Molecular Phylogenetics and Evolution xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Molecular phylogeography of canine distemper virus: Geographic origin and global spreading ☆

Yanina Panzera [a,*], Nicolás Sarute [a], Gregorio Iraola [a,b], Martín Hernández [a], Ruben Pérez [a]

[a] Sección Genética Evolutiva, Instituto de Biología, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay
[b] Unidad de Bioinformática, Institut Pasteur Montevideo, Montevideo, Uruguay

ARTICLE INFO

Article history:
Received 27 December 2014
Revised 9 May 2015
Accepted 23 June 2015
Available online xxxx

Keywords:
Paramyxovirus
CDV
Phylogeographic analysis
Hemagglutinin

ABSTRACT

Canine distemper virus (CDV) (*Paramyxoviridae-Morbillivirus*) is a worldwide spread virus causing a fatal systemic disease in a broad range of carnivore hosts. In this study we performed Bayesian inferences using 208 full-length hemagglutinin gene nucleotide sequences isolated in 16 countries during 37 years (1975–2011). The estimated time to the most recent common ancestor suggested that current CDV strains emerged in the United States in the 1880s. This ancestor diversified through time into two ancestral clades, the current America 1 lineage that recently spread to Asia, and other ancestral clade that diversified and spread worldwide to originate the remaining eight lineages characterized to date. The spreading of CDV was characterized by several migratory events with posterior local differentiation, and expansion of the virus host range. A significant genetic flow between domestic and wildlife hosts are displayed; being domestic hosts the main viral reservoirs worldwide. This study is an extensive and integrative description of spatio/temporal population dynamics of CDV lineages that provides a novel evolutionary paradigm origin and dissemination of the current strains of the virus.

1. Introduction

Canine distemper virus (CDV) is a member of the *Morbillivirus* genus within the *Paramyxoviridae* family. This global virus is the cause for canine distemper (CD), a severe systemic disease affecting a broad range of terrestrial and marine carnivores, and non-hominid primates (Appel, 1987; Deem et al., 2000; Mamaev et al., 1995; Sakai et al., 2013; Qiu et al., 2011). CDV was isolated in 1905 by Henri Carré (Carré, 1905) and for decades was responsible for large numbers of animal deaths worldwide (Appel and Summers, 1999). By the 1950s, the development of attenuated vaccines reduced considerably the mortality rates and partially controlled the disease (Haig, 1956; Rockborn, 1959). However, in the last years several outbreaks of CD were reported in dogs and wildlife hosts (Harder et al., 1996; Lednicky et al., 2004; Martella et al., 2010; Munson et al., 2008). The nature of the CDV genome (ssRNA) implies high mutation rates mainly to the error prone viral RNA polymerase, which determines the range of genetic variation upon which natural selection can act (Lauring and Andino, 2010). The CDV genome encodes six structural proteins being the hemagglutinin (H) and the fusion (F) glycoproteins involved in the cell attachment and fusion processes (Lamb and Parks, 2007). The H protein attaches to the signaling lymphocyte activation molecule (SLAM/CD150) cellular receptor, facilitating the fusion and virus entry into the host cell (von Messling et al., 2001). Variations in the H protein residues have been proposed as involved in cell tropism, and are associated with host shift and adaptability (McCarthy et al., 2007; Nikolin et al., 2011; von Messling et al., 2003). On the basis of the H protein genetic variability and its phylogenetic relationships, CDV is classified into genetic lineages (Bolt et al., 1997; Martella et al., 2006). The criterion for lineage assignment states that two strains belong to the same lineage if they cluster together in the phylogenetic tree, and show an H amino acid divergence less than 4% (Martella et al., 2006). At least, nine well-defined CDV lineages had been identified worldwide, denoted America 1 (NA1), America 2 (NA2), Europe 1/South America 1 (EU1/SA1), Europe 2/Europe-wildlife (EU2), Europe 3/Arctic like (EU3), Asia 1 (AS1), Asia 2 (AS2), South Africa (ZA) and South America 2 (SA2) (An et al., 2008; Martella et al., 2006; Panzera et al., 2012; Woma et al., 2009). It is assumed that most CDV lineages follow a geographical pattern; however the geographic origin and evolutionary history of the virus still remains uncertain.

In this study we reconstructed the spatio/temporal dynamics of the CDV employing a Bayesian framework that included 208 H

☆ This paper was edited by the Associate Editor C. Nesbo.

* Corresponding author. Fax: +598 2525 86 17.
E-mail addresses: ypanzera@fcien.edu.uy, yaninapanzera@gmail.com (Y. Panzera).

http://dx.doi.org/10.1016/j.ympev.2015.06.015
1055-7903/

nucleotide sequences of the nine lineages sampled during a time frame of 37 years. Our findings revealed the time to the most recent common ancestor, migration routes and viral population dynamics of current CDV lineages, providing novel information about the evolution and global distribution of this relevant virus.

2. Materials and methods

2.1. Sequence dataset

Complete sequences of CDV hemagglutinin gene (1824 bp) and associated meta-information indicating year of isolation, country and host, were retrieved from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov). The dataset consisted of 208 sequences isolated from 1975 to 2011 and belonging to 16 countries: Argentina, Austria, Brazil, China, Germany, Greenland, Hungary, Italy, Japan, Portugal, South Africa, South Korea, Spain, Taiwan, United States and Uruguay.

Sequences were aligned using MAFFT (Katoh et al., 2002), and the best-fit nucleotide substitution model (TN93 + Γ) was selected under the Akaike and Bayesian information criteria as implemented in jModelTest (Posada, 2008). Likelihood mapping analyses were performed using the TREE-PUZZLE program by analyzing 10,000 random quartets (Schmidt et al., 2002; Schmidt and von Haeseler, 2007).

Recombination events between sequences were analyzed using the Phi test implemented in SplitsTree4 program (Huson and Bryant, 2006). A Phi test value of $p < 0.05$ was considered statistical evidence for recombination in the dataset.

2.2. Spatio/temporal dynamics of viral population

The evolutionary rate (substitutions per site per year), the time to the most recent common ancestor (tMRCA), the geographical origin and the global spatial dynamics of major CDV clades were jointly inferred using a Bayesian Markov chain Monte Carlo approach implemented in the BEAUti/BEAST package v1.7.5 (Drummond et al., 2012).

Preliminary analyses were performed with an uncorrelated log-normal relaxed clock to test if a strict molecular clock can be rejected for the dataset. As the relaxed clock standard deviation (ucld.stdev parameter) was greater than 1 with a frequency histogram not abutting 0, a strict molecular clock was preferred for subsequent runs. In the same way, a Bayesian skyline plot revealed a constant population size. Therefore, subsequent runs were performed using a constant population prior.

Spatio/temporal dynamics were inferred implementing a discrete phylogeographic model and a Bayesian stochastic search variable selection procedure was employed to identify the most parsimonious description of the diffusion process (Lemey et al., 2009). Four independent Markov chain Monte Carlo processes were run for 1×10^8 generations, adequate chain mixing was ensured by effective sample size values greater than 200 for each sampled parameter using TRACER v1.4 (Drummond and Rambaut, 2007). The first 10% steps were removed as burn-in for combined chains.

Maximum Clade Credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator, visualized and annotated with FigTree v1.4.0 (available at http://tree.bio.ed.ac.uk/software). Spatial dynamics were summarized using SPREAD (Bielejec et al., 2011) and visualized with Google Earth (Lisle, 2006).

2.3. Ancestral states reconstruction

The states "wildlife" or "domestic dog" were reconstructed for all internal nodes in the posterior set of phylogenies implementing a symmetric continuous-time Markov Chain (CTMC) approach in BEAST v1.7.5 (Drummond et al., 2012). Probabilities higher than 0.7 were considered for determining host ancestor.

3. Results and discussion

3.1. Spatio/temporal dynamics of CDV viral population

The H gene showed an appropriate tree-like phylogenetic signal (97%) in the likelihood mapping analysis, pointing out the suitability of this dataset to perform robust and consistent phylogenetic and molecular clock inferences (Strimmer and von Haeseler, 1997). The estimated tMRCA was 125 years from the most recent analyzed strain (2011), corresponding to 1886 with a 95% highest posterior density (HPD) interval ranging from 1858 to 1913 (Table 1, Fig. 1). A previous study estimated a tMRCA for CDV of 58 years before the most recent sequences (2001) with a HPD ranging from 1894 to 1974 (Pomeroy et al., 2008). This estimation was based on the analysis of 35 H gene sequences collected during 19 years (1982–2001) from Japan, the United States, and Western Europe. This dataset is smaller and more geographic and temporally restricted than our dataset (208 H sequences from 16 countries during 1975–2011), which explains the differences in the tMRCAs. Historical records based on dog's clinical symptoms dated back canine distemper to 1761 in Peru (Howell, 1965; Blancou, 2004). These reports could have been referred to a disease with similar symptomatology than distemper, but caused by different etiological agents, leading to an incorrect diagnosis. In fact, it had been proposed that the disease was caused by bacteria such as *Pasteurella canis* and *Bacillus bronchisiptecus* (Bresalier and Worboys, 2014). We cannot rule out the existence of CDV variants before the emergence of current viruses (~1886), but our findings suggest that these "old" strains are now extinct or remain undetected in particular wildlife host.

According to the posterior probabilities for geographic location in the MCC tree root, the most recent common ancestor of current CDV strains emerged in the United States. This is the first molecular inference concerning the geographic origin of CDV. This American CDV ancestor diversified and split into two major ancestral clades, which spread worldwide (Fig. 1). The first clade (1923 year node) originated the eight current lineages: EU3, ZA, AS2, SA2, EU2, NA2, EU1/SA1 and AS1. The second ancestral clade diversified into the current NA1 lineage, which recently has spread to Asia (China) (1978 year node). The CDV ancestor spread through major migration routes as shown in the MCC tree, supported by a posterior probability close to 1 (Table 1, Fig. 1). To provide a spatial projection of CDV dissemination we employed the SPREAD software to generate suitable files that are visualized using Google Earth (Fig. 2).

The MCC tree allowed us to infer the most plausible direction of the migration events concerning the current CDV lineages. The first migration event occurred in 1923 (95% HPD 1904–1942) and reached Greenland, and from there to Europe (Italy) over 44 years later (~1967) (Table 1, Figs. 1 and 2A). The Italian strains exhibited a wide spreading capability and reached other European countries (Hungary and Austria), Asia (China), and the United States (Table 1, Figs. 1 and 2B–C). The first strains within this cluster date back to late 1980s (Blixenkrone-Moller et al., 1992; Bolt et al., 1997) and were later classified within the EU3/Arctic-like lineage (Martella et al., 2006; Demeter et al., 2007; Zhao et al., 2010). Accordingly, the unrevealed origin of the unusual Italian Arctic strains, raised

Please cite this article in press as: Panzera, Y., et al. Molecular phylogeography of canine distemper virus: Geographic origin and global spreading. Mol. Phylogenet. Evol. (2015), http://dx.doi.org/10.1016/j.ympev.2015.06.015