Virus Research

Origin and spreading of canine morbillivirus in South America

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Keywords:	Canine morbillivirus (CDV); hemagglutinin, South America; phylogenetic
Corresponding Author:	Yanina Panzera University of Science Montevideo, Uruguay
First Author:	Yanina Panzera
Order of Authors:	Yanina Panzera
	E. Fuques
	G. Tomás
	S. Grecco
	E. Condon
	C. Techera
	A. Marandino
	N. Sarute
	J. Aldaz
	J.E. Gutiérrez
	A. Benech
	R Pérez
	Y Panzera
Abstract:	Canine distemper virus (CDV) is a Morbillivirus (Canine morbillivirus) that greatly impacts domestic and wildlife carnivores worldwide. The CDV RNA genome has high genetic variability, evidenced by several lineages that follow a global geographic pattern. The evolutionary trajectories and population dynamics of CDV lineages are still unclear and debatable, particularly in South America, where relatively few sequences are available. We performed phylogenetic and Bayesian analyses using an updated dataset of the highly variable hemagglutinin (H) gene, including seven South American countries. The time to the most recent common ancestor (tMRCA) of the current CDV lineages was dated to the early 1900s in North America. Maximum likelihood and Bayesian maximum clade credibility phylogenies showed similar topologies with two main branches (L1 and L2) corresponding to the NA1 lineage (L1) and the remaining lineages worldwide (L2). The four circulating lineages in South America (EU1/SA1, SA2, SA3, NA4/SA4) arose from independent migration events from North America and Europe. North American strains colonized most northern South American countries via Ecuador and then Colombia and Peru, originating the SA3 and NA4/SA4 lineages during their spread. The entry and expansion in the southern part of South America (Argentina, Brazil, Chile, and Uruguay) occurred through three independent migration events and gave rise to the EU1/SA1 and SA2 lineages. South American lineages have specific combinations of amino acids under positive selection that constitute signatures of taxonomic and evolutionary relevance. Our findings provide a comprehensive scenario for the origin and migration routes of Canine morbillivirus in South America and highlight the importance of phylodynamics in understanding the geographic patterns of modern genetic variability.
Suggested Reviewers:	Rebecca Wilkes beckpen@uga.edu
	Michele Lunardi

	michelelunardi@gmail.com
	Alberto Pessia alberto.pessia@helsinki.fi
	Xijun Yan Chinese Academy of Agricultural Sciences yanxijun@caas.cn
Opposed Reviewers:	

Highlights

South America lineages arose from independent migration from North America and Europe.

The circulating South America lineages spreading by intra-continental migration routes.

South American lineages have amino acids under signatures of evolutionary relevance.

Origin and spreading of canine morbillivirus in South America

Fuques E.[1], Tomás G.[1], Grecco S.[1], Condon E.[1], Techera C.[1], Marandino A.[1], Sarute N.[1], Aldaz J.[2] Gutiérrez J.E.[3], Benech A.[4], Pérez R[1]. , Panzera Y[1*].

1. Sección Genética Evolutiva, Departamento de Biología Animal, Instituto de Biología, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400, Montevideo, Uruguay.

2. Escuela de Medicina Veterinaria y Zootecnia, Facultad de Ciencias Agropecuarias, Universidad Estatal de Bolı´var, Av. Ernesto Che Guevara s/n. Guaranda, Ecuador.

3. Grupo Medicina Regenerativa. Universidad Científica del Sur. Lima. Perú

4. Unidad de Clínica y Hospital Veterinario, Facultad de Veterinaria. Universidad de la República. Montevideo, Uruguay.

*Corresponding author. Fax: +598 2525 86 17. E-mail addresses:

ypanzera@fcien.edu.uy, yaninapanzera@gmail.com (Panzera Y).

Keywords: Canine morbillivirus (CDV); hemagglutinin, South America; phylogenetic.

Abstract

Canine distemper virus (CDV) is a *Morbillivirus* (*Canine morbillivirus*) that greatly impacts domestic and wildlife carnivores worldwide. The CDV RNA genome has high genetic variability, evidenced by several lineages that follow a global geographic pattern. The evolutionary trajectories and population dynamics of CDV lineages are still unclear and debatable,

particularly in South America, where relatively few sequences are available. We performed phylogenetic and Bayesian analyses using an updated dataset of the highly variable hemagglutinin (H) gene, including seven South American countries. The time to the most recent common ancestor (tMRCA) of the current CDV lineages was dated to the early 1900s in North America. Maximum likelihood and Bayesian maximum clade credibility phylogenies showed similar topologies with two main branches (L1 and L2) corresponding to the NA1 lineage (L1) and the remaining lineages worldwide (L2). The four circulating lineages in South America (EU1/SA1, SA2, SA3, NA4/SA4) arose from independent migration events from North America and Europe. North American strains colonized most northern South American countries via Ecuador and then Colombia and Peru, originating the SA3 and NA4/SA4 lineages during their spread. The entry and expansion in the southern part of South America (Argentina, Brazil, Chile, and Uruguay) occurred through three independent migration events and gave rise to the EU1/SA1 and SA2 lineages. South American lineages have specific combinations of amino acids under positive selection that constitute signatures of taxonomic and evolutionary relevance. Our findings provide a comprehensive scenario for the origin and migration routes of *Canine morbillivirus* in South America and highlight the importance of phylodynamics in understanding the geographic patterns of modern genetic variability.

1. Introduction

Canine distemper virus (CDV), renamed *canine morbillivirus* (*Paramyxoviridae*; *Morbillivirus*) by the International Committee on Taxonomy of Viruses (2016), is the etiological agent of canine distemper, a viral disease with a great impact on domestic dogs and wild carnivores. CDV was first reported to infect dogs, but now it is known that the virus can affect several carnivorous and non-carnivorous species (Martinez-Gutierrez and Ruiz-Saenz, 2016). New species are

frequently proposed as hosts for CDV, and many cases of re-emergence in naive populations of susceptible species have been described (Garigliany et al., 2018; Lunardi et al., 2018a).

