RESEARCH

The evolutionary dynamics history of canine distemper virus through analysis of the hemagglutinin gene during 1930–2020

Haoning Wang[1,2,3,4] · Hong Guo[1,2] · Van Gils Hein[5] · Yanchun Xu[4] · Shaopeng Yu[1,2] · Xiaolong Wang[3,4]

Abstract

Canine distemper virus (CDV) is a lethal viral disease of carnivores which is considered to be a serious threat to domestic and wild species. Despite the widespread use of vaccines, CDV still occurs in vaccinated animals and current vaccines does not guarantee complete protection. In this study, a total of 286 hemagglutinin (H) gene sequences of the virus isolated in 25 countries during 90 years (1930–2020) were analyzed by Bayesian maximum likelihood analysis to estimate the population dynamics. We identified the most recent common ancestor (TMRCA) of the virus in 1868 in the USA which arrived in continental Europe in 1948, and from there, the virus spread rapidly to other continents. The *Canidae* family was identified as the original host as well as a source of the subsequent spread. We identified 11 lineages of geographic co-circulating strains globally. The effective population size experienced a two-phase-exponential growth between 2000–2005 and 2010–2012. Our findings provide a novel insight into the epidemic history of canine distemper virus which may facilitate more effective disease management. This study uses a large set of sequencing data on the H gene of CDV to identify distinct lineages of the virus, track its geographic spread over time, analyze its likelihood of transmission within and between animal families, and provide suggestions for improved strategies to combat the virus.

Keywords Canine distemper virus · Hemagglutinin · Phylodynamics · Bayesian · Maximum likelihood

The authors Yanchun Xu and Shaopeng Yu contributed equally to this work.

✉ Xiaolong Wang
yttuhh@yeah.net

1 Heilongjiang Cold Region Wetland Ecology and Environment Research Key Laboratory, School of Geography and Tourism, Harbin University, 109 Zhongxing Road, Harbin 150086, Heilongjiang Province, People's Republic of China

2 School of Geography and Tourism, Harbin University, Harbin 150086, Heilongjiang Province, People's Republic of China

3 The Key Laboratory of Wildlife Diseases and Biosecurity Management of Heilongjiang Province, Harbin, Heilongjiang Province, People's Republic of China

4 Collage of Wildlife and Protected Area, Northeast Forestry University, No. 26, Hexing Road, Xiangfang District, Harbin 150040, Heilongjiang, People's Republic of China

5 Department of Geography, Geoinformatics & Meteorology, University of Pretoria, Pretoria 0002, Gauteng Province, Republic of South Africa

Introduction

Canine distemper virus (CDV) is a highly infectious pathogen in carnivores and represents a serious threat for both wild and domestic species, as it has a broad host range (Kličková et al. 2022 Feb 15). CDV is considered to be a model pathogen in cross-species transmission research. Recent CDV infections have been reported in wild species including snow leopard (*Panthera uncia*) and Amur tiger (*Panthera tigris altaica*), which emphasizes the ecological role of CDV in common veterinary service and conservation strategies (Sulikhan et al. 2018 Jan). CDV is a single-stranded, negative RNA virus of the *Mobillivirus* genus of the *Paramyxoviridae* virus family. Although wild primates can be a host, transmission into human hosts has not been recorded. The infection generates a wide array of clinical symptoms including fever, vomiting, coughing, diarrhea, sneezing, anorexia, and respiratory infection. In the later stages of CDV pathogenesis, the animal can be easily infected by other pathogenic microorganisms, leading to neurologic complications and death (Jiang et al. 2019).

The mortality rate of CDV may reach up to 100%, which is second only to rabies in dogs (Deem et al. 2000).

Due to the continued evolution of CDV, episodes of CD still occur in vaccinated animals (Zhao et al. 2010; Lee et al. 2010; Riley and Wilkes 2015) and administration of the current vaccines does not guarantee complete protection. Therefore, to develop more effectiveness vaccine, successful vaccination campaigns and real-time genetic monitoring of the CDV are essential for accurate prevention. Phylogenic and selection analysis has emerged as a powerful and fast tool for extracting molecular epidemiological information and enriching our understanding of virus origin, evolution, transmission dynamics, and adaptation to host populations (Ke et al. 2015; Panzera et al. 2015). The natural selection that determines genetic variation is a key aspect in assessing the probability of virus adaptation to host populations. This will allow for more targeted approaches in developing vaccines and therapies for various CDV genotypes.

Historically, CDV has shown a wide host range, including members of the families *Canidae*, *Mustelidae*, *Procyonidae*, *Urdidae*, *Hyanidae*, and *Felidae* (Beineke et al. 2015). The range of susceptible hosts has expand to *Macacae*, *Lacepedae*, *Phocidae*, and other families of the *Carnivora* order and now sustain to expand (Liu et al. 2016; Carpenter et al. 1998). In 1925, CDV was found in a silver fox ranches in the USA (Green 1925). Since then, CDV infection has mostly been diagnosed in domestic dogs and on fur animal farms worldwide (Alexander and Appel 1994; Hua and Wang 2004). In 1994, CDV was discovered in wild lions in Serengeti National Park (Tanzania), resulting in a serious population decline (Roelke-parker et al. 1996). Additionally, CDV was reported to have threatened the wild Siberian tiger population in 2000 in the Russian Federation (Quigley et al. 2010 Oct; Seimon et al. 2013). CDV has been reported in giant pandas at Chongqing Zoo, China in 1997 (Li et al. 1999).

The CDV genome encodes six proteins (fusion (F), matrix (M), nucleocapsid (N), hemagglutinin (H), phosphoprotein (P), and polymerase (L)), and each serves a specific role in virus replication and infectivity. For example, the helical N surrounded by an envelope structure and H and F proteins are the major targets of the host immune system and are involved in the process of cell attachment and fusion between virion and host cell (Lamb and Parks 2007). The H protein is a component of the virion envelope glycoprotein spikes that attaches onto cellular receptors such as the signaling lymphocyte activation molecule (SLAM/CD150) or Nectin 4 (PVRL4) (Noyce et al. 2013). H is the most variable protein among all members of the genus *Morbillivirus* (Nikolin et al. 2012), which may explain why CDV has a wider host range than the other *Morbillivirus* members (Pomeroy et al. 2008). Based on the high genetic variability of the H gene, it is commonly used for molecular classification of CDV

strains and it is a suitable target to investigate the genetic relationships between different strains (Demeter et al. 2007). CDV also has geographically distinct lineages, as full-length sequencing of the H gene has led to the identification of 18 CDV lineages (Table 1). In addition, some CDV lineages in Africa and Asia appear to have diverged substantially (Zhao et al. 2010; Woma et al. 2010), which suggests genetic drift of the H gene (Martella et al. 2006).

In this study, a set of Bayesian maximum likelihood methods was used to study the genetic divergence, selection, and viral population dynamics in the evolution of CDV among different outbreaks.

Materials and methods

Sequencing data set

All 526 complete genome H gene sequences of CDV with known sampling date, geographic location, and host between 1930 and 2020 were retrieved from the GenBank database of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). This study was limited by under-sampling of some ancestral strains that were not found in GenBank, especially the ancestral strains isolated before the 1980s.

