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

sequences derived from the same host taxa. To implement the phylogeny trait methods which require a posterior distribution of phylogenetic trees (see below), we inferred non-clock Bayesian trees for data set RdRp_CoV_1 using MrBayes v3.2.4 (Ronquist and Huelsenbeck, 2003) and assuming the GTR + Γ_4 nucleotide substitution model. This analysis was run for 70 million generations (25% discarded as burn-in) and sampled every 500 generations. The resultant tree was then edited using iTOL (Letunic and Bork, 2016).

Finally, the degree of temporal signal (i.e. clock-like structure) in these CoV data was explored by plotting root-to-tip genetic distances on the ML trees against year of sampling using the method implemented in TempEst (Rambaut et al., 2016).

2.3. Assessing the extent of clustering by bat hosts

We first investigated whether coronaviruses showed significant clustering by bat host genus/species or sampling location using the association index (AI), parsimony score (PS) and maximum monophyletic clade (MC) phylogeny-trait statistics available in the BaTS package (Parker et al., 2008). This analysis compared the posterior distribution of trees for data set RdRp_CoV_1 described above to a null distribution of 100 trait-randomized trees, with bat large-scale geographic area and country of sampling, bat genus, and bat species assigned as the character states of interest. The extent of clustering associated with the characters “host genus” and “host species” was investigated also in the reduced data set RdRp_CoV_2. In addition, we used the phylogenetic tree to identify bat RdRp genus-specific clusters of CoVs, defined as a minimum of two monophyletic RdRp sequences associated with the same bat genus, supported by Bayesian posterior probabilities > 0.90 and differing no > 4.8% and 5.1% at the amino acid level for the genera *Alphacoronavirus* and *Betacoronavirus*, respectively (see below).

The taxonomy of CoV clusters was determined based on pairwise amino acid distances calculated using 816 bp of the RdRp, when available, and reflecting the RdRp Group Units (RGU) defined previously for CoVs (Drexler et al., 2014). Following these criteria, we assigned distinctive RGUs to sequences differing by at least 4.8% and 5.1% at the amino acid level for the genera *Alphacoronavirus* and *Betacoronavirus*, respectively (Drexler et al., 2014).

2.4. Analysis of virus-host co-divergence

To assess the extent of virus-host co-divergence in the data we reconciled the CoV and host phylogenies using Jane 4.0 (Conow et al., 2010), which infers the nature and frequency of different evolutionary scenarios by finding the reconciliation with the lowest total cost. Accordingly, we assigned event costs = 1 for virus lineage duplication, host shift, and virus loss, or failure to diverge following host speciation, and costs of both 0 and 1 for co-divergence. Jane was run for 45 generations (G) with a set size of 23 (S). This analysis used the tree topologies based on RdRp and cytb for coronaviruses and their hosts, respectively. In addition, the congruence between the phylogeny of bat genera and their CoV clusters was depicted graphically using the tanglegram function of the DECIPHER package (Wright, 2016) from the *R* environment. To exclude the impact of possible host misclassification, this analysis was also run for datasets_2 in which hosts are assigned genetically.

2.5. Analyses of putative cross-species transmission events

Sequences included within bat RdRp genus-specific clusters of CoVs but associated with different host genera were considered as likely cross-species transmission events. For each such event we determined whether cross-species transmission was associated with the following variables: host taxonomy (genus, family and superfamily) of the donor and recipient hosts, CoV lineage, sampling location, sampling year, sampling location, and sampling year and location of the most closely related sequence.

To tentatively determine whether there may have been a sustained chain of CoV transmission in the new host species (as opposed to transient spill-overs), we assessed whether sequences associated with cross-species transmission events were significantly divergent from the donor cluster. Accordingly, for each putative cross-species transmission event, we compared the median nucleotide distance among sequences from the donor host with the median nucleotide distances between the donor sequences and those from the novel (i.e. recipient) host using a one-tailed Wilcoxon rank sum test. Genetic distances were again calculated using MEGA6 (Tamura et al., 2013) as described above. Although this analysis was based on the RdRp, spike protein sequences were also used when available.

2.6. Statistical analyses

The Spearman coefficient (rs) was used to determine the strength of the correlation between CoV diversity, expressed as number of detected CoV clusters, and the species richness or geographical range of CoV samples for each host genus. To test the influence of sampling effort on CoV diversity, the same correlation coefficient was determined between CoV diversity and the number of GenBank submissions and the total number of sequences per genus. Absolute values were (arbitrarily) interpreted as follows: 0.00–0.39 “weak” correlation, 0.40–0.59 “moderate” correlation, 0.60–0.79 “strong” correlation, and 0.80–1.0 “very strong” correlation. Coefficients were considered significant for p-value < 0.05.

3. Results

3.1. CoV sequences analyzed

A total of 650 CoV RdRp sequences were analyzed in this study. Among these, 541 were from bats (307 *Alphacoronavirus* and 234 *Betacoronavirus*), representing most of the CoV diversity currently described in bats, and 111 from other mammalian species (57 *Alphacoronavirus*, 50 *Betacoronavirus* and two *Gammacoronavirus*, with the latter used as outgroups) (data set “RdRp_CoV_1”). Sequence lengths generally ranged from 350 bp to 816 bp (n = 352), although a number (n = 298) were longer and trimmed to 935 bp prior to analysis (Table 1). Similarly, we analyzed 245 spike protein sequences, including 199 sequences from bats and 46 from other mammals. Sequence lengths ranged from 678 bp (including the receptor-binding domain RBD) and ~4000 bp corresponding to the full length sequence. For 94 bat CoVs both the RdRp and the spike protein sequences were available (Table 1). Although bat sequences obtained were identified worldwide, most came from Asia (n = 252 in five countries), Africa (n = 170 in six countries) and Europe (n = 116 in nine countries) (Fig. 1).

Bat-derived CoV sequences (either RdRp or spike) were sampled from 82 different species belonging to 25 genera (Table 1). Importantly, information on the criteria used for the classification of bat species in the data set was often lacking (69.5%); in some cases, classification was based on morphology (55.4%), and to a lesser extent on genetic identification (17.1%). Of note, 36 of the bats under study (representing 12 genera) are considered cryptic species as they are morphologically indistinguishable from other sympatric species (Table S1). In 18 cases, the host classification provided with the CoV sequence identified the host genus only, while in one case (*H. caffer_ruber*) the exact species was not defined (Pfefferle et al., 2009). Therefore, to help assess the robustness of our results, we performed a second analysis on 42 sequences for which the bat host species was confirmed genetically (data set RdRp_CoV_2), among which 26 were *Alphacoronavirus* and 16 were *Betacoronavirus*.

Table 1
Host association, length and classification of RdRp and spike protein CoV sequences used in this study.