CDV has a single-stranded, negative-sense RNA genome of approximately 16 kb in length that encodes for six structural and two non-structural proteins. The virion envelope contains two glycoproteins: hemagglutinin (H) and fusion protein (F), which are responsible for host cell receptor binding and virus-cell fusion, respectively (von Messling et al., 2001). The H protein's high genetic and antigenic variability makes it suitable for understanding CDV evolution patterns and the antigenic variation between field and vaccine strains developed around the 1950s (An et al., 2008; Espinal et al., 2014; Ke et al., 2015; Martella et al., 2006; Panzera et al., 2012; Radtanakatikanon et al., 2013; Riley and Wilkes, 2015; Woma et al., 2010; Zhao et al., 2010).

Paleopathological analysis, historical records and codon usage suggest that CDV originated in South America through the cross-species transfer of the Measles virus from humans to dogs during the XVIII century (Uhl et al. 2019). The South American countries Ecuador and Peru are the hypothetical geographical origin of the virus around 1746 (Juan and Ulloa, 1748). Two decades later, outbreaks of a disease presumably caused by CDV were described in Europe and North America (Flemming 1882). The implementation of molecular phylodynamic analysis allows studying another aspect of the CDV story based on analysis of modern, current circulating strains. A coalescence analysis showed that the most recent common ancestor of the present CDV strains emerged in the United States in mid-1880 (Panzera et al., 2015). The apparent discrepancy between the studies could be attributed to the lack of information about ancient strains and their descendants and sampling bias from different geographic areas and hosts. Phylodynamic studies reveal the origin and spreading of current viruses, while the

hypothesis of extinct or unsampled lineages demands other interdisciplinary approaches and can not be discarded (Uhl et al., 2019).

More sampling efforts are desirable to better understand the origin and spreading of worldwide CDV strains. There has been a significant increase in the number of H gene sequences available in the GenBank database, reaching over 700 sequences, tripling the number analyzed in previous studies. However, only four countries have reported full-length sequences of the H gene in South America (Argentina, Brazil, Colombia, and Uruguay). These H gene sequences correspond to four lineages with unevenly geographic distribution. South American strains from Argentina, Brazil, and Uruguay are of the widely distributed Europe 1/South America 1 (EU1/SA1) lineage. In addition, most Argentinean strains belong to the South America 2 lineage (SA2), and Colombian strains constitute the South America 3 (SA3) (Espinal et al., 2014; Fischer et al., 2016; Panzera et al., 2015). Recently, a new lineage that encompasses Colombian and North American strains was reported as South/North America 4 (NA4/SA4) (Espinal et al., 2014; Fischer et al., 2016; Panzera et al., 2015; Duque-valencia et al. 2019).

The present study obtained the complete H gene sequence from new CDV strains from Ecuador, Peru, and Uruguay and performed an extensive comparative analysis with the available South American sequences, including previously uncharacterized CDV strains from Chile. The inferred origin, migration routes, and amino acid characteristics of South American CDV strains reveal new findings on the evolutionary dynamics of CDV in South America.

2. Materials and methods

2.1 Amplification of the full-length H gene and dataset construction

The full-length H gene sequence (1824 nt) from Ecuadorian, Peruvian, and Uruguayan CDV strains was obtained by RT-PCR upon isolation of viral RNA from biological samples (conjunctival swabs and urine), according to Panzera et al. (2012). Consensus sequences were submitted to the GenBank database under the following accession numbers ON533741-ON533748.

2.2 Phylogenetic analysis

Two datasets (Dataset I and II) were built with the newly obtained (n=8) and all available H sequences of South American CDV.

Dataset I. This full-length H gene (1824 nt) dataset was built with 459 sequences described worldwide, including 56 South American strains retrieved from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov)

Dataset II. To encompass a more representative dataset, and according to their availability in the GenBank database, sequences of the almost complete (1742 nt) H gene was retrieved (n=482), including 23 additional sequences from South America (Brazil and Chile).

Nucleotide and deduced amino acid sequences were aligned using MEGA 7 software (Kumar et al., 2016). Possible recombinant sequences were scanned with RDP4 software using default parameters. The phylogenetic signal of dataset II was analyzed using the Likelihood Mapping method using TREE-PUZZLE 5.3 software (Schmidt, 2002). Phylogenetic trees were reconstructed using the maximum likelihood (ML) method with the general time-reversible plus gamma distribution (GTR + G) model selected by using the Akaike information criterion

(AIC) in jModelTest 2.1. and PHYML 3.3 software (Darriba et al., 2015; Guindon and Gascuel, 2003). Phocine Distemper Virus strains (PDV) (HQ007902.1) were used as an external group. Statistical support was obtained by bootstrap analysis of 1000 pseudoreplicates. Trees were visualized in FigTree v1.4.0 (http://tree.-bio.ed.ac.uk/software/figtree).

2.3 Phylodynamic analysis

Phylodynamic analysis was performed using H sequences (dataset II) and associated metadata (country, year of isolation, and host). The temporal signal was evaluated using TempEst software (Rambaut et al., 2016). Evolutionary relationships, the time to the most recent common ancestors (tMRCA), internal nodes, and migration routes were inferred using Bayesian Markov chain Monte Carlo (BMCMC) methods. BEAST 1.10.0 package (Drummond and Rambaut, 2007) was employed through the Cipres Science Gateway (http://www.phylo.org) using a strict clock model and a constant population model as a tree prior. MCMC length was 1×10^8 generations, and the convergence of the analysis was determined using Tracer 1.6 by considering values of Effective Sample Size (ESS) > 200. The Maximum Clade Credibility (MCC) tree was generated with TreeAnnotator and visualized using FigTree 1.4.3 software (http://tree.bio.ed.ac.uk/software/figtree/). From MCC phylogenies, posterior probability (PP), 95% highest probability density (HPD), and posterior state probability (PSP) values for each geographic location of origin were retrieved. Statistical support of each migration event was determined using Bayes Factor (BF) implemented in SPREAD software v1.0.7.

2.4 Bayesian method for amino acids analysis of the H protein

K-Pax 2 software was used to identify evolutionary signals in the amino acid sequences of the H protein (Pessia et al., 2015). K-Pax 2 was run with the default prior setting, and the

clustering result with the highest log posterior probability was chosen between 50 independent runs.

3. Results

3.1 CDV sequences

Eight novel H gene sequences were obtained from Ecuador (n=4), Peru (n=2), and Uruguay (n=2) by PCR amplification and sequencing of two overlapping fragments using primers and conditions previously described by Panzera et al. (2012).

