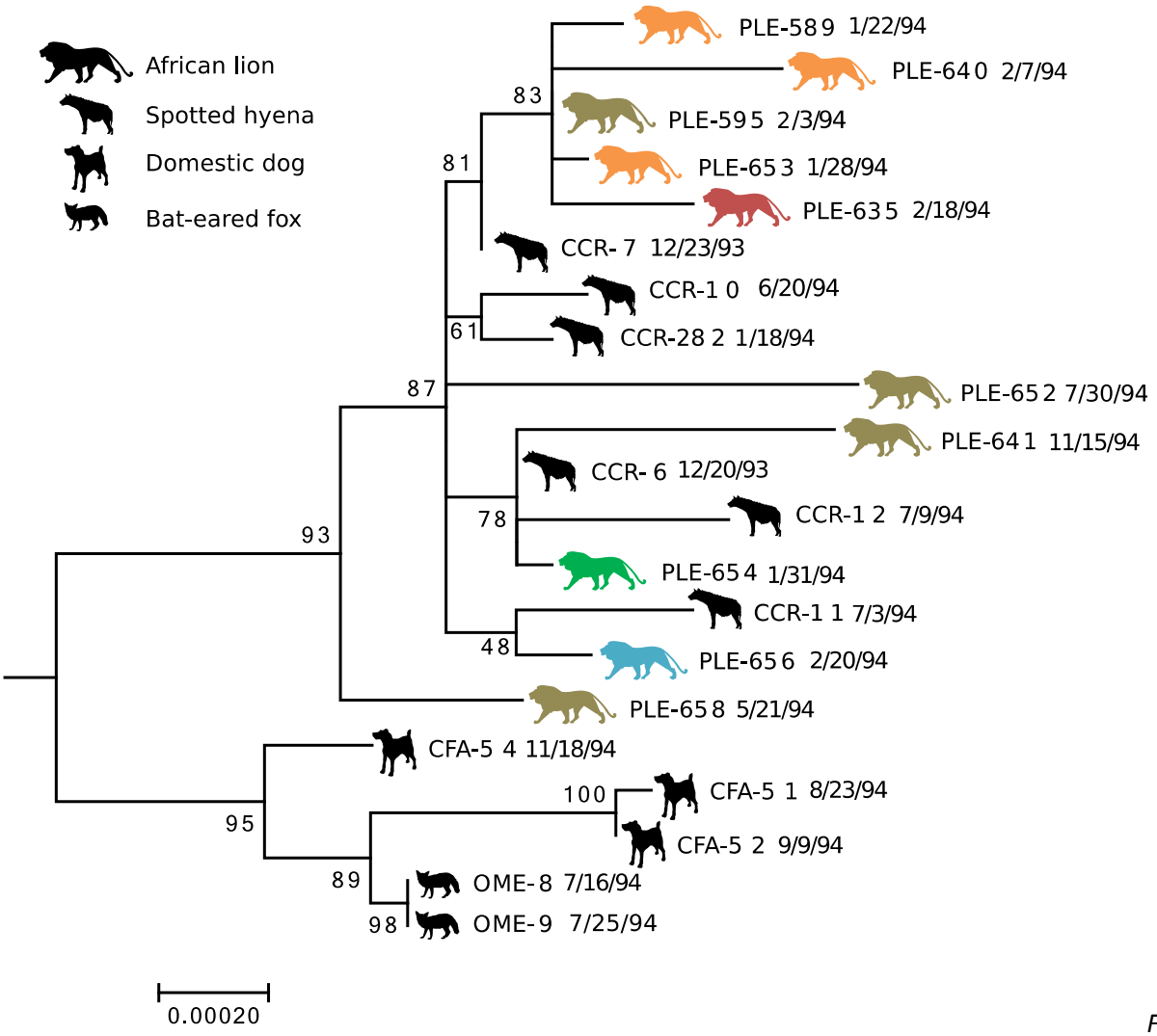

Figure 2. Phylogenetic relationship of near whole canine distemper virus (CDV) genome sequences (11,926 -15,547 bp) sampled during the 1993/1994 Serengeti CDV outbreak reconstructed using the Maximum Likelihood method in MEGA rooted using the Onderstepoort CDV strain (root not shown). Tree tips are annotated with sample name. Symbol shape indicates host species. African lion color indicates pride affiliation: orange = Campsites pride, red = Transects, green = Masai, blue = Simba Numbers, brown = unknown. Bootstrap support for nodes shown next to branches (BS = 1000). Branch lengths are measured by the number of substitutions per site and indicated by the scale bar.