Host superfamily	Host family	Host genus	Host species	RdRp (α; β)	RdRp > 816 bp	Spike (α; β)	RdRp + Spike (α; β)	RdRp clusters (α; β)
Emballonuroidea	Emballonuridae	*Taphozous*	1	2 (1; 1)	0			0
Molossoidea	Molossidae	*Chaerephon*	2	3 (2; 1)	3	3 (2; 1)	3 (2; 1)	0
Molossoidea	Molossidae	*Molossus*	1	2 (2; 0)	0			1 (1; 0)
Noctilionoidea	Phillostomidae	*Artibeus*	2	10 (9; 1)	7			2 (2; 0)
Noctilionoidea	Phillostomidae	*Carollia*	1	11 (10; 1)	8			1 (1; 0)
Noctilionoidea	Phillostomidae	*Sturnira*	1	3 (3; 0)	0			1 (1; 0)
Pteropodidae	Pteropodidae	*Cynopterus*	2	13 (0; 13)	0			1 (0; 1)
Pteropodidae	Pteropodidae	*Dobsonia*	1	3 (0; 3)	0			1 (0; 1)
Pteropodidae	Pteropodidae	*Eidolon*	1	60 (1; 59)	1	1 (0;1)	1 (0;1)	1 (0; 1)
Pteropodidae	Pteropodidae	*Eonycteris*	1	4 (0; 4)	0			0
Pteropodidae	Pteropodidae	*Epomophorus*	1	5 (2; 3)	0			1 (0; 1)
Pteropodidae	Pteropodidae	*Pteropus*	1	6 (0; 6)	0			1 (0; 1)
Pteropodidae	Pteropodidae	*Rousettus*	2	18 (3; 15)	14	13 (2; 11)	12 (2; 10)	2 (0; 2)
Rhinolophoidea	Hipposideridae	*Hipposideros*	10	65 (55; 10)	45	21 (19; 2)	12 (10;2)	3 (2; 1)
Rhinolophoidea	Hipposideridae	*Triaenops*	1	4 (4; 0)	4			2 (2; 0)
Rhinolophoidea	Nycteridae	*Nycteris*	1	3 (0; 3)	3	1 (0; 1)	1 (0; 1)	1 (0; 1)
Rhinolophoidea	Rhinolophidae	*Rhinolophus*	13	87 (24; 63)	61	74 (8; 66)	29 (4; 25)	3 (2;1)
Vespertilionoidea	Miniopteridae	*Miniopterus*	7	84 (84; 0)	34	30 (30; 0)	14 (14;0)	2 (2; 0)
Vespertilionoidea	Vespertilionidae	*Eptesicus*	3	14 (3; 11)	1			2 (1; 1)
Vespertilionoidea	Vespertilionidae	*Murina*	1	6 (6; 0)	6	2 (1; 1)	0	1 (1; 0)
Vespertilionoidea	Vespertilionidae	*Myotis*	16	80 (77; 3)	8	6 (6; 0)	5 (5; 0)	8 (8; 0)
Vespertilionoidea	Vespertilionidae	*Nyctalus*	3	7 (6; 1)	2	1 (1;0)	0	1 (1; 0)
Vespertilionoidea	Vespertilionidae	*Pipistrellus*	5	37 (9; 28)	20	22 (0; 22)	12 (0; 12)	6 (3; 3)
Vespertilionoidea	Vespertilionidae	*Scotophilus*	2	7 (4; 3)	1	3 (3; 0)	1 (1; 0)	1 (1; 0)
Vespertilionoidea	Vespertilionidae	*Tylonycteris*	1	7 (0; 7)	7	22 (0; 22)	4 (0; 4)	1 (0; 1)

3.2. Clustering of CoVs by host and sampling location

Phylogenetic analyses revealed a structuring of bat coronavirus diversity dependent on both their host taxa and the large-scale geographic area of sampling (Fig. 2). With the exception of CoV sequences associated with *Alphacoronavirus 1*, *Mink coronavirus 1*, *Human coronavirus HKU1* and *Betacoronavirus 1*, all other CoVs from non-flying mammals formed clades that either exhibited sister relationships with bat-associated viruses or were nested within them (Fig. 2). Regression analyses revealed no correlation between sampling times and root-to-tip genetic distances for both the RdRp and spike proteins ($R^2 = 0.0078$ and 0.0075 for the RdRp_CoV_1 and spike_CoV data sets, respectively) thereby precluding any molecular clock dating (Fig. S1A, B).

Phylogeny-trait analyses based on RdRp sequences using the BaTS

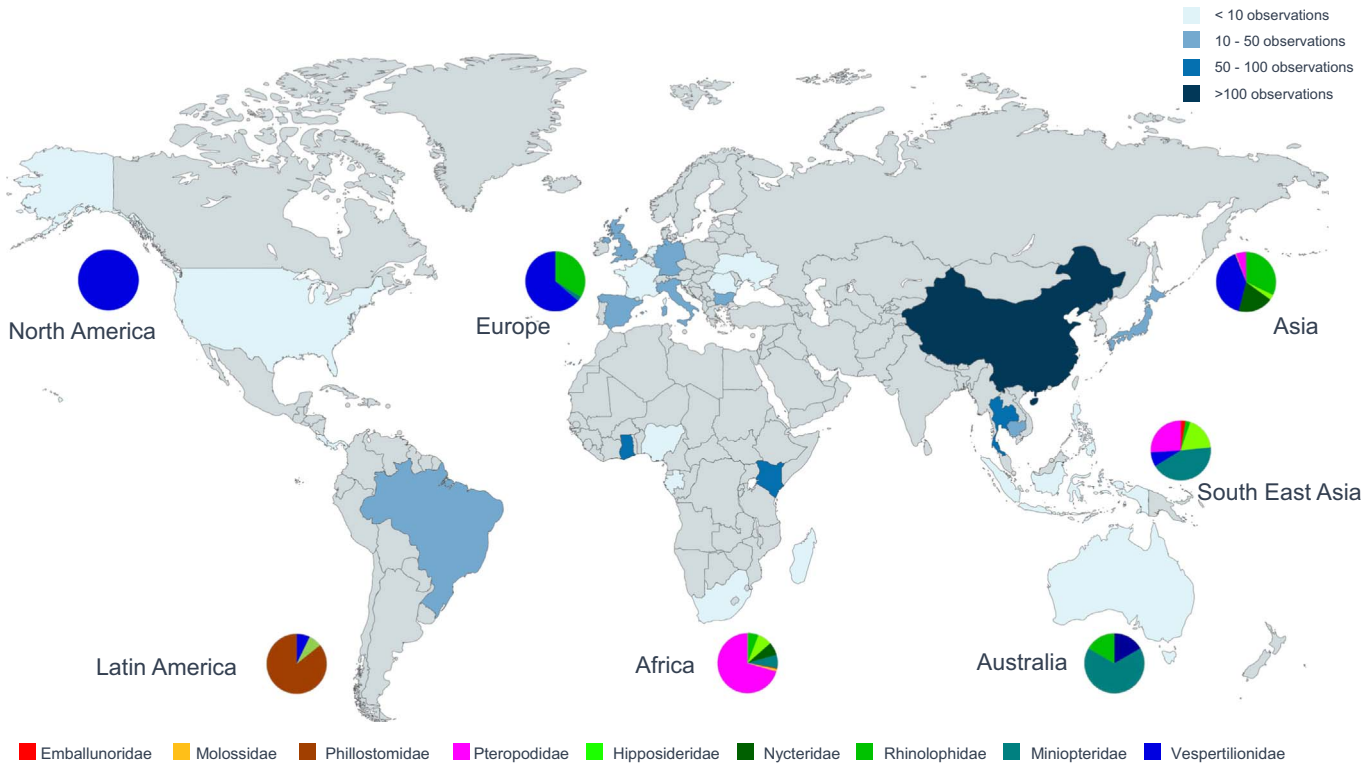

Fig. 1. Host association and geographical distribution of the CoV sequences analyzed here. Countries within large-scale geographical regions are colored according to the number of CoV analyzed. Pie charts indicate the host-association of the CoV sequences included within each geographical area, with the colors indicating the different families of bat hosts. The map was built using mapchart (https://mapchart.net).