No recombinant strains were identified in either of the datasets employed, according to RDP4 software results.

3.2 Phylogenetic analysis and lineage assignment using full-length H gene (dataset I)

The CDV strains analyzed were divided into two main intercontinental branches (L1 and L2) with significant bootstrap values (0.89) (Fig. 1A). The L1 branch comprises a few North American and Vietnamese strains of the North America 1 lineage (NA1) and previously unclassified Chinese and Kazakh strains. The L2 branch splits into the 14 already described geographical lineages (most bootstrap values > 0.97).

South American (SA) strains group within the four already characterized lineages (EU1/SA1, SA2, SA3, and NA4/SA4). The newly obtained Uruguayan strains associate with the EU1/SA1 lineage, while the new Ecuadorian and Peruvian groups associate with the Colombian and American NA4/SA4 lineage (Fig. 1B).

3.3 Phylogenetic analysis using nearly complete H sequences (dataset II)

The phylogenetic signal of dataset II, including unclassified South American strains, was assessed by likelihood-mapping analysis. The 92.4% of the quartets were equally distributed in the regions representing well-resolved phylogenies (Fig. 1 Suppl).

The phylogeny obtained from dataset II showed that the Chilean strains fall with high statistical support (0.99 bootstraps) within the EU1/SA1 lineage (Fig. 2 Suppl).

3.4 Ancestral state's reconstruction and migration routes

The phylodynamic analysis was performed with dataset II, covering a range of 43 years (1975–2017) and encompassing strains from 25 countries, including seven from South America. To avoid potential sampling bias, some countries were grouped into locations as follows: Europe a (Spain, Portugal, and Italy), Europe b (Austria, Hungary, and Germany), East Asia (Japan, Taiwan, and South Korea), Continental Asia (China, Vietnam, and Kazakhstan), Africa (South Africa, Tanzania, and Ethiopia) and North America (the United States and Mexico). Also, the number of strains in over-represented countries was decreased by removing random sequences. Thus, a balanced dataset was obtained (359 sequences, 13 locations) to represent CDV population dynamics better. Root-to-tip analysis carried out with TempEst produced a correlation coefficient of 0.09, and the graph cut the X-axis (an approximation to the tMRCA) at the value 1898 (data not shown).

The MCC tree has a branch topology supported with maximum posterior probabilities (PP~1). The tMRCA was dated 108 years ago from the most recently analyzed strain (2017) and had North America as the most probable geographic location. From that North American ancestor, CDV strains split into two major branches (L1 and L2). The L1 branch includes some continental Asia strains, while L2 encompasses all the remaining strains (Fig. 2).

3.5 Spreading in South America

In South America, the first CDV strains arrived from the Europe b region to Argentina (PSP=0.75) in 1960, with a 95% HPD ranging from 1955 to 1967. This migration event is statistically supported (BF > 250) and originates the strains of the SA2 lineage that is unique to Argentina (Fig. 2 and 3).

A second migration event occurred in Ecuador, coming from the North American strain (BF >178) approximately in 1961 (95% HPD: 1954–69) and spread to the bordering countries Peru (BF=397) and Colombia (BF=270). The entry of CDV to Colombia occurred firstly in 1994 (95% HPD: 1989–1999), giving rise to the SA3 lineage. Following entries in 2001 and 2006 originated the NA4/SA4 lineage, comprising Colombian, Ecuadorian, Peruvian, and North American strains (Fig. 2 and 3).

New entries (BF> 214) reached southern South America from the Europe b region's countries (Austria, Hungary, and Germany) (PSP = 0.77). The first strains came to Brazil in 1980 (PSP 0.93) and spread to the bordering countries Argentina and Uruguay; entry to Uruguay occurred in 2002 and 2014 (BF=259).

The third migratory event from the Europe b region (BF >87) reached Chile in 1999 with a 95% HPD from 1995 to 2003. These South American and European strains are now grouped within the EU1/SA1 lineage (Fig. 2 and 3).

3.6 Bayesian H amino acid analysis (K-Pax)

South American CDV strains were classified into 9 major groups or Bayesian clusters (C-1 to C-9) with amino acid markers and signatures (Fig. 2, Table 1).

C-1 and C-2 include the NA4/SA4 lineage strains from Ecuador, Peru, Colombia, and North America. C-1 comprises three Ecuadorian strains, one from Peru and four from Colombia, and

has 14 aa markers. C-2 includes five strains from North America (n=2), Colombia (n=1), Peru (1), and Ecuador (N=1) and is characterized by 14 amino acids. C-3 includes a strain from Colombia belonging to the SA3 lineage and has 21 characteristic residues. C-4 to C-8 include strains of the EU1/SA1 lineage. C-4 groups Chilean strains with 11 amino acid residues. C-5 is exclusive from Brazil (Fig. 2) and contains four dog strains isolated in 2012 and three strains isolated from *Tamandua tetradactyla* (anteater) with 14 aa marker residues. C-6 included several strains from Brazil and one from Argentina and was characterized by six amino acid markers. C-7 and C-8 include Uruguayan strains. C-7 comprises strains collected from 2008–to 2012 and has eleven amino acid markers. C-8 associates two Uruguayan strains herein characterized and one strain from Brazil with 16 amino acid markers. C-9 comprises Argentinean strains from the SA2 lineage sharing 14 amino acid markers.

4. Discussion

Accurate sampling of genomic sequences is a key parameter to estimating time-scale virus evolution by phylodynamic analysis, especially for those viruses with high evolutionary rates. Limited datasets' usage may lead to inexact evolutionary estimations (Baele et al., 2017). The high variability of the CDV H gene and the steady growth of sequences available on Genbank demands constant analysis, especially if viral sequences from under-analyzed countries become available. To expand and update the demographic history of current CDV strains, we obtained the full-length H gene from Uruguayan, Ecuadorian and Peruvian strains, and constructed a worldwide dataset, including strains from seven South American countries (Argentina, Chile, Colombia, Brazil, Ecuador, Peru, and Uruguay). Further, we identified evolutionary signals in the amino acid sequences of the H protein potentially associated with specific lineages or geographical regions (Table 1).

The tMRCA of the current CDV strains was dated to the early 1900s in North America, which agrees with previous phylogenetic analyses (Panzera et al. 2015; Ke et al. 2015; Fischer et al. 2016, Duque-valencia et al. 2019). Accordingly, strains of all circulating CDV lineages can be traced back to a single common ancestor that emerged in North America and spread worldwide.