All sequences were aligned with Multiple Alignment Fast Fourier Transform program [MAFFT, v7.222] (Katoh et al. 2002). Recombination events among sequences were screened using the Recombination Detection Program [RDP,

Table 1 Lineages of the canine distemper virus during 1997–2020

Lineage	References
America 1 (vaccine strain)	Woma et al. (2010)
America 2	Woma et al. (2010)
Europe 1/South America 1	Woma et al. (2010)
Arctic	Woma et al. (2010)
Europe 2/European wildlife	Bhatt et al. (2019); Piewbang et al. (2020)
South America 2	Bhatt et al. (2019); Piewbang et al. (2020)
Rockborn-like	Bhatt et al. (2019); Piewbang et al. (2020)
Europe 3/Arctic-like	Bhatt et al. (2019); Piewbang et al. (2020)
Africa 1	Bhatt et al. (2019); Piewbang et al. (2020)
Africa 2	Bhatt et al. (2019); Piewbang et al. (2020)
South America 3	Bhatt et al. (2019); Piewbang et al. (2020)
South Africa	Bhatt et al. (2019); Piewbang et al. (2020)
Asia 1	Bhatt et al. (2019); Piewbang et al. (2020)
Asia 2	Bhatt et al. (2019); Piewbang et al. (2020)
Asia 3	Bhatt et al. (2019); Piewbang et al. (2020)
Asia 4	Bhatt et al. (2019); Piewbang et al. (2020)
North American 1	Bhatt et al. (2019); Piewbang et al. (2020)
North American 2	Bhatt et al. (2019); Piewbang et al. (2020)

v4.95] (Martin et al. 2015) and SplitsTree [v4.14.6] (Huson and Bryant 2006). Values of $p < 0.05$ were considered to be statistically significant evidence of recombination. The sequences were stratified by the host family, then were manually adjusted using BioEdit [v7.2.5] (Hall 1999). Finally, 286 H gene sequences were obtained by moving vaccine strains (CDV3, Onderstepoor, SnyderHill, Convac) and potential recombinant sequences were removed from the previous dataset of 526 sequences (Supplementary Table 1) spread over 25 countries (Table 2).

Phylogenetic tree construction

The most suitable nucleotide substitution model was selected using jModelTest 2 programs (Darriba et al. 2012). Likelihood-mapping analysis was performed using TREE-PUZZLE [v5.3] to analyze the evolutionary information in the dataset (Schmidt et al. 2002) by analyzing 10,000 randomly chosen quartets for the entire tree. The phylogenetic relationship was deduced from the full-length H gene sequences by the maximum likelihood (ML) method, which was constructed using RAxML [v8.2.10] (Stamatakis 2014) with the GTR + G nucleotide substitution model.

Table 2 Sample number per continent and country

Continent	Sub-continental	Country	Sample (n)
South America	Continental	Argentina	2
Europe	Continental	Austria	4
South America	Continental	Brazil	4
Asia	East	China	139
South America	Continental	Colombia	12
Europe	Continental	Denmark	5
Africa	Sub-Saharan	Ethiopia	1
Africa	Sub-Saharan	Gabon	1
Europe	Continental	Germany	8
North America	Island	Greenland	1
Europe	Continental	Hungary	4
Asia	South	India	5
Europe	Continental	Italy	22
Asia	Island	Japan	17
Asia	Central	Kazakhstan	2
Africa	Sub-Saharan	South Africa	5
Asia	East	South Korea	5
Europe	Continental	Spain	2
Europe	Continental	Sweden	1
Europe	Continental	Switzerland	6
Asia	Island	Taiwan	6
Africa	Sub-Saharan	Tanzania	6
North America	Continental	US	19
South America	Continental	Uruguay	4
Asia	Southeast	Vietnam	5

Bootstrap analysis with 1000 replicates was performed to estimate the reliability of the ML tree. Finally, the ML tree was visualized using FigTree [v1.4.3] (http://tree.bio.ed.ac.uk/software) and Evolview v2 software (http://www.evolgenius.info/evolview) (He et al. 2016) to annotate phylogenetic tree with the location and phylogenetic lineage. Therefore, a lineage was accepted when two assumed conditions were met; the bootstrap cut-off reached 100% of each potential lineage and more than two sequences were classified into a potential lineage. In addition, inter-lineage and intra-lineage evolutionary divergence estimations were calculated for each different lineage using the maximum composite likelihood model (Tamura et al. 2004) with gamma distribution in MEGA [v7.0.14] (Kumar et al. 2016).

Molecular clock and phylodynamic analyses

To investigate the temporal signal of the dataset, TempEst [v1.5] (Rambaut et al. 2016) was used to analyze the correlation between root-to-tip genetic distance and sampling year on the maximum-likelihood tree. Next, the Bayesian Markov chain Monte Carlo (BMCMC) approach was adopted to estimate the evolutionary rate and the mean time period to the most recent common ancestor (TMRCA). The geographic origin was calculated by BEAST [v1.8.2] (Drummond et al. 2012), in which the nodal support was estimated by calculating the posterior probability (PP). In addition, the BEAGLE parallel computation library was used to improve the speed of the likelihood calculations (Ayres et al. 2012).

The nucleotide substitution process was modeled with GTR + G. Next, an uncorrelated lognormal relaxed-clock mode (Drummond et al. 2006) and a Bayesian skyline coalescent model (Drummond et al. 2005) were selected. Then, BMCMC chains were run for 1×10^{10} generations, 10% of which was removed as burn-in, and sampled every 100,000 steps.

Convergence and adequate sampling were assessed by calculating the effective sample size (ESS) of the parameters using Tracer [v1.5] (http://beast.bio.ed.ac.uk/software/tracer). Finally, maximum clade credibility (MCC) trees from all MCMC samples were summarized using TreeAnnotator [v1.8.2], then visualized using Evolview v2 (He et al. 2016).

Lineages were identified by the Bayesian MCC tree based on a posterior probability cut-off of 95%, and the historical population dynamics of the H gene sequences of CDV are depicted in a skyline plot (Drummond et al. 2005). In addition, the major global dispersal routes of major CDV lineages were analyzed using SPREAD (Bielejec et al. 2011) and visualized by Google Earth (Ke et al. 2015). The reliable spreading routes were shown using SPREAD software with support values of the Bayesian factor (BF) > 3 (Bielejec et al. 2011).

The symmetric continuous-time Markov chain (CTMC) method in BEAST [v1.8.2] (Drummond et al. 2012) was implemented to analyze the CDV transmission probability among hosts and in specific geographic locations (i.e., the most infected animal families (*Canidae*, *Felidae*, *Cercopithecidae*, *Mustelidae*, *Procyonidae*, *Hyaenidae*, *Ursidae*, *Paguma*, *Suidae*, and *Sciuridae*) and the five continents (South America, continental Europe, East Asia, sub-Saharan Africa, and North America)) for which the data were obtained (Table 2). Transmission probability was calculated as the interval between 95% of the highest posterior density and 95% of the lowest posterior density.