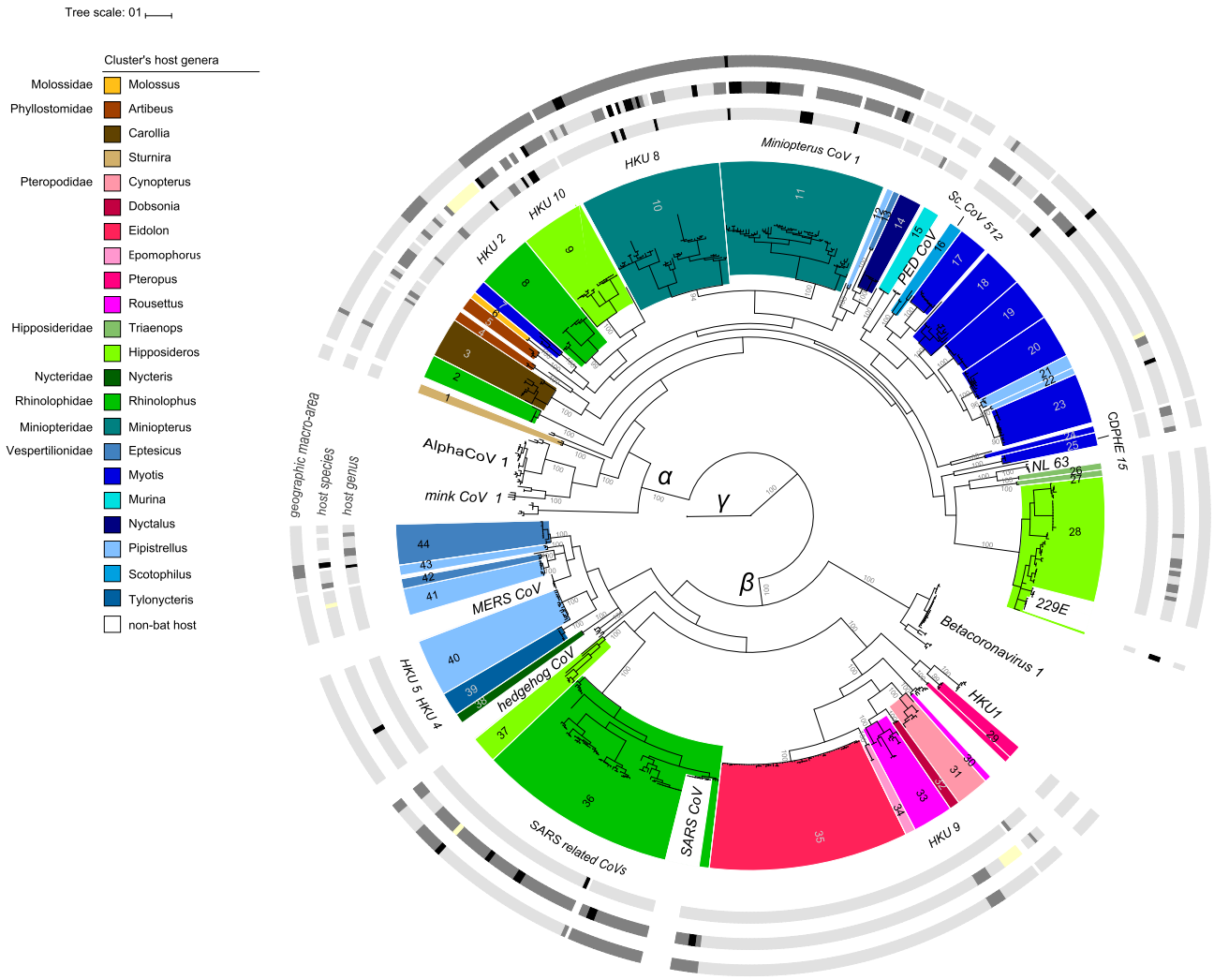

Fig. 2. Phylogenetic overview of CoV sequences analyzed here. The tree reflects a Bayesian analysis of 935 bp of the RdRp gene (data set RdRp_CoV_1), rooted using two sequences from gamma coronaviruses (GenBank accession numbers EF584911-2). Genus specific clusters identified in our study are colored based on the host genus, as indicated. Posterior probabilities > 0.90 supporting each cluster are shown. Branch lengths are scaled according to the number of substitutions per site. The three bars around the tree show the frequency within each cluster of (i) host genera, (ii) host species and (iii) sampling locations, from the innermost to the most exterior. Sequences showing characters with frequency < 10%, between 10 and 50%, and > 50% are colored black, grey and light grey, respectively. For the "host species" bar, only sequences belonging to the host genus characterizing the cluster (frequency > 50%) have been colored; sequences associated with hosts only characterized at the genus level are indicated in yellow. The ICTV classification of virus clusters is indicated when available. The figure was generated using iTOL.

program provided statistical support ($p < 0.01$) for the clustering of CoVs in all the traits investigated. The percentage of individual traits significantly (i.e. < 0.05) supporting CoV clustering was 100% for "large-scale geographic area of sampling", 88.46% for "bat genus", 77.4% for "country of sampling", and 75.4% for "bat species" (Table S2). Individual traits represented by unique sequences (e.g. single countries) always gave non-significant results and were excluded from the overall percentages given above. A similar trend was confirmed for sequences associated with genetically defined host species (data set RdRp_CoV_2), for which clustering was significant in 80% of host genera and 50% of host species (Table S3).

Our phylogenetic analyses identified 44 RdRp bat genus-specific clusters, 28 in the alpha CoVs (denoted C1α-C28α) and 16 in the beta CoVs (denoted C29β-C44β) (Fig. 2). As expected, sequences from NL63, CoV 229E and SARS CoV from other mammalian hosts fell within bat-associated clusters (between C26α and C27α, within C26α and within C36β, and associated with bats belonging to the genera *Trianops*, *Hipposideros* and *Rhinolophus*, respectively) (Fig. 2). Conversely, it was not possible to identify species-specific or geographically structured clusters across all bat genera. This was particularly evident for clusters of CoVs found in the bat genera *Miniopterus*, *Rhinolophus* and *Hipposideros* associated with different host species sampled from different geographical macro-areas (Fig. 2). That these results were not due to errors in host classification was confirmed by phylogenetic analyses of data set RdRp_CoV_2, which also revealed a clustering by host genus rather than host species (Fig. S2).

A single CoV phylogenetic cluster was associated with most bat genera. However, two or three clusters were observed in *Artibeus*, *Rousettus*, *Hipposideros*, *Triaenops*, *Rhinolophus*, *Eptesicus* and *Miniopterus*, and more than three clusters were present in *Myotis* and *Pipistrellus*. The bat genera *Pipistrellus*, *Eptesicus*, *Rhinolophus* and *Hipposideros* were associated with both alpha and beta CoVs, while all genera of fruit bats (the Pteropodidae) were found to only harbor beta-CoVs assigned to lineage D (Table 1, Fig. 2). The number of host species included within genus-specific clusters varied between one and 10, with the most observed in *Rhinolphus* (n = 10), *Miniopterus* (n = 7) and *Hipposideros* (n = 4) (Table 2).