The phylogenetic trees, inferred by Maximum likelihood (ML) and Bayesian maximum clade credibility (MCC), showed similar topologies and revealed the existence of two main branches, L1 and L2, both with North American ancestors (Fig. 1 and 2). L2 strains successfully spread worldwide, giving rise to most lineages, including the four South American lineages EU1/SA1, SA2, SA3, and NA4/SA.

South American lineages were structured according to their continental origin and migratory dynamics (Fig. 2 and 3). The SA2 lineage, restricted to Argentina, emerged in 1960 due to an intercontinental migration from Europe. Although SA2 strains are not distributed outside Argentina, they are highly prevalent and infect domestic and wild carnivores within the country (Calderón et al., 2007; Ferreyra et al., 2009). The SA2 lineage has specific amino acid signatures (C-9 in Fig. 2 and Table 1). The Bayesian clustering identified the 530D residue, which is located in the β-5-sheet of the H protein ectodomain and associated with host switching: G/E (domestic dogs) to R/D/N (non-dog hosts) [8, 41, 42] (Mccarthy et al. 2007; Ke et al. 2015; Nikolin et al. 2017). According to ML and MCC trees, the SA2 and the European wildlife strains are associated with a sister group. This phylogenetic relationship supports a possible origin of the SA2 lineage from wildlife strains or a strong gene flow between wild and domestic host populations.

Almost simultaneously with the introduction of CDV from Europe to Argentina, an independent migratory event of North American strains reached Ecuador (Fig. 3). This country served as a secondary hub of dispersion to bordering countries Colombia and Peru. The migratory event that reached Colombia in 1994 created the SA3 lineage comprising CDV strains collected in 2011-12 from a single province of Colombia (Espinal et al. 2014). The SA3 lineage shares unique residue signatures associated with the C-3 group (Fig. 2, Table 1). There are no other Colombian strains available or records about the first description of distemper in the country to confirm this migratory scenario. Following these events, in the 2000s, two successive intracontinental migrations occurred from Ecuador to Colombia and Peru, giving rise to the NA4/SA4 lineage. Recent reports estimate the tMRCA of the SA3 and NA4/SA 4 lineages in 1964 and 1925, respectively (Duque-valencia et al., 2019). These discrepancies may be due to the limited dataset used in the study, which included only Colombian and North American strains, excluding more ancestral representatives. The number of sequences,

particularly from heterochronic strains with different geographic origins, is needed to obtain more robust information. Our Bayesian clustering analysis supports these different migratory routes by identifying amino acid signatures in each lineage. The SA3 lineage forms the entire C-3 group, while the NA4/SA4 lineage is grouped into two sets (C-1 and C-2) according to different migratory events (Fig. 2, Table 1). The only Ecuadorian strain that belongs to C-2 has a basal position on the MCC clade, supporting that the Ecuadorian strains are ancestral. C1 and C2 shared several amino acids undergoing positive selection, but the only H protein amino acid marker shared by the SA3 and NA4/SA4 lineages is 315V, which should be regarded as an ancestral marker. Our results reflect the origin and years of divergence between both lineages and specific selective local pressures.

The next two independent migratory events also occurred from Europe b to the southern South American countries Brazil and Chile (Fig. 3). Both European and South American strains fell within the EU1/SA1 lineage (PP=1). This lineage is the most widely distributed in southern South America (Argentina, Brazil, Chile, and Uruguay). It is highly structured according to its geographic location, migratory routes, and unique amino acid signatures. Some residues (51T, 103V, and 217T) of the SA and European strains (data not shown) can be considered ancestral states. In contrast, others are unique for SA, for example, 445S, which is present in all SA strains of the EU1/SA1 lineage but absent in the European ones.

The first migratory event from Europe occurred in Brazil in 1982 and later spread and reached Uruguay and Argentina (Fig. 3). The occurrence of Brazil as the most probable hub for southern South America is consistent with the expansion of the Brazilian virus populations in recent decades (Fischer et al., 2016). A specific substitution, R580Q, located on the β-5-sheet ectodomain of the H protein (Messling et al., 2004; Langedijk et al., 2011), was associated with strains with greater spread capacity (Fischer et al., 2016). Although this substitution was

detected in some Brazilian strains, Bayesian clustering identified 580Q as a positively selected residue in the “old” Uruguayan strains (C-7 in Fig. 2) and the C-8 group, including one Brazilian strain and the herein characterized Uruguayan strains. This evidences that the Uruguayan strains have emerged by two independent migratory events from Brazil around 2000 (old Uruguayan strains) and 2015 (more recent ones). These results evidence the ongoing migratory events occurring between these two countries and highlight the importance of constantly characterizing new strains. The Brazilian strains are highly divergent (Fig. 1 and 2) and were previously classified into sub-genotypes, based on H protein amino acid p-distance, within the EU1/SA1 lineage (Budaszewski et al., 2014). Our analysis showed that these strains gather into three Bayesian clusters (C-5, C6, and C-8, Fig. 2), reflecting local migratory and adaptative events. C-5 is composed of CDV strains from domestic and recently identified non-dog-hosts (Lunardi et al., 2018b), revealing the genetic flow between these carnivore hosts and highlighting the risk that CDV represents to wildlife. The serine residue in position 530 (530S) supports the genetic flow between species, but the relationship of this substitution with the host jump to non-dog hosts requires further analysis. C-6 is also restricted to Brazil, except for one Argentinean strain (Calderón et al. 2007). The C-6 cluster exhibits a basal position within the MCC tree, which may represent the core of the migrant strains in Brazil. Lastly, in another independent migratory event, CDV strains from Europe reached Chile approximately 20 years later than the Brazilian migratory event (Fig. 3). Phylogenetic, coalescence and Bayesian clustering analyses support that the Chilean strains comprise a monophyletic clade of recent emergence (the 2000s). This is concordant with the first reports of the disease in Chile, which date from 2003 (Gonzalez-Acuña et al. 2003).

In conclusion, our findings support that present CDV lineages have the same ancestral origin in North America but experienced particular migratory dynamics. The current CDV lineages in South America emerged due to heterochronic migratory waves from the primary (United

States) and secondary (Europe) distribution hubs that entered different countries to spread further and undergo local diversification.