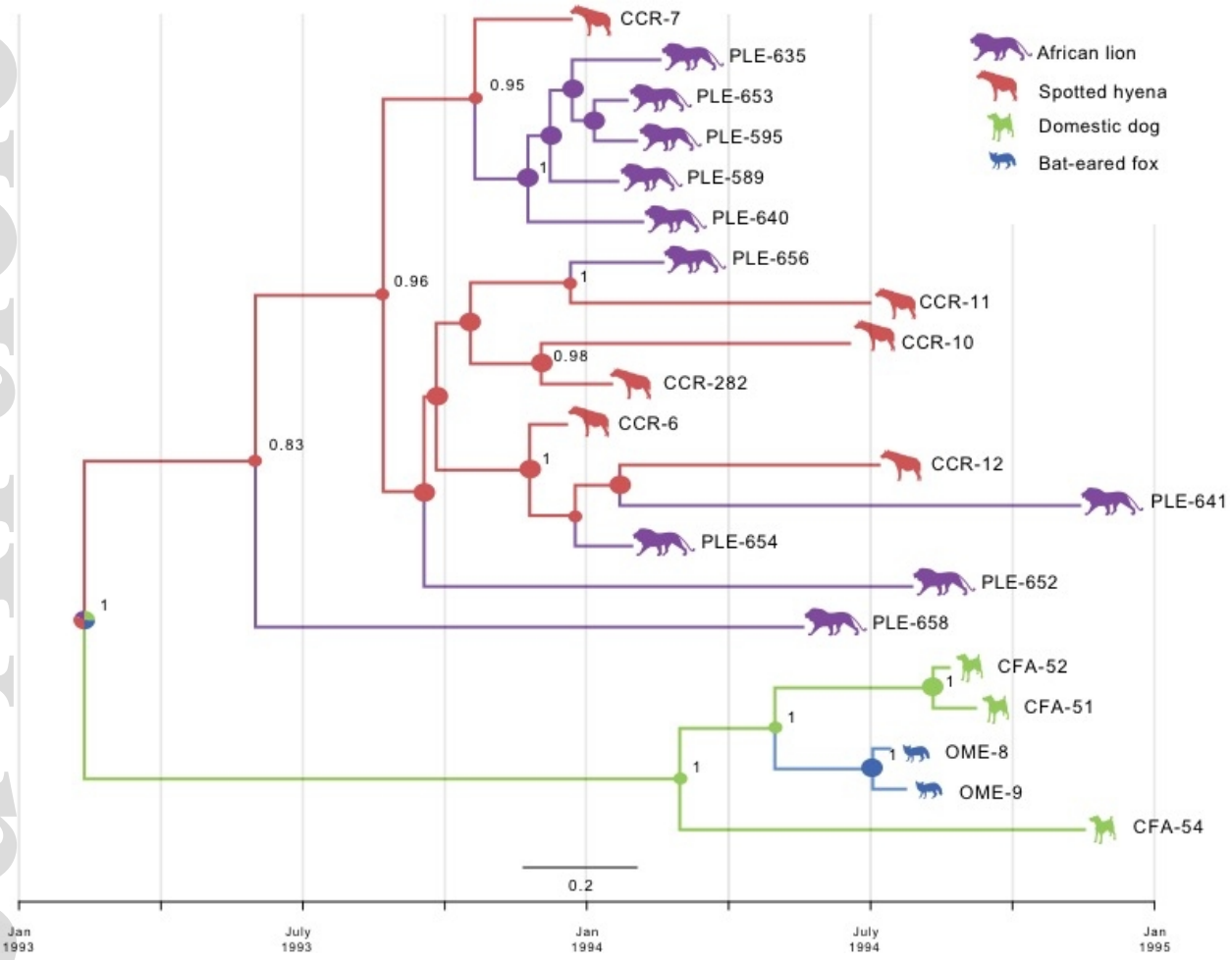

Figure 3. Ancestral host-state reconstruction with asymmetric model of host species transitions over the maximum clade credibility tree generated from near whole CDV genomes sampled in the Serengeti ecological region from 1993-1994. Branch and node color indicate the most probable host state. Host state posterior probabilities ≥0.70 at the nodes are indicated by larger circles with the exception of the root where the probability of each host is represented with a pie graph at the node. Nodes are labeled with the posterior probability of estimated node age ≥0.80.