A strong correlation was observed between the number of RdRp specific clusters described for each bat genus and either its species richness (rs = 0.69, p = 0.0001) or geographic distribution (rs = 0.67,

Table 2
Summary of CoV sequences included in RdRp genus-specific phylogenetic clusters. For each cluster, the table indicates the best represented host genus (> 50%), host family, number of host species, sampling location, length of the longest RdRp sequence and the presence of a corresponding cluster in the spike protein sequences.

Cluster	Host genus	Host family	Host species	Sampling location[a]	RdRp max length	Spike protein[b]
C1α	*Sturnira*	Phillostomidae	1	LAM	393 bp	
C2α	*Rhinolophus*	Rhinolophidae	2	AS, SEA	935 bp	X (AS)
C3α	*Carollia*	Phillostomidae	1	LAM	935 bp	
C4α	*Artibeus*		2	LAM	816 bp	
C5α	*Artibeus*		2	LAM	816 bp	
C6α	*Molossus*	Molossidae	1	LAM	393 bp	
C7α	*Myotis*	Vespertilionidae	4	AS, EU	935 bp	X (AS)
C8α	*Rhinolophus*	Rhinolophidae	3	AFR, EU	816 bp	
C9α	*Hipposideros*	Hipposideridae	4	AS, SEA	935 bp	X (AS)
C10α	*Miniopterus*	Miniopteridae	7	AS, AUS, EU, SEA	935 bp	X (AS)
C11α	*Miniopterus*		6	AFR, AS, SEA	935 bp	X (AFR, AS)
C12α	*Pipistrellus*	Vespertilionidae	1	EU	412 bp	
C13α	*Eptesicus*		1	AS	408 bp	
C14α	*Nyctalus*		2	EU	817 bp	
C15α	*Murina*		1	AS	935 bp	
C16α	*Schotophilus*		2	AS-SEA	935 bp	X (AS)
C17α	*Myotis*		2	AS	415 bp	
C18α	*Myotis*		3	AS, EU	935 bp	X (AS)
C19α	*Myotis*		1	EU	392 bp	
C20α	*Myotis*		1	EU	816 bp	
C21α	*Pipistrellus*		1	EU	403 bp	
C22α	*Pipistrellus*		1	EU	403 bp	
C23α	*Myotis*		2	EU	403 bp	
C24α	*Myotis*		2	LAM	393 bp	
C25α	*Myotis*		2	NAM	935 bp	X
C26α	*Triaenops*	Hipposideridae	1	AFR	935 bp	
C27α	*Triaenops*		1	AFR	935 bp	
C28α	*Hipposideros*		4	AFR	816 bp	X
C29β	*Pteropus*	Pteropodidae	1	AFR	805 bp	
C30β	*Rousettus*		1	AS	935 bp	X
C31β	*Cynopterus*		2	SEA	422 bp	
C32β	*Dobsonia*		1	SEA	394 bp	
C33β	*Rousettus*		2	AFR, AS	935 bp	X
C34β	*Epomophorus*		1	AFR	416 bp	
C35β	*Eidolon*		1	AFR	935 bp	X
C36β	*Rhinolophus*	Rhinolophidae	10	AFR, AS, EU	935 bp	X (AS, EU)
C37β	*Hipposideros*	Hipposideridae	5	AFR, AS, SEA	935 bp	X (AFR, AS)
C38β	*Nycteris*	Nycteridae	1	AFR	816 bp	X
C39β	*Tylonicteris*	Vespertilionidae	1	AS	935 bp	X
C40β	*Pipistrellus*		2	AS	935 bp	X
C41β	*Pipistrellus*		1	EU	392 bp	
C42β	*Eptesicus*		1	AS	408 bp	
C43β	*Pipistrellus*		2	EU	903 bp	
C44β	*Eptesicus*		2	AS, EU	895 bp	

[a] Sampling locations are indicated according to their large-scale geographic area, comprising Europe (EU), Africa (AFR), North America (NAM), Latin America (LAM) (Central and South America), Asia (AS), South East Asia (SEA), and Australia (AUS).
[b] Indicates the presence of a spike protein sequence for one or more of the RdRp sequences included within the cluster; the sampling macro-area is indicated in brackets.

$p = 0.0002$). However, the correlations between CoV diversity and the number of GenBank submissions and the total number of CoV sequences within each bat genus were also very strong ($rs = 0.64$, $p = 0.0006$; $rs = 0.73$, $p < 0.0001$). Hence, it is possible that these sampling biases have had a strong impact on the results.

The majority of genus-specific clusters (32/44) originated from a single large-scale geographic area. Of the remaining 12/44 clusters showing a broader geographical spread, 11 were identified in two or three geographic areas and one in more than three geographic areas (Table 2, Fig. 2). Of note is the frequent clustering of coronaviruses from Asia and South-East Asia (China and Thailand) associated with *Miniopterus* (C10-11α), *Hipposideros* (C9α, C27β) and *Schotophilus* (C16α) (Table 2).

The availability of RdRp sequences > 816 bp enabled the classification of 27/44 genus-specific clusters using the RdRp group units (RGU) previously defined for CoVs (Drexler et al., 2010) (C2α-C5α, C7α-C11α, C15α-C16α, C18α, C20α, C25α-C28α, C30β, C33β, C35β-C40β, C43β-C44β) (Table 3). The taxonomy of all other clusters was not resolved due to short fragment lengths. In all but two cases, the pairwise amino acid distance between sequences > 816 bp was consistent with the association between RdRp genus-specific clusters identified in this study and distinct RGUs. The exceptions, C18α and C20α, both associated with *Myotis* bats, and C43β and C44β, associated with *Pipistrellus* and *Eptesicus*, respectively, shared 96% similarity at the amino acid level in 816 bp of RdRp and hence should be classified as a single RGU (Table 3). Mean nucleotide divergence between sequences from *Eptesicus* and *Pipistrellus* was 39.3% (SE = 0.11), compared to only 1.6% (SE = 0.005) mean divergence within sequences from *Pipistrellus* bats only. Interestingly, our analyses identified these RdRp genus-specific clusters as belonging to the same putative species of MERS-CoV, with amino acid identities of 98.2% and 96.3%, respectively (Table 3).

Spike protein sequences were available for 17/44 RdRp genus-specific clusters (C2α, C7α, C9α-C11α, C16α, C18α, C25α, C28α, C30β, C33β, C35β-C40β), all of which were taxonomically resolved based on RGU classification (Table 3). Mean amino acid divergence in spike sequences from CoVs included within the same RGU ranged from 0.46% (C39β) to 48.29% (C37β) (Table 3). Tree topologies based on the spike protein sequences were largely consistent with RdRp clustering detected in this study (Fig. S3). However, there were discordant tree topologies between the RdRp and spike proteins for the two clusters

Table 3
Amino acid diversity within and between clusters based on the RdRp and the spike protein, expressed as percentages (with SE). Spike protein data are only shown sequences for whose clusters correspond with those obtained from the RdRp.