Legends to Figures and Tables

Figure 1. Maximum Likelihood phylogeny based on 1824 nt of the length of CDV hemagglutinin gene sequences (dataset I). Main intercontinental lineages L1 and L2 are shown (A), and worldwide lineages characterized to date are identified by colors (B). Asia 1 to 4: AS1-4, East Africa: EA; South Africa: ZA, Europe 1 to 3: EU1-3, North America 1 to 4: NA1-4. Bootstrap support values are indicated for each main node.

Figure 2. Maximum Clade Credibility (MCC) tree based on 1742 nt of the length of CDV hemagglutinin gene sequences (dataset II). CDV circulating lineages are indicated and compressed, except for South American strains. Posterior probability values and locations are shown for main nodes (Brazil: BR, Ecuador: EC, Eurb: Europe B, North America: NA). Colors represent the different South American countries, and the X-axis represents the time scale. Clusters obtained with K-pax2 Software are indicated as C_1 to C_9.

Figure 3. South American migrations routes. Location, year, and BF values are indicated in the migration events.

Table 1. List of the hemagglutinin amino acid sites to each Bayesian cluster determined by K-Pax software.

Supplementary Figs

Figure 1 Suppl. Likelihood mapping analysis based on dataset II. The upper triangle shows the distribution of patterns. The left triangle indicates the tree topologies T1, T2, and T3. The right triangle shows probabilities: the corners of the triangle represent well-resolved topologies, while the center and the sides represent unresolved and partially unresolved, respectively.

Figure 2 Suppl. Maximum Likelihood phylogeny based on dataset II. Lineages that do not include South American sequences are collapsed and indicated. Bootstrap values are shown for the main nodes. The Chilean strains (red clade) within EU1/SA1 lineage are indicated. Bootstrap support values are displayed for each main node.

Acknowledgements

This work was supported in part by Agencia Nacional de Investigación e Innovación (ANII-FCE) from Uruguay and the Programa de Desarrollo de las Ciencias Básicas (PEDECIBA).

References

An, D.J., Yoon, S.H., Park, J.Y., No, I.S., Park, B.K., 2008. Phylogenetic characterization of canine distemper virus isolates from naturally infected dogs and a marten in Korea. Veterinary Microbiology 132, 389–395. https://doi.org/10.1016/j.vetmic.2008.05.025

Baele, G., Suchard, M.A., Rambaut, A., Lemey, P., 2017. Emerging Concepts of Data Integration in Pathogen Phylodynamics. Syst Biol. 66(1):e47-e65. doi: 10.1093/sysbio/syw054. PMID: 28173504; PMCID: PMC5837209

Budaszewski, F., Dubina, L., Nunes, M., Teles, E., Diniz, C., Travassos, B., Martella, V., Ikuta, N., Ricardo, V., Wageck, C., 2014. Genotyping of canine distemper virus strains circulating in Brazil from 2008 to 2012. Virus Research 180, 76–83. https://doi.org/10.1016/j.virusres.2013.12.024

Calderon, M.G., 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

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2015. jModelTest 2: more models, new heuristics and high-performance computing Europe PMC Funders Group. Nature Methods 9, 772. https://doi.org/10.1038/nmeth.2109

Duque-Valencia J., N.R. forero-Muñoz, F.J. Díaz, E. Martins, P. Barato & Julian Ruiz-Saenz, 2019. Phylogenetic evidence of the intercontinental circulation of a canine distemper virus lineage in the Americas. Scientific RepoRtS 9:15747 https://doi.org/10.1038/s41598-019-52345-9

Drummond, A.J., Rambaut, A., 2007. Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 8, 79–96. https://doi.org/doi:10.1186/1471-2148-7-214

Espinal, M.A., Díaz, F.J., Ruiz-Saenz, J., 2014. Phylogenetic evidence of a new canine distemper virus lineage among domestic dogs in Colombia, South America. Veterinary Microbiology 172, 168–176. https://doi.org/10.1016/j.vetmic.2014.05.019

Ferreyra, H., M.G. Calderón, D. Marticorena, C. Marull, Barrios Caro L., 2009. Canine distemper infection in crab-eating fox (Cerdocyon thous) from Argentina. Journal of Wildlife Dis. 45, 1158–1162. DOI: 10.7589/0090-3558-45.4.1158

Fischer, C.D.B., Gräf, T., Ikuta, N., Lehmann, F.K.M., Passos, D.T., Makiejczuk, A., Silveira, M.A.T., Fonseca, A.S.K., Canal, C.W., Lunge, V.R., 2016. Phylogenetic analysis of canine distemper virus in South America clade 1 reveals unique molecular signatures of the local epidemic. Infection, Genetics and Evolution 41, 135–141 https://doi.org/10.1016/j.meegid.2016.03.029

Fleming, G., 1882. A Chronological History of Animal Plagues; Their History, Nature, and Prevention. Chapman and Hall, London

Garigliany, M., Sarlet, M., Franssen, M., Desmecht, D., Volpe, R., Lesenfant, C., Paternostre, J., Linden, A., 2018. Re-emergence of canine distemper in wildlife in Belgium. Veterinary Record 182, 439. https://doi.org/10.1136/vr.k1610

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic biology 52, 696–704. https://doi.org/10.1080/10635150390235520

Juan, J., Ulloa, A., 1748. Relación histórica del viage a la América Meridional hecho de orden de S. Mag. para medir algunos grados de meridiano Terrestre, y venir por ellos en conocimiento de la verdadera figura, y Magnitud de la Tierra, con otras varias Observaciones Astronómicas, y Phísicas. Antonio Marin, Madrid

Ke, G.M., Ho, C.H., Chiang, M.J., Sanno-Duanda, B., Chung, C.S., Lin, M.Y., Shi, Y.Y., Yang, M.H., Tyan, Y.C., Liao, P.C., Chu, P.Y., 2015. Phylodynamic analysis of the canine distemper virus hemagglutinin gene. BMC Veterinary Research 11, 1–15. https://doi.org/10.1186/s12917-015-0491-9

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7 : Molecular Evolutionary Genetics Analysis Version 7 . 0 for Bigger Datasets Brief communication. Molecular Biology and Evolution. https://doi.org/10.1093/molbev/msw054