Selection analysis

Evolutionary analysis of positive selection was performed by the sequences of the operation, as shown in Table 3. The selection codon sites on the H gene was evaluated via the Datamonkey website (http://www.datamonkey.org), and a *p*-value threshold of 0.05 was considered statistically significant. The Bayesian graphical model (BGM) (http://www.datamonkey.org) was used to reconstruct the evolutionary history of each individual codon site to find evidence of co-evolution between sites. A posterior probability (pp) support ≥ 0.95 was considered statistically significant.

The resulting ratio of synonymous (dS) and non-synonymous (dN) mutations was classified into three categories, positive (1), neutral (2), and negative (3), and was determined as follows (Ke et al. 2015):

$$dN - dS > 0 \text{ or } dN/dS > 1 \tag{1}$$

$$dN - dS = 0 \text{ or } dN - dS < 0 \tag{2}$$

$$dN - dS < 0 \text{ or } dN/dS < 1 \tag{3}$$

Comparing the ratio of synonymous (dS) and non-synonymous (dN) mutations is an important indicator of selective pressure at the codon level.

Table 3 The six methods for positive selection and their abbreviations

Method	Abbreviation
Single likelihood ancestor counting	SLAC
Fixed effects likelihood	FEL
Fast Unconstrained Bayesian AppRoximation	FUBAR
Mixed effects model of evolution	MEME
Branch-site unrestricted statistical test for episodic diversification	BUSTED
Bayesian graphical model	BGM

Results

Likelihood-mapping and phylogenetic analyses

The likelihood-mapping showed that the quartets were distributed in the center (69.3%) rather than at the sides (1.9%) or corners (28.7%) of the triangle from the H gene dataset, which indicates a strong star-like phylogenetic signal, which indicates that this dataset is suitable for a phylogenetic reconstruction based on the method of the maximum likelihood analysis (Supplementary Fig. 1). This strong star-like phylogenetic signal also reveals that the spread of CDV population does not follow a slow progress, but a suddenly and sharply increased process in a short time.

Furthermore, the phylogenetic analysis based on the ML method resulted in 11 lineages. To distinguish these lineages, we grouped them according to the maximum-like phylogenic tree analysis with strong bootstrap support and named these lineages lineage 1–lineage 11 (Fig. 1 and Supplementary Fig. 2). Notably, geographically co-circulating lineages are very common (Fig. 2). Nearly half of our lineages are found in more than one continent (lineage 1, lineage 5, lineage 7, lineage 10, and lineage 11), while the others are limited to a single continent (lineage 2, lineage 3, lineage 4, lineage 6, lineage 8, and lineage 9). The remaining eight strains (not grouped) were scattered in the major lineages and had been collected from Argentina (*n* = 1), China (*n* = 2), Denmark (*n* = 1), Hungary (*n* = 1), South Africa (*n* = 1), Greenland (*n* = 1), and the USA (*n* = 1). Lineage 10 was the most widely distributed which is across the five continents. In addition, the genetic diversity within and between each of 11 CDV lineages is provided in Supplementary Fig. 3. The largest genetic distance identified was between lineages 1 and 11 (10.04%) and the smallest genetic distance within a lineage was in lineage 8 (0.82%). Our analysis found that the smallest and largest genetic distances to other lineages were lineage 7 and lineage 11, respectively.

Demographic analysis

The plot of root-to-tip genetic divergence with sampling year revealed a strong temporal signal without obvious outlier sequences, reflecting molecular evolution in relatively clock-like pattern (Fig. 3). The evolution rate was calculated to be $4.433\mathrm{e}^{-4}$ substitutions per site per year. We also estimated a substitution rate of $4.613\mathrm{e}^{-4}$ substitutions per site per year (95% credibility interval: $3.9548\mathrm{e}^{-4}$–$5.4104\mathrm{e}^{-4}$) by Bayesian phylogenetic analysis. It is noteworthy that the substitution rate estimates obtained by these two methods were in close agreement.
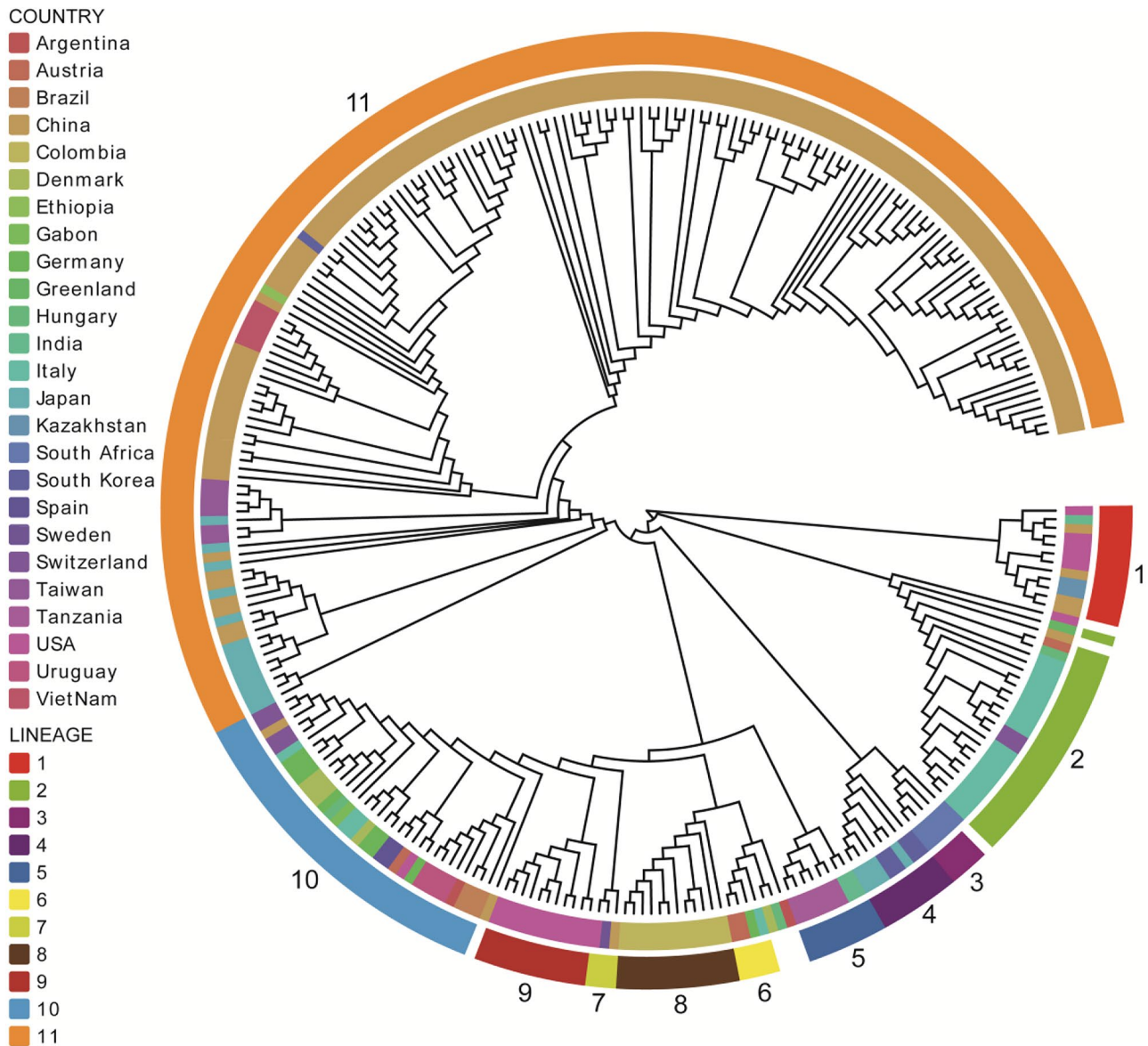
Fig. 1 Maximum-likelihood phylogenic tree for the H gene of the CDV strains. The inner circle represented the tree branches. The ring in the middle is colored by country. The outer ring is colored according to the lineages of the CDV strains

According to posterior probabilities for geographic location in the MCC tree root, TMRCA of the CDV strains originated from the USA. The estimated TMRCA was 1868.055 (95% highest posterior density (HPD) interval: 1829.968–1908.917) (Fig. 4). In addition, the skyline plot suggests that the effective population size has undergone complex dynamics characterized by three exponential growth phases (1930–1950, 2000–2005, and 2010–2013) separated by periods of either declining and constant population size (Fig. 5).

We estimated that there were seven major dispersal routes between continents. CDV originated in the USA and began to spread to the other continents in 1923, first arrived in Europe in 1948, and then rapidly spread to the other continents (Fig. 2). The highest transmission probabilities within the five continents were found in Europe (Supplementary Fig. 4). The transmission probabilities from Europe to East Asia, North America, and South America were very similar.

The transmission probability between species ranking among the 10 animal families (Supplementary Fig. 5) showed that the highest probability was in *Canidae*, followed by *Mustelidae*. Furthermore, the transmission probability of *Canidae* to *Mustelidae* was higher than *Canidae* to others.
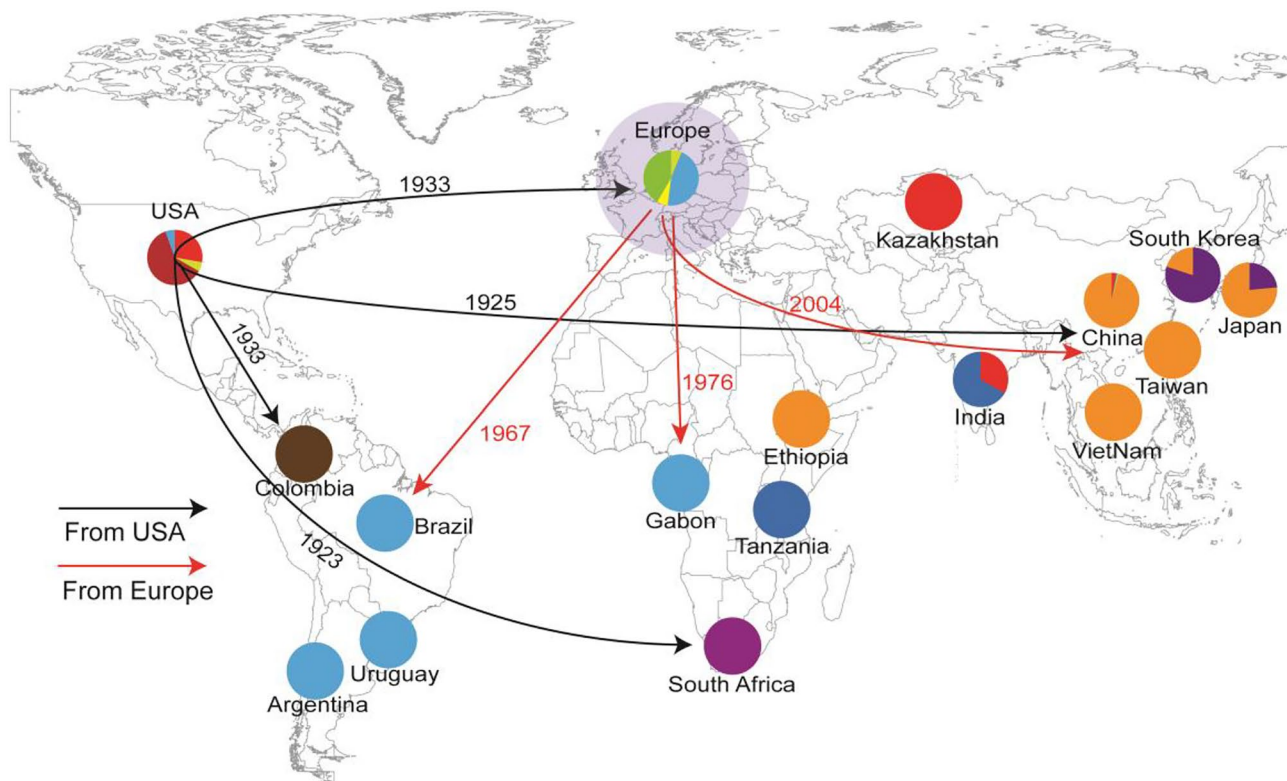
Fig. 2 Geographic distribution of the H gene of the CDV lineages as identified in the present study. Each lineage is color-coded as shown in Fig. 1. The black lines are the dispersal routes from the source in North America to the other continents while the red lines depict the rapid expansion from Europe to other continents. The gray circle represents the source of the rapid expansion from Europe. The eight countries in Europe area are as follows: Sweden, Denmark, Germany, Austria, Switzerland, Hungary, Spain, and Italy, but these countries are too close together to show in this figure, so we use "Europe" to represent these countries. The background map was obtained from http://www.craftmap.box-i.net

Selection analysis

The selection analysis showed strong evidence of evolutionary positive selection (Table 3). Two sites with a positive selection were detected, which were in codons 103 and 549 (Supplementary Table 2). The selection analysis indicated that the positive selection drove amino acid substitutions at position 549 (H to Y) and 103 (I to V) within the signaling lymphocytic activation molecule (SLAM) binding region. The results of the BGM analysis revealed eight pairs of co-evolution sites (Supplementary Fig. 6).

Discussion

The concept of phylodynamics assumes that viral phylogenies are formed by both epidemiological and evolutionary processes (Gog and Grenfell 2002). Therefore, we can identify the migration routes of CDV. We reconstructed the time-scale phylogeographic maximum clade credibility (MCC) trees, indicating that the USA is the original source of CDV

through the root state posterior probability (Fig. 4), which is consistent with a previous report (Panzera et al. 2015). A previous study estimated TMRCA for CDV to be 1886 with a HPD range from 1858 to 1913 (Ke et al. 2015). This estimate was based on the analysis of 208 CDV sequences collected between 1975 and 2011 from 16 countries. One reason for the disagreement between our findings and the previous report is that Ke and colleagues had a data from a smaller geographic area and was more temporally restricted than the dataset that we used. Our dataset contains all of the sequences that Ke et al. (Ke et al. 2015) used in addition to 286 sequences from 25 countries between 1930 and 2020, which likely explains the differences in TMRCA. French veterinarian Henri Carré (Carré 1905) described the CDV in 1905 and successfully transmitted CDV to healthy dogs by inoculation with tissue samples from infected animals. The year 1923 is an important historical milestone in the study of the canine distemper vaccine, as researchers Edward C. Holmes (Holmes and Shope 1923) and Richard E. Shope (Shope 1923) successfully isolated the CDV from infected dogs in the USA during their research and began developing

Fig. 3 Root-to-tip regression of a maximum likelihood (ML) phylogenetic tree of the H gene sequences of the CDV. The colors are the same as shown in Fig. 1 to indicate sampling countries for each H gene sequence
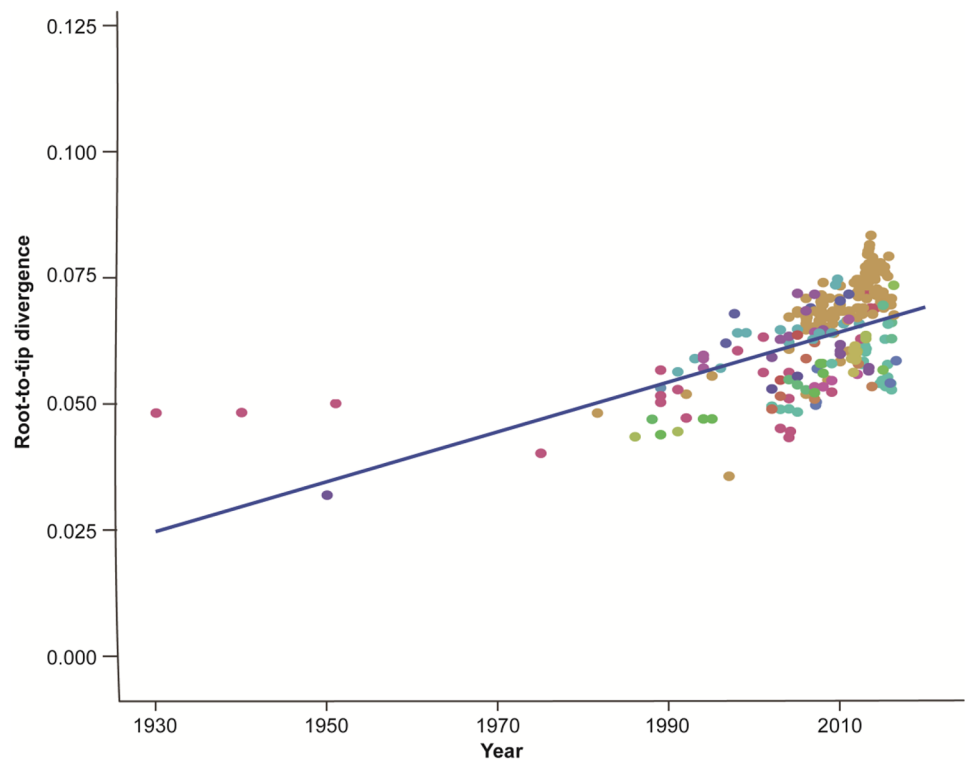
a corresponding vaccine. This early work laid the foundation for future vaccine development, which would eventually lead to the creation of effective vaccines in the 1950s. In our study, the historical CDV sequences before 1930 are unavailable, which leaving a research gap before 1930 in our analysis.

Spatial and temporal dynamic analysis for the geographic spread of CDV revealed more details about the migration patterns. The early overall migration patterns of CDV are roughly estimated to have spread from the USA to the other continents. For European, the later spreading center, CDV left America in 1933 and landed on European continent in 1948 (Fig. 2, the arrived date is not shown on the figure). From Europe, CDV began to spread rapidly to the other continents, as supported by the transmission probability analysis among continent indicating that Europe was higher than others (Supplementary Fig. 5). Figure 2 shows that CDV outbreaks have occur worldwide, which increases the possibility of cross-species transmission in the world. Based on transmission probability analysis among species, canids were the main source of infection among susceptible species, and the transmission probability of canids to mustelids is higher than the others, indicating that canids and mustelids are likely to have frequent contact (Akdesir et al. 2018; Oleaga et al. 2022). The fact that an increasing number of parasite diseases have been found in both canids and mustelids in the world is further evidence for frequent contact between canids and mustelids (Santoro et al. 2019).

Furthermore, likelihood-mapping analysis showed that 69.3% of the quartets were distributed in the center of the triangle, which resulted in a star-like topology phylogenetic signal. Therefore, the CDV outbreak is accompanied by an exponential expansion of its population size. A Bayesian skyline plot is also consistent with this conclusion. We clearly found that the virus underwent exponential epidemic spread until 1950 (Fig. 5). One plausible explanation for the previous assumption about CDV is linked to its first documented arrival in Europe from the USA, after which it underwent a large-scale spread to the other continents so that the viral population increased.