Cluster	Host genus	Mean within cluster amino acid divergence		Mean amino acid divergence from the closest group[b]	
		RdRp[a]	Spike	RdRp	Spike
C1α	*Sturnira*	0.62 (0.005)			
C2α	*Rhinolophus*	**0 (0)**		17.2 (0.023) – C28α	60.97 (0.015) – Suncus_α
C3α	*Carollia*	**0.75 (0.003)**		13.3 (0.019) – C4α	
C4α	*Artibeus*	**0.24 (0.002)**		13.3 (0.019) – C3α	
C5α	*Artibeus*	**0 (0)**		12.9 (0.018) – C8α	
C6α	*Molossus*	0 (0)			
C7α	*Myotis*	**3.8 (0.009)**	11.72 (0.007)	17 (0.019) – PEDV	40.22 (0.01) – C10α
C8α	*Rhinolophus*	**0.14 (0.001)**		7.8 (0.017) – C9α	
C9α	*Hipposideros*	**0.3 (0.002)**	19.42 (0.007)	7.8 (0.017) – C8α	47.45 (0.013) – C7α
C10α	*Miniopterus*	**1.96 (0.003)**	31.45 (0.009)	6.8 (0.013) – C11α	40.22 (0.01) – C7α
C11α	*Miniopterus*	**3.44 (0.005)**	20.59 (0.007)	6.8 (0.013) – C10α	41.45 (0.012) – C7α
C12α	*Pipistrellus*	0 (0)			
C13α	*Eptesicus*	0 (0)			
C14α	*Nyctalus*	0 (0)			
C15α	*Murina*	**0.14 (0.001)**		7.4 (0.018) – PEDV	
C16α	*Schotophilus*	3.73 (0.015)		8.5 (0.018) – C18α	37.81 (0.012) – PEDV
C17α	*Myotis*	1.86 (0.005)			
C18α	*Myotis*	2.60 (0.008)		4.1 (0.017) – C20α[c]	39.64 (0.013) – PEDV
C19α	*Myotis*	0.86 (0.003)			
C20α	*Myotis*	0.69 (0.003)		4.1 (0.017) – C18α[c]	
C21α	*Pipistrellus*	0 (0)			
C22α	*Pipistrellus*	0 (0)			
C23α	*Myotis*	0.69 (0.004)			
C24α	*Myotis*	0 (0)			
C25α	*Myotis*	0 (0)		11 (0.02) – PEDV	46.68 (0.012) – C16α
C26α	*Triaenops*	**0 (0)**		6.3 (0.014) – NL63	
C27α	*Triaenops*	**0 (0)**		9.6 (0.02) – C26α	
C28α	Hipposideros	**1.40 (0.004)**	19.07 (0.007)	1.5 (0.006) – 229E[c]	17.46 (0.008) – 299E
C29β	*Pteropus*	1.52 (0.005)			
C30β	*Rousettus*	**0 (0)**	1.03 (0.002)	6.7 (0.014) – C33β	36.63 (0.011) – C33β
C31β	*Cynopterus*	0.60 (0.005)			
C32β	*Dobsonia*	0 (0)			
C33β	*Rousettus*	**1.36 (0.004)**	28.59 (0.007)	6.7 (0.014) – C30β	36.63 (0.011) – C30β
C34β	*Epomophorus*	0 (0)			
C35β	*Eidolon*	0.07 (0.00)		10.3 (0.015) – C33β	47.98 (0.013) – C30β
C36β	*Rhinolophus*	**1.00 (0.003)**	17.72 (0.01)	4.6 (0.005) – SARSV[c]	21.96 (0.011) – SARSV
C37β	*Hipposideros*	**4.16 (0.009)**	48.29 (0.01)	17.2 (0.021) – C36β	61.14 (0.011) - SARSV
C38β	*Nycteris*	**0 (0)**		6.6 (0.013) – C44β	44.89 (0.013) – Hedgehog_CoV
C39β	*Tylonicteris*	**0.21 (0.001)**	0.46 (0.001)	6.2 (0.016) – C40β	31.23 (0.013) – C40β
C40β	*Pipistrellus*	**0.04 (0.00)**	16.3 (0.003)	5.2 (0.013) – C43β	31.23 (0.013) – C39β
C41β	*Pipistrellus*	0 (0)			
C42β	*Eptesicus*	0.72 (0.007)			
C43β	*Pipistrellus*	**0 (0)**		1.8 (0.007) – MERSV[c]	
C44β	*Eptesicus*	**0 (0)**		3.7 (0.01) – MERSV/C43β[c]	

[a] Distances calculated between RdRp sequences longer than 816 bp are indicated in bold.
[b] Distances are only calculated between clusters containing sequences equal or longer than 816 bp, including those from non-flying mammals.
[c] Clusters compatible with the inclusion within a single RGU.

associated with miniopterus bats which did not cluster together in the latter, with mean amino acid divergence of 42.4% (SE = 0.01) in the spike protein compared to only 6.8% (0.013) for the RdRp (Figs. S3, 2, Table 3). Furthermore, three sequences from *Miniopterus fuliginosus* (accession numbers KJ473800, KJ473799, KJ473797) belonging to RdRp genus-specific cluster C10α fell into spike protein cluster C11α (Fig. S3). A similar pattern was observed for C30β associated with rousettus bats; these fell within cluster 33β on the RdRp tree but were distinct from it in the spike protein tree (Figs. S3, 2). Finally, our analyses confirmed a different evolutionary history for the spike protein of *human Coronavirus NL63*; this grouped with CoVs from hipposideros bats (C28α) and *human Coronavirus 229E* rather than being nested within sequences from bats of the genus *Trienops* (RdRp genus-specific cluster C26α and C27α) as seen in the RdRp phylogenies (Figs. S3, 2).

3.3. Co-phylogenetic analyses of CoVs and their hosts

Despite the likely antiquity of CoVs in bats, genus-specific clusters showed a low level of phylogenetic congruence in respect to their hosts (Fig. S4). Indeed, a full reconciliation analysis (using Jane) suggested that the evolutionary history of CoVs was best explained by more frequent cross-species transmission than co-divergence, with the latter accounting for 0–18.3% of all the events observed (Table 4). For example, across the CoVs as a whole, there were 0–11 co-divergence events compared to 38–47 host-shift events. Importantly, this result was independent of the cost associated with co-divergence (assigned as either equal or lower than that associated with cross-species transmission) and was the same when the alpha and beta CoVs were analyzed independently, such that there was no difference in the frequency of cross-species transmission between these viral groups. Equivalent results were obtained using the genetically confirmed hosts in data sets_2 (Table S4).

3.4. Analyses of putative cross-species transmission events

A total of 27 CoV sequences were identified as likely cross-species

Table 4
Frequency of different evolutionary scenarios following co-phylogenetic reconciliation analysis (Jane) of the full set of sequences. Results are shown assuming equal or lower costs for co-divergence compared to the other possible evolutionary events.

	Association tested	Co-divergence	Virus lineage duplication	Host-shift	Virus loss	Failure of virus divergence	Co-divergence vs. other events (%)
Co-divergence cost = other events	Mammals: α-βCoVs	0	10	47	0	0	0
	Mammals: αCoVs	0	6	27	0	0	0
	Mammals: βCoVs	0	2	21	0	0	0
Co-divergence cost < other events	Mammals: α-βCoVs	11	8	38	3	0	18.3
	Mammals: αCoVs	5	5	23	1	0	14.7
	Mammals: βCoVs	3	3	17	0	0	13

Table 5
Summary of biological information shared between the donor and recipient CoV hosts.