Langedijk J.PM, J. Janda, F. C Origgi, C. Örvell, M. Vandevelde, A. Zurbriggen, P. Plattet, 2011. Canine distemper virus infects canine keratinocytes and immune cells by using overlapping and distinct regions located on one side of the attachment protein. Journal of Virology, 11242–11254. Doi:10.1128/JVI.05340-11

Lunardi, M., Darold, G.M., Amude, A.M., Headley, S.A., Sonne, L., Yamauchi, K.C.I., Boabaid, F.M., Alfieri, A.F., Alfieri, A.A., 2018a. Canine distemper virus active infection in order Pilosa, family Myrmecophagidae, species Tamandua tetradactyla. Veterinary Microbiology 220, 7–11. https://doi.org/10.1016/j.vetmic.2018.04.030

Lunardi, M., Darold, G.M., Amude, A.M., Headley, S.A., Sonne, L., Yamauchi, K.C.I., Boabaid, F.M., Alfieri, A.F., Alfieri, A.A., 2018b. Canine distemper virus active infection in order Pilosa, family Myrmecophagidae, species Tamandua tetradactyla. Veterinary Microbiology 220, 7–11. https://doi.org/10.1016/j.vetmic.2018.04.030

Martella, V., Cirone, F., Elia, G., Lorusso, E., Decaro, N., Campolo, M., Desario, C., Lucente, M.S., Bellacicco, A.L., Blixenkrone-Møller, M., Carmichael, L.E., Buonavoglia, C., 2006. Heterogeneity within the hemagglutinin genes of canine distemper virus (CDV) strains detected in Italy. Veterinary Microbiology 116, 301–309. https://doi.org/10.1016/j.vetmic.2006.04.019

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, 1–11. 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. Proc Biol Sci 274, 3165–3174, https://doi.org/10.1098/rspb.2007.0884

von Messling, V., Milosevic, D., Devaux, P. and Cattaneo, R., 2004. Canine distemper virus and measles virus fusion glycoprotein trimers: partial membrane-proximal ectodomain cleavage enhances function. Journal of virology 78, 7894-903. DOI: 10.1128/JVI.78.15.7894-7903.2004

Nikolin V.J., X. A Olarte-Castillo, N. Osterrieder, H. Hofer, E. Dubovi, C.J Mazzoni, E. Brunner, K.V Goller, R.D Fyumagwa, P.D Moehlman, D. Thierer, M.L East. 2017. Canine distemper virus in the Serengeti ecosystem: molecular adaptation to different carnivore species. Mol Ecol 26, 2111–2130, https://doi.org/10.1111/mec.13902 (2017).

Panzera, Y., Calderón, M.G., Sarute, N., Guasco, S., Cardeillac, A., Bonilla, B., Hernández, M., Francia, L., Bedó, G., La Torre, J., Pérez, R., 2012. Evidence of two co-circulating genetic lineages of canine distemper virus in South America. Virus Research 163, 401–404. https://doi.org/10.1016/j.virusres.2011.10.008

Panzera, Y., Sarute, N., Iraola, G., Hernández, M., Pérez, R., 2015. Molecular phylogeography of canine distemper virus: Geographic origin and global spreading. Molecular Phylogenetics and Evolution 92, 147–154. https://doi.org/10.1016/j.ympev.2015.06.015

Pessia, A., Grad, Y., Cobey, S., Puranen, J.S., Corander, J., 2015. K-Pax2: Bayesian identification of cluster-defining amino acid positions in large sequence datasets. Microbial Genomics 1. https://doi.org/10.1099/mgen.0.000025

Radtanakatikanon, A., Keawcharoen, J., Charoenvisal, N. taya, Poovorawan, Y., Prompetchara, E., Yamaguchi, R., Techangamsuwan, S., 2013. Genotypic lineages and restriction fragment length polymorphism of canine distemper virus isolates in Thailand. Veterinary Microbiology 166, 76–83. https://doi.org/10.1016/j.vetmic.2013.05.015

Rambaut, A., Lam, T.T., Carvalho, L.M., Oliver, G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution. 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, 1–10. https://doi.org/10.1186/s12985-015-0445-7

Schmidt, H.A., 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. https://doi.org/10.1093/bioinformatics/18.3.502

Uhl Elizabeth W., C. Kelderhouse, J. Buikstra, J.P. Blickd, B. Bolon, R.J. Hogan, 2019. New world origin of canine distemper: Interdisciplinary insights. International Journal of Paleopathology 24, 266–278. https://doi.org/10.1016/j.ijpp.2018.12.007

von Messling, V., Zimmer, G., Herrler, G., Haas, L., Cattaneo, R., 2001. The Hemagglutinin of Canine Distemper Virus Determines Tropism and Cytopathogenicity. Journal of Virology 75, 6418–6427. https://doi.org/10.1128/JVI.75.14.6418-6427.2001

Woma, T.Y., van Vuuren, M., Bosman, A.M., Quan, M., Oosthuizen, M., 2010. Phylogenetic analysis of the haemagglutinin gene of current wild-type canine distemper viruses from South Africa: Lineage Africa. Veterinary Microbiology 143, 126–132. https://doi.org/10.1016/j.vetmic.2009.11.013

Zhao, J.J., Yan, X.J., Chai, X.L., Martella, V., Luo, G.L., Zhang, H.L., Gao, H., Liu, Y.X., Bai, X., Zhang, L., Chen, T., Xu, L., Zhao, C.F., Wang, F.X., Shao, X.Q., Wu, W., Cheng, S.P., 2010. Phylogenetic analysis of the haemagglutinin gene of canine distemper virus strains detected from breeding foxes, raccoon dogs and minks in China. Veterinary Microbiology 140, 34–42. https://doi.org/10.1016/j.vetmic.2009.07.010