Another interesting finding in our study was noted in the Bayesian skyline plot that the viral population decreased considerably in two periods (1995–2000 and 2005–2010), which is mostly due to the continuous improvement in international cooperation to combat the disease and effective vaccination programs. However, in 2000–2005 and 2010–2013, we found that the viral population increased dramatically, indicating that despite the vaccination procedures adopted around the world, CDV is still considered to be a serious threat (Fig. 5). However, vaccination coverage may vary greatly between different countries and regions although most countries provide CD vaccination. In places with insufficient veterinary service, vaccination rates may be low, and stray dogs in these areas may not receive adequate vaccination protection against CDV which are susceptible infected in CDV (Del et al. 2010). By the same reason, the prevalence
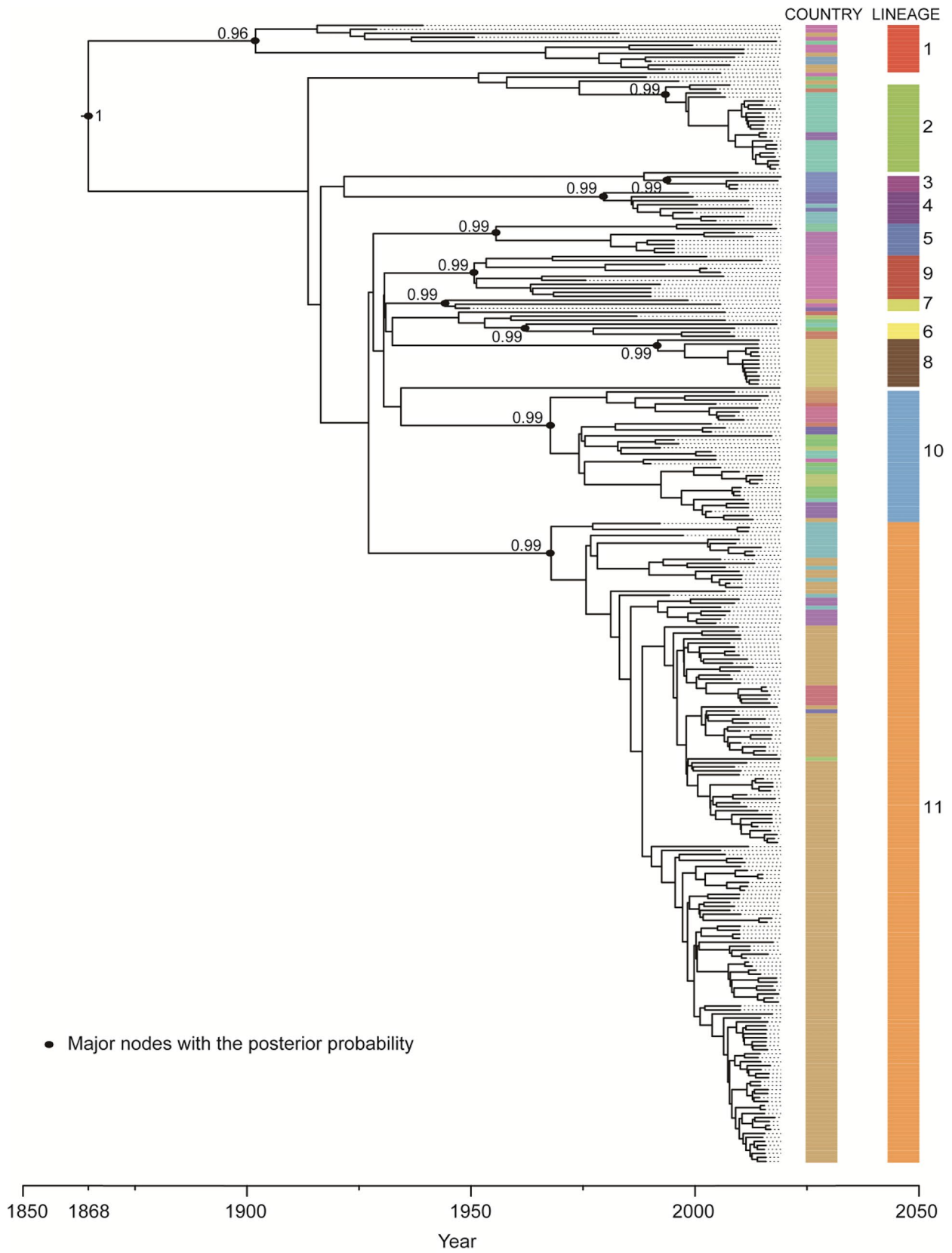
◄**Fig. 4** Bayesian maximum clade credibility (MCC) in the phylogenetic tree of H gene sequences of the CDV. The geographic locations (COUNTRY) and lineages (LINEAGE) are shown by two bars on the right, and each color are shown in Fig. 1. The posterior probability for the major nodes is included at the black dot and the temporal scale bar is provided at the bottom

rate in non-vaccinated dogs is under report. Therefore, insufficient vaccination rates could also be a reason for the surge in viral population during this time. In addition, the offspring of non-vaccinated dogs continue to join the population, and if not immunized, it will inevitably lead to a further increase in population susceptibility to CDV. When an infected dog comes into contact with the susceptible population, it can lead to an outbreak of CD. For example, the total number of CDV-positive samples submitted to the Clinical Virology Laboratory of the Veterinary College of Tennessee University increased from 5% in 2010 to 27% in 2013 (Wilkes et al. 2014) and an outbreak was found in the USA from 2011 to 2013 (Riley and Wilkes 2015). Therefore, it is necessary to consider why the vaccination program is not effective. One possible reason is that most of the current epidemic strains are geographically different from the strains used for vaccine production. The America 1 lineage that was isolated in the USA was used for vaccine production in the 1950s, and the vaccines have remained largely unchanged since that time (Demeter et al. 2010). In recent years, many people have questioned the effectiveness of these vaccines, as CDV occurrences have been reported in previously vaccinated dogs in Argentina (Calderon et al. 2007), Japan (Lan et al. 2006), Mexico (Simon-Martı´nez et al. 2008), and the USA (Kapil et al. 2008). Thus, selection analysis could be employed to guide the vaccination program. Five

methods were used to investigate the selective pressures and co-evolution codon sites acting on H gene sequences of CDV in addition to the BGM approach. The results of this study are generally in agreement with a previous study (Liao et al. 2015) that reported a substitution at residue 549 (H to Y) which has emerged in the population through positive selection. Sawatsky and von Messling (2010) reported that the amino acid residue at site 549 is directly related to host specificity and cross-species transmission. Nikolin et al. (2012) (Nikolin et al. 2012) also found that this substitution may be closely related to the transmission of CDV from dogs to the carnivores. We point out that the characteristics of mutations may be related not only to host tropism but also to different geographical lineages. The key residues in the H protein may be helpful to maintain the important function for host cell entry. Therefore, co-evolution of codon sites are increasingly important for CDV intervention strategies. Eight pairs of co-evolution codon sites have been found and will further guide research on vaccines and therapeutic targets in the future research work.

Improved understanding the spatial distribution of the lineages suggests that we should choose a vaccine strategy that takes the different geographical locations into account. Part of this strategy should include attempts to prevent and control new lineage transmission between different locations. Additionally, the risk level of genetic recombination in virus laboratories must be reduced through strict biosafety management protocols. These efforts will help prevent the emergence of new lineages through genetic engineering and re-emergence of new virus epidemics. One source of long-distance transmission is the breeding of canines, which can involve bringing animals to different continents, where they may encounter strains of the virus that they are vulnerable to and generate new

Fig. 5 Bayesian skyline demographic reconstruction of H gene of the CDV. The vertical axis represents the effective number of infections (Ne) multiplied by the mean viral generation time (τ). The horizontal axis represents the sample collection year. The black line and shaded region represent the median and 95% a high posterior density (HPD) interval of Neτ over time, respectively
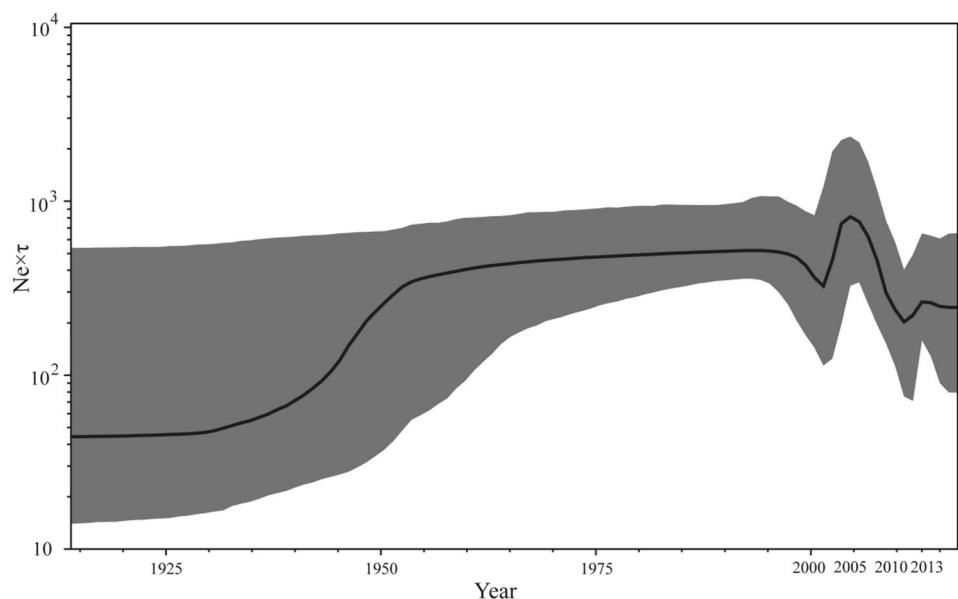
virus lineages. Therefore, actions should be taken to strengthen supervision by customs officials and the use specific primers should be implemented to identify lineages of CDV, which will improve the detection rate.

Many questions in evolutionary biology require immunology and molecular epidemiology perspectives. As mentioned above, most of the CDV sequences were collected from the GenBank database, but not all existing strains are included in this repository. To minimize the sampling bias as much as possible, all lineages were evenly distributed worldwide to maintain sufficient phylogeographic information. We hope to improve the Bayesian framework of this analysis so that it can provide insights into the evolutionary molecular epidemiology of CDV. However, a more comprehensive and reliable CDV database is needed. In addition to evolutionary biology approaches to predict the migration of CDV, more comprehensive testing for the virus may be help us gaining new insight into viral migration and evolutionary dynamics in the future.

Declarations

Conflict of interest The authors declare no competing interests.

References

Akdesir E, Origgi FC, Wimmershoff J et al (2018) Causes of mortality and morbidity in free-ranging mustelids in Switzerland: necropsy data from over 50 years of general health surveillance. BMC Vet Res 14:195. https://doi.org/10.1186/s12917-018-1494-0

Alexander KA, Appel MJ (1994) African wild dogs (Lycaon pictus) endangered by a canine distemper epizootic among domestic dogs near the Masai Mara National Reserve. Kenya J Wildl Dis 30:481–485. https://doi.org/10.7589/0090-3558-30.4.481

Ayres DL, Darling A, Zwickl DJ, Beerli F, Holder MT, Lewis PO et al (2012) BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst Biol 61:170–173. https://doi.org/10.1093/sysbio/syr100

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

Bhatt M, Rajak KK, Chakravarti S, Yadav AK, Kumar A, Gupta V, Chander V, Mathesh K, Chandramohan S, Sharma AK, Mahendran K, Sankar M, Muthuchelvan D, Gandham RK, Baig M, Singh RP, Singh RK (2019) Phylogenetic analysis of haemagglutinin gene deciphering a new genetically distinct lineage of canine distemper virus circulating among domestic dogs in India. Transbound Emerg Dis 66(3):1252–1267. https://doi.org/10.1111/tbed.13142. (Epub 2019 Feb 27 PMID: 30725534)

Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27:2910–2912. https://doi.org/10.1093/bioinformatics/btr481

Calderon MG, Remorini P, Periolo O, Iglesias M, Mattion N, La Torre J (2007) Detection by RT-PCR and genetic characterization of canine distemper virus from vaccinated and non-vaccinated dogs in Argentina. Vet Microbiol 125:341–349. https://doi.org/10.1016/j.vetmic.2007.05.020

Carpenter MA, Appel MJ, Roelke-Parker ME, Munson L, Hofer H, East M et al (1998) Genetic characterization of canine distemper virus in Serengeti carnivores. Vet Immunol Immunopathol 65:259–266. https://doi.org/10.1016/s0165-2427(98)00159-7

Carré H (1905) Sur la maladie des jeunes chiens. C R Acad Sci 140(689–690):1489–1491

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772. https://doi.org/10.1038/nmeth.2109

Deem SL, Spelman LH, Yates RA, Montali RJ (2000) Canine distemper in terrestrial carnivores: a review. J Zoo Wildl Med 31:441–451. https://doi.org/10.1638/1042-7260(2000)031

Del PHL, Vasconcelos AC, Luciana M et al (2010) Canine distemper virus detection in asymptomatic and non vaccinated dogs. Pesqui Vet Bras 2:30. https://doi.org/10.1590/s0100-736x2010000200007

Demeter Z, Lakatos B, Palade EA, Kozma T, Forgách P, Rusvai M (2007) Genetic diversity of Hungarian canine distemper virus strains. Vet Microbiol 122:258–269. https://doi.org/10.1016/j.vetmic.2007.02.001

Demeter Z, Palade EA, Hornyak A, Rusvai M (2010) Controversial results of the genetic analysis of a canine distemper vaccine strain. Vet Microbiol 142:420–426. https://doi.org/10.1016/j.vetmic.2009.10.017

Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLOS Biol 4:e88. https://doi.org/10.1371/journal.pbio.0040088

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22:1185–1192. https://doi.org/10.1093/molbev/msi103

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973. https://doi.org/10.1093/molbev/mss075

Gog JR, Grenfell BT (2002) Dynamics and selection of many-strain pathogens. Proc Natl Acad Sci USA 99:17209–17214. https://doi.org/10.1073/pnas.252512799

Green RG (1925) Distemper in the silver fox. Proc Soc Exp BoilMed 22:546–548. https://doi.org/10.3181/00379727-22-261

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98. https://doi.org/10.1021/bk-1999-0734.ch008

He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S (2016) Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res 44:236–241. https://doi.org/10.1093/nar/gkw370

Holmes EC, Shope RE (1923) A filterable virus, the cause of a second common cold. Science 58(1497):55–56

Hua YP, Wang XL (2004) Diagnosis study on canine distemper outbreak in some fur animal farms in the fall and winter of 2002. J

Northeast Forest Univ 32:84–85. https://doi.org/10.3969/j.issn.1000-5382.2004.01.026

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. https://doi.org/10.1093/molbev/msj030

Jiang Y, Jia S, Zheng D, Li F, Wang S, Wang L, Qiao X, Cui W, Tang L, Xu Y (2019) Don't forget to number the four documents and add them to the references Xia X, Li Y. Protective immunity against canine distemper virus in dogs induced by intranasal immunization with a recombinant probiotic expressing the viral H protein. Vaccines (Basel) 7(4):213. https://doi.org/10.3390/vaccines7040213. PMID: 31835572; PMCID: PMC6963260

Kapil S, Allison RW, Johnston L III, Murray BL, Holland S, Meinkoth J et al (2008) Canine distemper virus strains circulating among North American Dogs. Clin Vaccine Immunol 15:707–712. https://doi.org/10.1128/CVI.00005-08

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436

Ke GM, Ho CH, Chiang MJ, Sanno-Duanda B, Chung CS, Lin MY et al (2015) Phylodynamic analysis of the canine distemper virus hemagglutinin gene. BMC Vet Res 11:164. https://doi.org/10.1186/s12917-015-0491-9

Kličková E, Černíková L, Dumondin A, Bártová E, Budíková M, Sedlák K (2022) Canine distemper virus in wild carnivore populations from the Czech Republic (2012–2020): occurrence, geographical distribution, and phylogenetic analysis. Life (basel) 12(2):289. https://doi.org/10.3390/life12020289.PMID:35207575;PMCID:PMC8874654

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054

Lamb RA, Parks GD (2007) Paramyxoviridae: The viruses and their replication. In: Fields virology (4th edn). Lippincott Williams & Wilkins Press, New York, pp 130–135