CoV donor cluster	Host genus		No. of cross-species transmissions (RdRp)	Availability of corresponding spike sequences	Characteristics shared between recipient and donor hosts			Co-roosting of hosts (documented)[a]	Co-roosting of hosts (potential)[b]	RdRp	Spike
	Donor	Recipient			Host family	Host superfamily	Sampling country			CoV divergent from donor cluster	
3α	*Carollia*	*Artibeus*	2	na	X	X	–	–	–	X	na
7α	*Myotis*	*Miniopterus*	1	0	–	X	X	X	X	–	na
8α	*Rhinolophus*	*Epomophorus*	2	na	–	–	X	–	–	–	na
8α	*Rhinolophus*	*Rousettus*	1	na	–	–	X	–	X	–	na
9α	*Hipposideros*	*Cynopterus*	1	0	–	–	X	–	–	–	na
9α	*Hipposideros*	*Thapozous*	1	0	–	–	X	X	X	–	na
9α	*Hipposideros*	*Rousettus*	2	2	–	–	X	–	X	–	X
9α	*Hipposideros*	*Myotis*	1	0	–	–	X	–	X	–	na
10α	*Miniopterus*	*Thapozous*	1	0	–	–	X	X	X	X	na
10α	*Miniopterus*	*Rhinolophus*	1	0	–	–	X	–	X	–	na
10α	*Miniopterus*	*Murina*	1	0	–	X	X	–	–	X	na
11α	*Miniopterus*	*Eidolon*	1	0	–	–	X	–	–	X	na
11α	*Miniopterus*	*Eptesicus*	1	0	–	X	–	–	–	–	na
11α	*Miniopterus*	*Hipposideros*	3	0	–	–	X	X	X	–	na
14α	*Nyctalus*	*Myotis*	1	na	X	X	X	–	–	–	na
18α	*Myotis*	*Rhinolophus*	1	na	–	–	X	–	X	–	na
31β	*Cynopterus*	*Hipposideros*	1	na	–	–	X	–	–	–	na
36β	*Rhinolophus*	*Chaerephon*	1	1	–	–	X	–	X	–	–
44β	*Eptesicus*	*Nyctalus*	1	na	X	X	X	–	–	X	na
44β	*Eptesicus*	*Myotis*	3	na	X	X	X	–	–	–	na

na: not applicable.
[a] Documented by associated literature.
[b] Potential co-roosting was based on roost sharing of sympatric species, based on information provided by IUCN (http://www.iucnredlist.org consulted on January 2016).

transmission events based on RdRp topology, as they were nested within clusters associated with a different host genus with strong bootstrap support (Fig. 2). These cross-species transmission events involved 20 different combinations of recipient and donor hosts (13 recipient and eight donor genera, respectively), among which the association between *Cynopterus* and *Hipposideros* was bi-directional. In five cases, cross-species transmission events involved more than one highly similar CoV from the same recipient bat species. Up to six cross-species transmissions were recorded for a single donor host, with the highest frequency in clusters associated with the genera *Miniopterus*, *Hipposideros* and *Rhinolophus* (Table 5). In 17/27 cases recipient and donor hosts belonged to different bat superfamilies. Notably, the bats involved in cross-species transmission were largely sampled from the same geographic location, with a sharing of roosts between host species documented in only 6/27 cases and possible in 13/27 (Table 5). In only two cases did cross-species transmission involve related hosts from different geographical areas, namely phyllostomidae bats of the genera *Carollia* and *Artibeus* sampled in Costa Rica and Panama, respectively, and bats of the genera *Miniopterus* and *Eptesicus* from China and USA, both belonging to the superfamily Vespertilionoidea. Unfortunately, the host classification was not confirmed genetically for any of these events so that more accurate analyses could not be performed.

Sequences identified as likely cross-species transmissions were generally located at tree tips and were not significantly divergent from those of their putative donor cluster suggesting that they most likely represent recent host jumps ($p > 0.05$ in 21/27 cases) (Table 5). Among sequences exhibiting divergence from the donor cluster, 4/6 were sampled from related hosts, belonging at least to the same superfamily.

Spike sequences were available for 3/27 CoVs identified as cross-species transmission events, involving jumps from the bat genera *Hipposideros* to *Rousettus* and from *Rhinolophus* to *Chaerepon*. Interestingly, spike protein sequences from rousettus bats were significantly divergent from hipposideros CoVs and constituted a sister group to those from this genus (Table 5). Although this is compatible with host adaptation, this clearly needs to be investigated in greater detail.

4. Discussion and conclusions

Alpha and beta coronaviruses exhibit both substantial genetic diversity and a wide geographical range in bats, such that these mammals are important virus hosts (Anthony et al., 2017b; Drexler et al., 2014). Our data indicate that bats are indeed associated with several distinct clusters of alpha and beta CoVs, and that most CoVs from non-flying mammals also fell within these clusters on phylogenetic trees. However, we also noted the existence of monophyletic groups of CoVs associated with mammalian species other than bats, such as *Alphacoronavirus 1* and *Betacoronavirus 1*. Interestingly, we revealed a positive association between CoV diversity and the species richness and geographical

distribution of samples from each bat genus, with more virus clusters in genera for which more species have been sampled across a wide geographic area, such as *Myotis*, *Pipistrellus*, *Rhinolophus*, and *Hipposideros*. This suggests that the high phylogenetic diversity of CoVs likely reflects the large number of different bat species and their global distribution (Fenton and Simmons, 2015; Woo et al., 2012), and supports the central role for these animals in CoV evolution. Furthermore, we detected a strong impact of sampling effort on CoV diversity, suggesting that this still likely underestimated in bats as a whole (Anthony et al., 2017b).

Our data also provide evidence for the close phylogenetic relationship of coronaviruses from hosts of the same genus, with distinct putative CoV species tending to be associated with different bat genera. Indeed, single RGUs (which are considered indicative of distinctive CoV species) were associated with more than one bat species, particularly in those bats that live sympatrically or that belong to the same genus. Similarly, despite the evidence for geographical structuring, CoVs belonging to the same RGUs were also described in different locations, as previously described (Leopardi et al., 2016). Hence, as a general rule, CoVs exhibit genus-specificity rather than species-specificity. Conversely, CoV clusters associated with bat genera *Pipistrellus* and *Eptesicus* showed low divergence at the amino acid level based on the 816 bp fragment, suggesting their classification within the same CoV putative species. Interestingly, these clusters were also highly similar to MERS-CoV, such that they may be included within a single CoV species, although more data are needed to confirm this hypothesis.

It was notable that co-specific *Miniopterus*, *Rousettus*, *Pipistrellus*, *Rhinolophus* and *Hipposideros* bats sampled from the same location harbor more than one CoV cluster, which will increase the likelihood of virus recombination (Parrish et al., 2008). This hypothesis is supported by our finding of discordant topologies between trees estimated using the RdRp and the spike protein for CoVs detected in *Rousettus* and *Miniopterus* bats and thereby supporting previous results (Huang et al., 2016; Wu et al., 2015). In this context, it is noteworthy that bats from the genera *Pipistrellus*, *Rhinolophus* and *Hipposideros* are also those that show the highest identity with three human coronaviruses, namely MERS CoV, SARS CoV and hCoV 229E, respectively.

A key observation of our study was the frequency with which cross-species transmission has occurred in the evolution history of coronaviruses, reflected in the general incongruence between the virus and host phylogenies, the lack of species-specificity, and the presence of phylogenetically divergent viruses in some bat genera suggesting multiple introductions of CoVs (Cui et al., 2007). This hypothesis is consistent with the finding of highly related CoVs in different species non-flying mammals; for example, feline coronavirus (FCoV) and canine coronavirus (CCoV) define two sister clades within the *Alphacoronavirus 1* species, and are likely a result of recent interspecies jumping (Woo et al., 2009).