Figure 1

Click here to access/download;Figure;Figure 1.tif

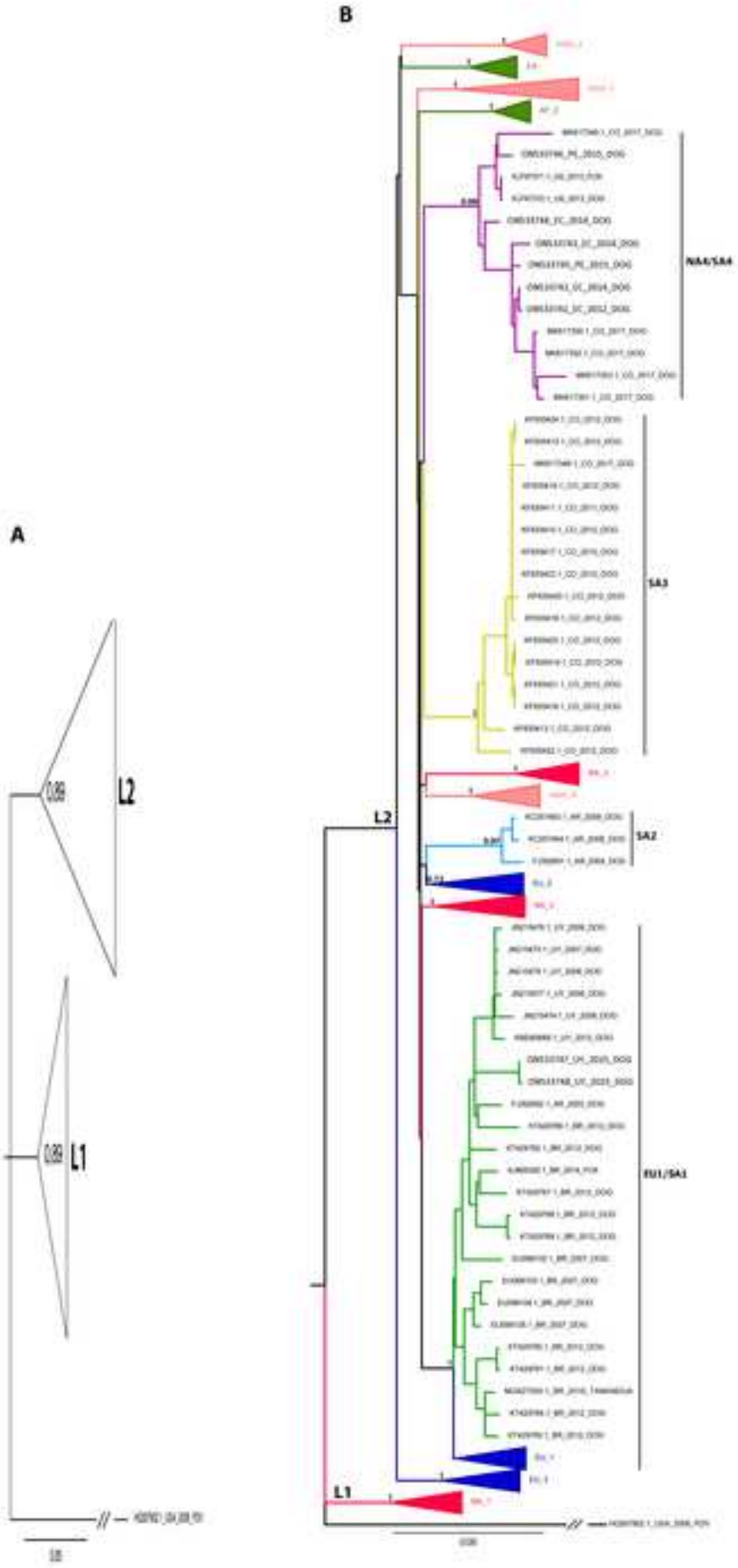

Figure 2

Click here to access/download;Figure;Figure 2.tif

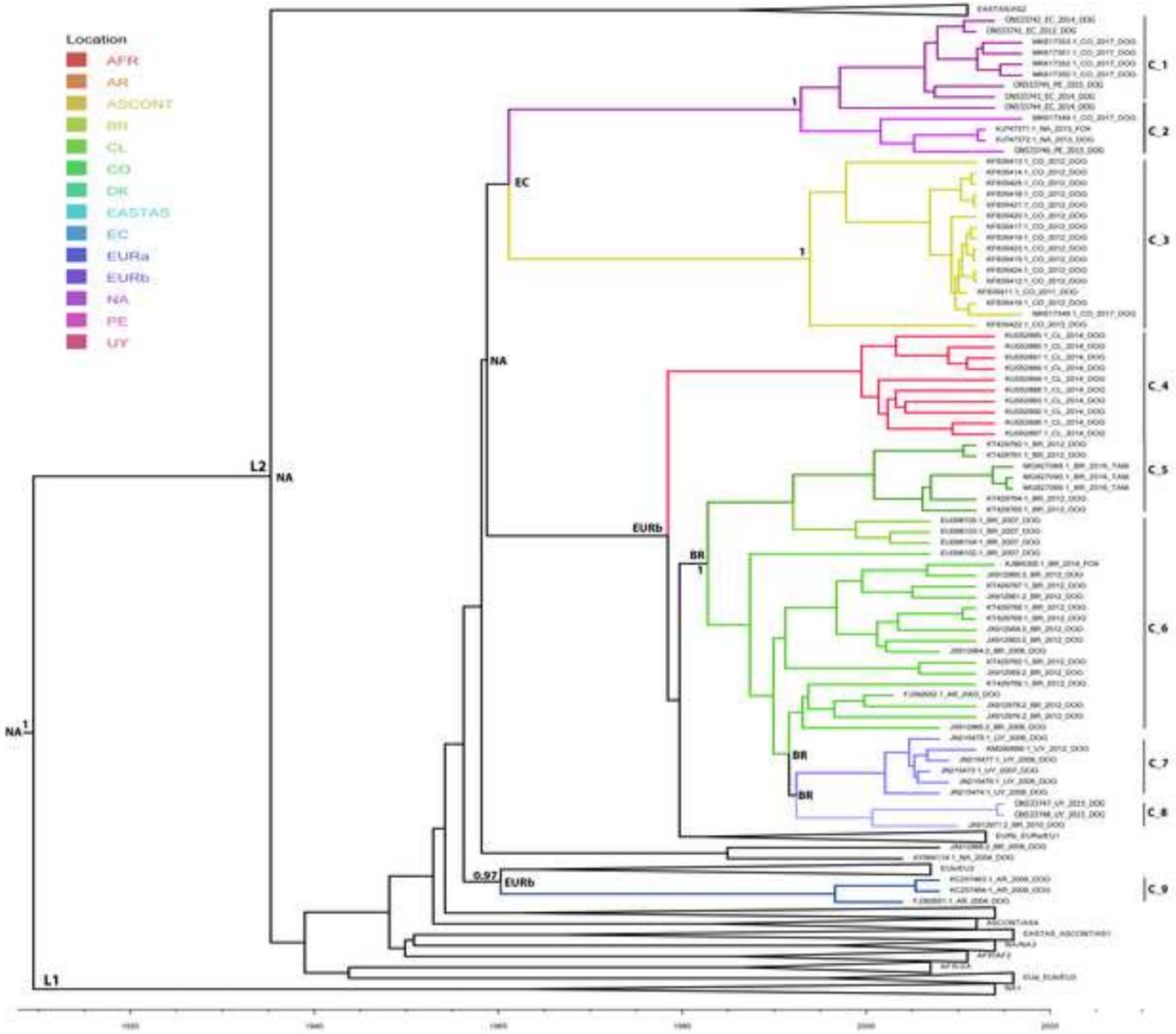

Figure 3

Click here to access/download;Figure;Fig 3.tif

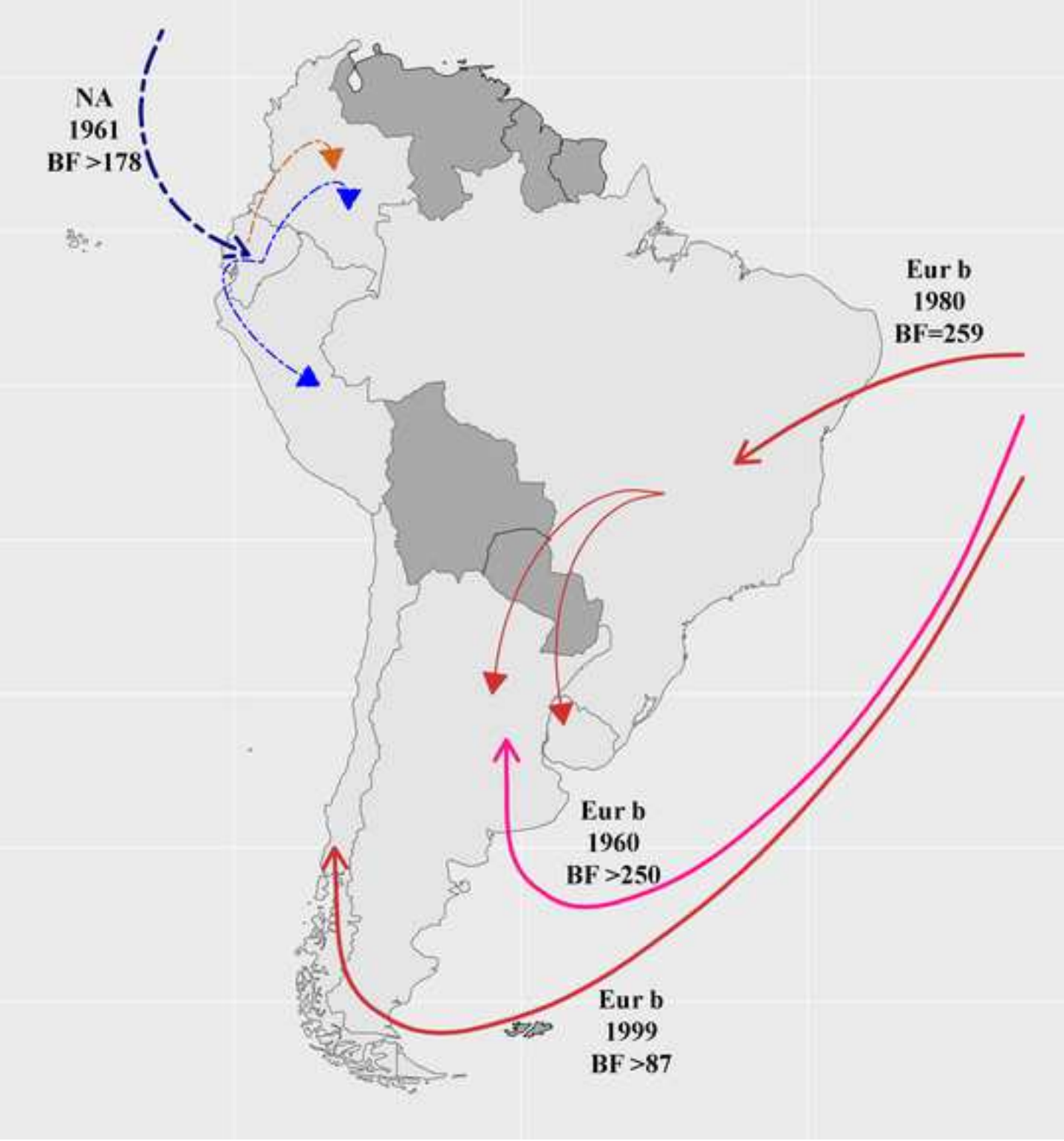

Table 1

Click here to access/download;Table;Table 1.xlsx

	22	38	41	51	78	83	89	103	128	145	146	151	157	172	177	189	192	193	195	197	198	212	213	214
C1		S																I		K	I			
C2					L											N		I		K	I			
C3	R		I					V	S														A	
C4				T				V																
C5				T	I	D		V						S					M			K		
C6				T				V																
C7				T				V			V													
C8				T				V		R			V		R									
C9				T			M	V				S				N	A		I					A

215	217	218	235	238	241	244	261	263	281	287	291	302	303	309	314	315	327	329	333	335	337	339	341	345
			I		E						M					V			V				L	
			M													V	S					D	L	
							S		R						S	V					I			V
	T				E			T																
	T	T		D										S										
	T												S											
	T											I								V				
G	T				E			I		M			S					N						
						E																		

349	353	358	360	416	424	427	430	435	443	445	459	487	488	515	506	522	523	530	542	544	580	582
	I						I					G				V						
	I						I									V		D				
	L					H			S	S			R		I	T		N		S		K
S	L				T			N		S				I								
										S								S	N			
										S	I											
				N						S	I						S				Q	
			I							S	I										Q	
		K	I									G			A			D				

Figure 2 S


Click here to access/download
Supplementary Material
Figure 2 Suppl.tif

Figure 1 S


Click here to access/download
Supplementary Material
Fig 1 Suppl.tif

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: