Infection, Genetics and Evolution 58 (2018) 279–289

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Interplay between co-divergence and cross-species transmission in the evolutionary history of bat coronaviruses

Stefania Leopardi[a,*], Edward C. Holmes[b], Michele Gastaldelli[a], Luca Tassoni[a], Pamela Priori[c], Dino Scaravelli[c], Gianpiero Zamperin[a], Paola De Benedictis[a]

[a] National Reference Centre, OIE Collaborating Centre for Diseases at the Animal-Human Interface, Istituto Zooprofilattico Sperimentale delle Venezie, Viale dell'Universita' 10, Legnaro, Padova 35020, Italy
[b] Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, Australia
[c] S.T.E.R.N.A., Forlì, Italy

ARTICLE INFO

Keywords:
Virus
Bats
Evolution
Coronaviruses
Phylogeny co-divergence
Cross-species transmission

ABSTRACT

Coronaviruses (CoVs) have been documented in almost every species of bat sampled. Bat CoVs exhibit both extensive genetic diversity and a broad geographic range, indicative of a long-standing host association. Despite this, the respective roles of long-term virus-host co-divergence and cross-species transmission (host-jumping) in the evolution of bat coronaviruses are unclear. Using a phylogenetic approach we provide evidence that CoV diversity in bats is shaped by both species richness and their geographical distribution, and that CoVs exhibit clustering at the level of bat genera, with these genus-specific clusters largely associated with distinct CoV species. Co-phylogenetic analyses revealed that cross-species transmission has been more common than co-divergence across coronavirus evolution as a whole, and that cross-species transmission events were more likely between sympatric bat hosts. Notably, however, an analysis of the CoV RNA polymerase phylogeny suggested that many such host-jumps likely resulted in short-term spill-over infections, with little evidence for sustained onward transmission in new co-roosting host species.

1. Introduction

Since the isolation of Hendra virus from pteropid bats in 2000 (Halpin et al., 2000), bats have been implicated in the emergence of a number of other human infectious diseases, most notably Nipah, Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS) and Ebola (Calisher et al., 2006; Moratelli and Calisher, 2015). In turn, the notion that these viral diseases likely have their ultimate ancestry in bats triggered a major increase in the sampling of bat viruses, leading to the progressive uncovering of a diverse bat virome and further fueling the idea that these animals are major reservoirs of emerging pathogens (Moratelli and Calisher, 2015; Young and Olival, 2016).

Successful cross-species transmission and emergence depends on a variety of biological, ecological and epidemiological factors. Although RNA viruses commonly jump species boundaries, in part reflecting their ability to rapidly generate important adaptive variation (Geoghegan et al., 2017; Holmes, 2009; Woolhouse and Gowtage-Sequeria, 2005), coronaviruses (CoVs) seem to exhibit a strong zoonotic potential and demonstrated by the emergence SARS-CoV and MERS-CoV in humans in 2002 and 2012, respectively (Graham et al., 2013). Coronaviruses are single-strand RNA viruses of the order *Nidovirales* that are classified in four genera: *Alpha-*, *Beta-*, *Gamma* and *Deltacoronavirus*. Among these, gamma and delta CoVs are largely associated with avian hosts, while alpha and beta CoVs include several pathogens of humans and domestic animals, and whose emergence is likely associated with cross-species transmission events (Drexler et al., 2014).

Both SARS-CoV and MERS-CoV belong to the genus *Betacoronavirus* and are associated with severe lower respiratory tract infection characterized by mortality rates of 10% and 35%, respectively (Hu et al., 2015). The SARS pandemic was promptly controlled through an unprecedented global containment effort and the virus has not been reported in humans since May 2004 (Graham et al., 2013). Despite this rapid eradication, SARS-CoV caused almost 800 deaths in 27 countries, with sustained outbreaks in 18 countries on three continents (WHO). There is increasing evidence that rhinolophid bats act as natural reservoirs for SARS-related CoVs, with direct spill-over to non-flying mammals. For example, like the SARS coronavirus, some bat CoVs are

* Corresponding author.
E-mail address: Sleopardi@izsvenezie.it (S. Leopardi).

https://doi.org/10.1016/j.meegid.2018.01.012
Received 25 September 2017; Received in revised form 10 January 2018; Accepted 12 January 2018
Available online 30 January 2018

able to utilize the angiotensin converting enzyme 2 (ACE2) as a cell receptor (Ge et al., 2014; Menachery et al., 2016; Yang et al., 2016; Zeng et al., 2016). Conversely, the role of bats in the epidemiology of MERS-CoV is less well understood as the human viruses are clearly mostly related to those viruses found in dromedary camels (Sabir et al., 2016). Indeed, although related viruses have been found in bats, these are divergent in their spike sequences and seem to be inefficient in the use of human dipeptidyl peptidase 4 (DPP4) as cell a receptor (Anthony et al., 2017a; Reusken et al., 2016; Yang et al., 2014). The MERS epidemic is ongoing in the Middle East and travel-associated cases have been reported in 27 countries worldwide (WHO, 2017). Finally, *Alphacoronavirus* 229E and NL63, which cause a mild influenza-like syndrome in humans, share a common ancestor with viruses sampled from the bat genus *Hipposideros* and *Triaenops*, respectively (Corman et al., 2015, 2016; Tao et al., 2017).