Lan NT, Yamaguchi R, Inomata A, Furuya Y, Uchida K, Sugano S et al (2006) Comparative analyses of canine distemper viral isolates from clinical cases of canine distemper in vaccinated dogs. Vet Microbiol 115:32–42. https://doi.org/10.1016/j.vetmic.2006.01.010

Lee MS, Tsai KJ, Chen LH, Chen CY, Liu YP, Chang CC et al (2010) The identification of frequent variations in the fusion protein of canine distemper virus. Vet J 183:184–190. https://doi.org/10.1016/j.tvjl.2008.10.001

Li JZ, Xia XZ, He HB, Yu C, Hu GX, Fan QS et al (1999) Gene sequence analysis diagnosis of giant pandas infected by canine distemper virus. J Chin Vet Sci 19:448–450. https://doi.org/10.16303/j.cnki.1005-4545.1999.05.010

Liao P, Guo L, Wen YJ, Yang YL, Cheng SP (2015) Phylogenetic features of hemagglutin gene in canine distemper virus strains from different genetic lineages. 8(4):6607–6612. PMID: 26131292; PMCID: PMC4484007

Liu PC, Chen CA, Chen CM, Yen CH, Lee MH, Chuang CK et al (2016) Application of xenogeneic anti-canine distemper virus antibodies in treatment of canine distemper puppies. J Small Anim Pract 57:626–630. https://doi.org/10.1111/jsap.12557

Martella V, Cirone F, Elia G, Lorusso E, Decaro N, Campolo M, Desario C et al (2006) Heterogeneity within the hemagglutinin genes of canine distemper virus (CDV) strains detected in Italy. Vet Microbiol 116:301–309. https://doi.org/10.1016/j.vetmic.2006.04.019

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1:1–5. https://doi.org/10.1093/ve/vev003

Nikolin VM, Wibbelt G, Michler FU, Wolf P, East ML (2012) Susceptibility of carnivore hosts to strains of canine distemper virus from distinct genetic lineages. Vet Microbiol 156:45–53. https://doi.org/10.1016/j.vetmic.2011.10.009

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. https://doi.org/10.1016/j.virol.2012.11.011

Oleaga Á, Vázquez CB, Royo LJ, Barral TD, Bonnaire D, Armenteros JÁ, Rabanal B, Gortázar C, Balseiro A (2022) Canine distemper virus in wildlife in south-western Europe. Transbound Emerg Dis 69:e473–e485. https://doi.org/10.1111/tbed.14323

Panzera Y, Sarute N, Iraola G, Hernández M, Pére R (2015) Molecular phylogeography of canine distemper virus: geographic origin and global spreading. Mol Phylogenet Evol 8:1055–7903. https://doi.org/10.1016/j.ympev.2015.06.015

Piewbang C, Chansaenroj J, Kongmakee P, Banlunara W, Poovorawan Y, Techangamsuwan S (2020) Genetic adaptations, biases, and evolutionary analysis of canine distemper virus Asia-4 lineage in a fatal outbreak of wild-caught civets in Thailand. Viruses 12(4):361. https://doi.org/10.3390/v12040361. PMID:32224857; PMCID:PMC7232145

Pomeroy LW, Bjornstad ON, Holmes EC (2008) The evolutionary and epidemiological dynamics of the paramyxoviridae. J Mol Evol 66:98–106. https://doi.org/10.1007/s00239-007-9040-x

Quigley KS, Evermann JF, Leathers CW, Armstrong DL, Goodrich J, Duncan NM, Miquelle DG (2010) Morbillivirus infection in a wild Siberian tiger in the Russian Far East. J Wildl Dis 46(4):1252–1256. https://doi.org/10.7589/0090-3558-46.4.1252. (PMID: 20966275)

Rambaut A, Lam TT, Max Carvalho L, Pybus OG (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2:vew007. https://doi.org/10.1093/ve/vew007

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. Virol J 12:219. https://doi.org/10.1186/s12985-015-0445-7

Roelke-parker ME, Munson L, Packer CA (1996) Canine distemper virus epidemic in Serengeti lions (Panthera leo). Nature 379:441–445. https://doi.org/10.1038/379441a0

Santoro M, Auriemma C, Lucibelli MG, Borriello G, D'Alessio N, Sgroi G, Veneziano V, Galiero G, Fusco G (2019) Molecular detection of Babesia spp. (Apicomplexa: Piroplasma) in free-ranging canids and mustelids from Southern Italy. Front Vet Sci 6:269. https://doi.org/10.3389/fvets.2019.00269

Sawatsky B, von Messling V (2010) Canine distemper viruses expressing a hemagglutinin without N1-glycans lose virulence but retain immune-suppression. J Virol 84:2753–2761. https://doi.org/10.1128/JVI.01813-09

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504. https://doi.org/10.1093/bioinformatics/18.3.502

Seimon TA, Miquelle DG, Chang TY, Newton AL, Korotkova I, Ivanchuk G et al (2013) Canine distemper virus: an emerging disease in wild endangered Amur tiger (Panthera tigris altaica). Mbio 4:e00410-413. https://doi.org/10.1128/mBio.00410-13

Shope RE (1923) The detection of filterable viruses in animal tissues. J Exp Med 38(2):229–244

Simon-Martı´nez J, Ulloa-Arvizu R, Soriano VE, Fajardo R (2008) Identification of a genetic strain of canine distemper virus from clinical cases in two vaccinated dogs in Mexico. Vet J 175:423–6. https://doi.org/10.1016/j.tvjl.2007.01.015

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Sulikhan NS, Gilbert M, Blidchenko EY, Naidenko SV, Ivanchuk GV, Gorpenchenko TY, Alshinetskiy MV, Shevtsova EI, Goodrich JM, Lewis JCM, Goncharuk MS, Uphyrkina OV, Rozhnov VV, Shedko SV, McAloose D, Miquelle DG, Seimon TA (2018) Canine distemper virus in a wild far Eastern Leopard (Panthera Pardus Orientalis). J Wildl Dis 54(1):170–174. https://doi.org/10.7589/2017-03-065. (Epub 2017 Oct 20 PMID: 29053427)

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 101:11030–11035. https://doi.org/10.1073/pnas.0404206101

Wilkes RP, Sanchez E, Riley MC, Kennedy MA (2014) Real-time reverse transcription polymerase chain reaction method for detection of canine distemper virus modified live vaccine shedding for differentiation from infection with wild-type strains. J Vet Diag Invest 26:27–34. https://doi.org/10.1177/1040638713517232.8

Woma TY, van Vuuren M, Bosman AM, Quan M, Oosthuizen M (2010) Phylogenetic analysis of the haemagglutinin gene of current wild-type canine distemper viruses from South Africa: lineage Africa. Vet Microbiol 143:126–132. https://doi.org/10.1016/j.vetmic.2009.11.013

Zhao JJ, Yan XJ, Chai XL, Martella V, Luo GL, Zhang HL et al (2010) Phylogenetic analysis of the haemagglutinin gene of canine distemper virus strains detected from breeding foxes, raccoon dogs and minks in China. Vet Microbiol 140:34–42. https://doi.org/10.1016/j.vetmic.2009.07.010