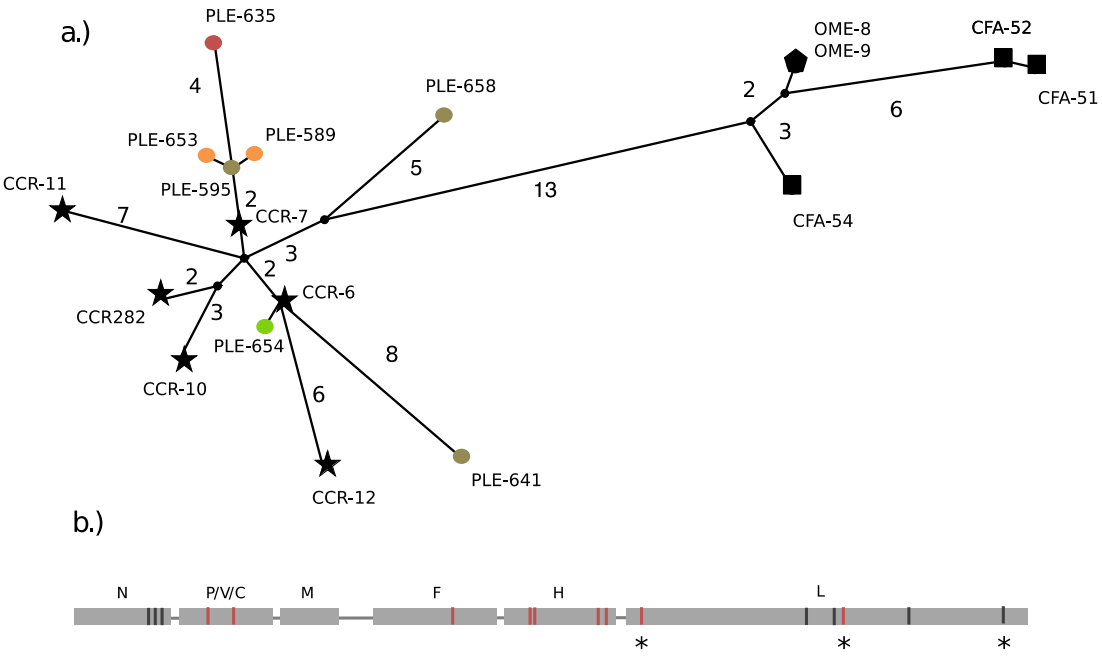

Figure 4. Mutational steps separating canine distemper virus (CDV) genomes sampled during the 1993/1994 Serengeti CDV outbreak from canid and non-canid carnivore hosts. a. Branch lengths in a haplotype network constructed from near whole genome CDV sequences are annotated with the number of mutational steps separating nodes in the graph. Symbol shape indicates host species: star = spotted hyena, circle = African lion, pentagon = bat-eared fox, and square = domestic dog. Circle symbol color indicates pride affiliation: orange = Campsites pride, red =Transects, brown = unknown, green = Masai, blue = Simba Numbers. Two bat-eared foxes have identical sequence, all other symbols represent a single sequence. Thirteen nucleotide substitutions separate all non-canid CDV sequences from all canid CDV sequences. b. Illustration of the position of nucleotide substitutions on the CDV genome separating viruses associated with non-canid hosts from those associated with canid hosts. Grey boxes represent open reading frames of the CDV N, P/V/C, M, F, H, and L genes. The connecting horizontal lines indicate intergenic or untranslated regions. Vertical lines indicate the position of nucleotide substitutions between viruses from non-canid and canid hosts; red = nonsynonymous, black = synonymous. Three additional substitutions separating CDV genomes from canid hosts and all but one non-canid host (PLE-658) are also shown, indicated with asterisks.