Despite the frequency of cross-species transmission, it was striking that only one recent host-jump between different bat genera was confirmed in our data set. This involved two distinct clusters of CoVs from closely related Vespertilionid bats from the genera *Pipistrellus* and *Eptesicus*, likely associated with the same CoV species. The availability of spike sequences also allowed us to identify a clear diversification of CoVs in bats of the genus Rousettus following their jump from a donor cluster associated with the co-roosting bat genus *Hipposideros*. Conversely, however, our data suggest that cross-species transmissions between distantly related hosts often result in transient spill-over infections, reflected as an absence of daughter lineages in phylogenetic trees (although this may also reflect a lack of appropriate sampling).

The multiple introductions of CoVs in Rhinopomatidae bats suggests that cross-species transmission events may be favored by their ecology of sharing roosts with different species, including *Myotis*, *Miniopterus* and *Rousettus* bats. However, it is important to note that we also detected cross-species transmission events between species for which no interaction is suspected based on their ecologies, such as the fruit bats *Eidolon helvum* that mainly roost in big colonies in trees, and the cave dwelling insectivorous bat *Miniopterus natalensis* (Fenton and Simmons, 2015). Although this suggests that we have not sampled the key intermediate species, we necessarily cannot exclude cross-contamination or incorrect classification of hosts as confounding factor in these cases.

Overall, our results confirm the long-term evolution of mammalian coronaviruses within bats, seemingly representing a complex interplay between co-divergence and cross-species transmission, a pattern that is seeming common among RNA viruses (Geoghegan et al., 2017). In addition, we found evidence that likelihood cross-species transmission increased with sympatry.

Despite its large-scale, this study has several limitations, mostly related to the quality of the available data. Of particular note is the short length of most fragments of coronaviruses used for analyses, with less than half compressing the 816 bp necessary for RGU classification (Drexler et al., 2010), and which obviously limit phylogenetic resolution. Similarly, the uncertain classification of certain host species should not be underestimated, due to variable species assignments as well as the cryptic nature of several bat species that cannot be readily identified based on obvious morphological features but whose correct assignation often require the use of genetic or echolocation studies (Kingston et al., 2001) (Table S1). Indeed, the hosts included in our database represent only about 20% of the bat species (6% of the bat genera) described worldwide. Studies of CoV diversity have mainly been performed in China and Europe, while important hot spots for bat biodiversity (Richardson, 2002), such as South East Asia and Latin America, are under-represented. Furthermore, sequences collected from different areas are generally weakly representative for their specific continent with, for example, most African samples being collected from Ghana and Kenya (Fig. 1). While the reliability of our conclusions was confirmed by analyses performed on a much smaller data set with more accurate host assignments, we encourage a more comprehensive sampling, the collection of longer CoV sequences, and the accurate genetic attribution of the host species, all of which provide the information needed to better reveal the evolution and ecology of coronaviruses in bats.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2018.01.012.

Acknowledgements

We are thankful to Mrs. Alice Fusaro for support and two anonymous reviewers for suggestions.

Funding

This study was partially supported by the Italian Ministry of Health through the Ricerca Finalizzata GR-2011-02350591 “An epizootiological survey of bats as reservoirs of emerging zoonotic viruses in Italy: implications for public health and biological conservation” and by the European Commission through FP7, project PREDEMICS (Grant agreement no 278433). GZ was supported by the Italian Ministry of Health through the RC IZS VE 05/14. ECH is supported by an ARC Australian Laureate Fellowship (FL170100022).

References

Anthony, S.J., Ojeda-Flores, R., Rico-Chávez, O., Navarrete-Macias, I., Zambrana-Torrelio, C.M., Rostal, M.K., Epstein, J.H., Tipps, T., Liang, E., Sanchez-Leon, M., Sotomayor-Bonilla, J., Aguirre, A.A., Ávila-Flores, R.A., Medellín, R.A., Goldstein, T., Suzán, G., Daszak, P., Lipkin, W.I., 2013. Coronaviruses in bats from Mexico. J. Gen. Virol. 94, 1028–1038. http://dx.doi.org/10.1099/vir.0.049759-0.

Anthony, S.J., Gilardi, K., Menachery, V.D., Goldstein, T., Ssebide, B., Mbabazi, R., Navarrete-Macias, I., Liang, E., Wells, H., Hicks, A., Petrosov, A., Byarugaba, D.K., Debbink, K., Dinnon, K.H., Scobey, T., Randell, S.H., Yount, B.L., Cranfield, M., Johnson, C.K., Baric, R.S., Lipkin, W.I., Mazet, J.A.K., 2017a. Further evidence for bats as the evolutionary source of Middle East Respiratory Syndrome coronavirus. MBio 8, 1–13. http://dx.doi.org/10.1128/mBio.00373-17.

Anthony, S.J., Johnson, C.K., Greig, D.J., Kramer, S., Che, X., Wells, H., Hicks, A.L., Joly,

D.O., Wolfe, N.D., Daszak, P., Karesh, W., Lipkin, W.I., Morse, S.S., Consortium, P., Mazet, J.A.K., Goldstein, T., 2017b. Global patterns in coronavirus diversity. J. Trop. Pediatr. Evol. 3, 1–15. http://dx.doi.org/10.1093/tropej/fmw080.

Calisher, C.H., Childs, J.E., Field, H.E., Holmes, K.V., Schountz, T., 2006. Bats: important reservoir hosts of emerging viruses. Clin. Microbiol. Rev. 19, 531–545. http://dx.doi.org/10.1128/CMR.00017-06.

Conow, C., Fielder, D., Ovadia, Y., Libeskind-Hadas, R., 2010. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms Mol. Biol. 5, 16. http://dx.doi.org/10.1186/1748-7188-5-16.

Corman, V.M., Rasche, A., Diallo, T.D., Cottontail, V.M., Stöcker, A., Souza, B.F.D.C.D., Corrêa, J.I., Carneiro, A.J.B., Franke, C.R., Nagy, M., Metz, M., Knörnschild, M., Kalko, E.K.V., Ghanem, S.J., Morales, K.D.S., Salsamendi, E., Spínola, M., Herrler, G., Voigt, C.C., Tschapka, M., Drosten, C., Drexler, J.F., 2013. Highly diversified coronaviruses in neotropical bats. J. Gen. Virol. 94, 1984–1994. http://dx.doi.org/10.1099/vir.0.054841-0.

Corman, V.M., Baldwin, H.J., Tateno, A.F., Zerbinati, R.M., Annan, A., Owusu, M., Nkrumah, E.E., Maganga, G.D., Oppong, S., Adu-Sarkodie, Y., Vallo, P., da Silva Filho, L.V.R.F., Leroy, E.M., Thiel, V., van der Hoek, L., Poon, L.L.M., Tschapka, M., Drosten, C., Drexler, J.F., 2015. Evidence for an ancestral association of human coronavirus 229E with bats. J. Virol. 89, 11858–11870. http://dx.doi.org/10.1128/JVI.01755-15.

Corman, V.M., Eckerle, I., Memish, Z.A., Liljander, A.M., Dijkman, R., Jonsdottir, H., Juma Ngeiywa, K.J.Z., Kamau, E., Younan, M., Al Masri, M., Assiri, A., Gluecks, I., Musa, B.E., Meyer, B., Müller, M.A., Hilali, M., Bornstein, S., Wernery, U., Thiel, V., Jores, J., Drexler, J.F., Drosten, C., 2016. Link of a ubiquitous human coronavirus to dromedary camels. Proc. Natl. Acad. Sci. U. S. A. 201604472. http://dx.doi.org/10.1073/pnas.1604472113.