Bats are known to harbor high levels of CoV diversity with impressive geographical range and prevalence in almost every species investigated, again supporting the idea that they have played a major role in CoV evolution (Anthony et al., 2017b; Drexler et al., 2014). In addition, bat CoVs are phylogenetically interspersed with those associated with other mammals, including humans and domestic species, compatible with the idea that bats are an important genetic reservoir (Tao et al., 2017; Woo et al., 2012). The long-term evolutionary interactions between bats and coronaviruses is also supported by phylogenetic evidence that CoVs exhibit some species- and genus-specific tropism (Cui et al., 2007; Vijaykrishna et al., 2007), and that phylogenetically related viruses are found in related bat species independent of sampling location. In contrast, that CoVs are not always shared among bat species that co-roost suggests that there are some barriers to cross-species transmission (Anthony et al., 2013; Corman et al., 2013; Cui et al., 2007; Drexler et al., 2010; Smith et al., 2016; Tang et al., 2006).

Because of the topological similarity between the phylogenetic trees of CoVs and their mammalian hosts, it has been suggested that the diversity of CoVs largely reflects the long-term co-divergence between bats and CoVs (Cui et al., 2007). However, recent studies on specific bat taxa from particular locations suggests that the role of virus-host co-divergence in the evolutionary history of CoVs may have been overestimated relative to other events including host-jumping (Anthony et al., 2017b; Lin et al., 2017; Tao et al., 2017). Indeed, as well as strict virus-host co-divergence, topological congruence could also arise from preferential host switching, in which viruses most often successfully jump from closely related hosts (De Vienne et al., 2013). The analysis of the long-term evolutionary history of bat CoVs is also complicated by frequent multiple substitution at deep evolutionary distances that prevents the accurate estimation of divergence times (Wertheim et al., 2013).

To obtain a more complete picture of the evolutionary history of alpha and beta coronaviruses in their natural hosts, which is essential for understanding the fundamental mechanisms of virus emergence, we performed a statistical analysis of co-phylogenetic relationships on a large data set of mammalian CoVs. Not only did this suggest that cross-species transmission has played a major role in the evolution of alpha and beta CoVs in bats, but also that differences in bat host ecology, biology and geographical range have a strong impact on coronavirus evolution.

2. Materials and methods

2.1. Source and selection of CoV and host sequences

We retrieved all bat CoV sequences representing the partial ORF1b that encodes the RNA-dependent RNA polymerase (RdRp) available on GenBank (as of March 2017). These were combined with 109 CoV sequences from other mammals. Two gamma CoVs were used to root the phylogeny. Only sequences > 350 bp in length and associated with a bat genus for which at least two sequences were available were retained. Unique sequences associated with a particular species were included, but solely used for analyses based on the host genus. Similarly, we retrieved CoV sequences encoding the spike (S) protein, including those from bats and 46 CoV sequences sampled from other mammals. For each CoV sequence we recorded the collection date, location and host (genus and species) based on information available in GenBank and/or in the associated literature. CoV sequences for which the sampling location and/or host genus were unavailable were discarded. Sampling locations were retrieved at the country level, and were categorized according to their large-scale geographic area of sampling: Europe, Africa, North America, Latin America (Central and South America), Asia, South East Asia, and Australia.

The most comprehensive CoV data set encoding the RdRp (denoted "RdRp_CoV_1") comprised 541 CoV sequences from bats plus 111 sequences from other mammalian genera, including three randomly chosen representatives of known monophyletic groups of CoVs as well as all unclassified mammalian sequences. This data set was used for the phylogenetic and host clustering analyses (see below). CoV sequences encoding the spike protein comprised a data set, denoted "spike_CoV", which included 199 sequences from bats plus 46 CoV sequences from other mammals.

We also constructed sub-sampled data sets from the comprehensive RdRp_CoV_1 data set based on results from the genus-specific clustering (see below) to minimize errors associated with the non-independence of data. Specifically, reduced data sets CoV (n = 58), CoVα (n = 34) and CoVβ (n = 24) included the longest sequence for each genus-specific cluster. These reduced data sets were used in the co-phylogenetic analysis (see below).

To help assess the validity of our results we constructed a second group of data sets ("data sets_2") which only included sequences from bat hosts whose species was confirmed genetically (and hence more confidently), thereby removing any error due to host misclassification. These data sets were termed RdRp_CoV_2 (n = 42 sequences), which was used for phylogenetic and host clustering analyses, and CoV_2 (n = 11), CoV_2α (n = 8), CoV_2β (n = 3) and host_2, used in the co-phylogenetic analysis (see below).

Host sequences targeting the full mitochondrial cytochrome b (cytb) gene were also retrieved from GenBank and visually inspected to ensure that they agreed with previously published bat phylogenies. The host data set (denoted "host") included one cytb gene sequence for each genus associated with the CoV data set.

2.2. Phylogenetic analysis

All sequences were aligned with MAFFT utilizing the L-INS-i routine (Katoh et al., 2002), manually adjusted. Longer sequences were then trimmed to 935 bp (RdRp) and 1440 bp (spike protein) using MEGA6 (Tamura et al., 2013). Sequence alignments utilized nucleotide sequences for the host, CoV, CoVα and CoVβ data sets and all the data sets_2, amino acid sequences for the spike_CoV data set, and both nucleotide and amino acid sequences for the comprehensive data set RdRp_CoV_1. Best-fit models of nucleotide and amino acid substitution for each data set were determined using MEGA6 (Tamura et al., 2013). Pairwise genetic distances among nucleotide and amino acid sequences were computed using the Maximum Composite Likelihood method in MEGA6.

Maximum likelihood nucleotide phylogenetic trees were inferred using PhyML (version 3.0), employing the GTR + Γ_4 substitution model, a heuristic SPR branch-swapping algorithm and 1000 bootstrap replicates (Dereeper et al., 2008). Similarly, amino acid ML trees were estimated for data sets RdRp_CoV_1 and spike_CoV using RAXML (version 8.1.17) assuming the LG + Γ_4 and the LG + Γ_4 + I models of amino acid substitution, respectively, and 1000 bootstrap replicates.

Topological congruence between the RdRp and spike-based amino acid trees was determined based on the phylogenies of RdRp and spike