Cui, J., Han, N., Streicker, D., Li, G., Tang, X., Shi, Z., Hu, Z., Zhao, G., Fontanet, A., Guan, Y., Wang, L., Jones, G., Field, H.E., Daszak, P., Zhang, S., 2007. Evolutionary relationships between bat coronaviruses and their hosts. Emerg. Infect. Dis. 13, 1526–1532. http://dx.doi.org/10.3201/eid1310.070448.

De Vienne, D.M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M.E., Giraud, T., 2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. New Phytol. 198, 347–385. http://dx.doi.org/10.1111/nph.12150.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, 465–469. http://dx.doi.org/10.1093/nar/gkn180.

Drexler, J.F., Gloza-Rausch, F., Glende, J., Corman, V.M., Muth, D., Goettsche, M., Seebens, A., Niedrig, M., Pfefferle, S., Yordanov, S., Zhelyazkov, L., Hermanns, U., Vallo, P., Lukashev, A., Müller, M.A., Deng, H., Herrler, G., Drosten, C., 2010. Genomic characterization of Severe Acute Respiratory Syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. J. Virol. 84, 11336–11349. http://dx.doi.org/10.1128/JVI.00650-10.

Drexler, J.F., Corman, V.M., Drosten, C., 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antivir. Res. 101, 45–56. http://dx.doi.org/10.1016/j.antiviral.2013.10.013.

Fenton, M.B., Simmons, N.B., 2015. Bats a World of Science and Mistery. University of Chicago Press, Chicago.

Ge, X.-Y., Li, J.-L., Yang, X.-L., Chmura, A.A., Zhu, G., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C., Zhang, Y.-J., Luo, C.-M., Tan, B., Wang, N., Zhu, Y., Crameri, G., Zhang, S.-Y., Wang, L.-F., Daszak, P., Shi, Z.-L., 2014. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503, 535–538. http://dx.doi.org/10.1038/nature12711.

Geoghegan, J.L., Duchêne, S., Holmes, E.C., 2017. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog. 13, 1–17. http://dx.doi.org/10.1371/journal.ppat.1006215.

Graham, R.L., Donaldson, E.F., Baric, R.S., 2013. A decade after SARS. Strategies for controlling emerging coronaviruses. Nat. Rev. Microbiol. 11, 836–848. http://dx.doi.org/10.1038/nrmicro3143.

Halpin, K., Young, P.L., Field, H.E., Mackenzie, J.S., 2000. Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus. J. Gen. Virol. 81, 1927–1932. http://dx.doi.org/10.1099/0022-1317-81-8-1927.

Holmes, E.C., 2009. The Evolution and Emergence of RNA Viruse. Oxford Series in Ecology and Evolution (OSEE). Oxford University Press, Oxford.

Hu, B., Ge, X., Wang, L.-F., Shi, Z., 2015. Bat origin of human coronaviruses. Virol. J. 12, 221. http://dx.doi.org/10.1186/s12985-015-0422-1.

Huang, C., Liu, W.J., Xu, W., Jin, T., Zhao, Y., Song, J., Shi, Y., Ji, W., Jia, H., Zhou, Y., Wen, H., Zhao, H., Liu, H., Li, H., Wang, Q., Wu, Y., Wang, L., Liu, D., Liu, G., Yu, H., Holmes, E.C., Lu, L., Gao, G.F., 2016. A bat-derived putative cross-family recombinant coronavirus with a reovirus gene. PLoS Pathog. 12, 1–25. http://dx.doi.org/10.1371/journal.ppat.1005883.

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. http://dx.doi.org/10.1093/nar/gkf436.

Kingston, T., Lara, M.C., Jones, G., Akbar, Z., Kunz, T.H., Schneider, C.J., 2001. Acoustic divergence in two cryptic *Hipposideros* species: a role for social selection? Proc. R. Soc. B Biol. Sci. 268, 1381–1386. http://dx.doi.org/10.1098/rspb.2001.1630.

Leopardi, S., Oluwayelu, D., Meseko, C., Marciano, S., Tassoni, L., Bakarey, L., Monne, I., Cattoli, G., De Benedictis, P., 2016. The close genetic relationship of lineage D betacoronavirus from Nigerian and Kenyan straw-colored fruit bats (*Eidolon helvum*) is consistent with the existence of a single epidemiological unit across sub-Saharan Africa. Virus Genes 52, 573–577. http://dx.doi.org/10.1007/s11262-016-1331-0.

Letunic, I., Bork, P., 2016. Interactive Tree Of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245. http://dx.doi.org/10.1093/nar/gkw290.

Lin, X.D., Wang, W., Hao, Z.Y., Wang, Z.X., Guo, W.P., Guan, X.Q., Wang, M.R., Wang, H.W., Zhou, R.H., Li, M.H., Tang, G.P., Wu, J., Holmes, E.C., Zhang, Y.Z., 2017. Extensive diversity of coronaviruses in bats from China. Virology 507, 1–10. http://dx.doi.org/10.1016/j.virol.2017.03.019.

Menachery, V.D., Yount, B.L., Sims, A.C., Debbink, K., Agnihothram, S.S., Gralinski, L.E., Graham, R.L., Scobey, T., Plante, J.A., Royal, S.R., Swanstrom, J., Sheahan, T.P., Pickles, R.J., Corti, D., Randell, S.H., Lanzavecchia, A., Marasco, W.A., Baric, R.S., 2016. SARS-like WIV1-CoV poised for human emergence. Proc. Natl. Acad. Sci. 113, 201517719. http://dx.doi.org/10.1073/pnas.1517719113.

Moratelli, R., Calisher, C.H., 2015. Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? Mem. Inst. Oswaldo Cruz 110, 1–22. http://dx.doi.org/10.1590/0074-02760150048.

Parker, J., Rambaut, A., Pybus, O.G., 2008. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. Infect. Genet. Evol. 8, 239–246. http://dx.doi.org/10.1016/j.meegid.2007.08.001.

Parrish, C.R., Holmes, E.C., Morens, D.M., Park, E., Burke, D.S., Calisher, C.H., Laughlin, C.A., Saif, L.J., Daszak, P., 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol. Mol. Biol. Rev. 72, 457–470. http://dx.doi.org/10.1128/MMBR.00004-08.

Pfefferle, S., Oppong, S., Drexler, J.F., Gloza-Rausch, F., Ipsen, A., Seebens, A., Müller, M.A., Annan, A., Vallo, P., Adu-Sarkodie, Y., Kruppa, T.F., Drosten, C., 2009. Distant relatives of Severe Acute Respiratory Syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. Emerg. Infect. Dis. 15, 1377–1384. http://dx.doi.org/10.3201/eid1509.090224.

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

Reusken, C.B.E.M., Raj, V.S., Koopmans, M.P., Haagmans, B.L., 2016. Cross host transmission in the emergence of MERS coronavirus. Curr. Opin. Virol. 16, 55–62. http://dx.doi.org/10.1016/j.coviro.2016.01.004.

Richardson, P., 2002. Bats, Natura His. Natural History Museum, London.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574. http://dx.doi.org/10.1093/bioinformatics/btg180.

Sabir, J.S.M., Lam, T.T.-Y., Ahmed, M.M.M., Li, L., Shen, Y., Abo-Aba, E.M., Qureshi, M.I., Abu-Zeid, M., Zhang, Y., Khiyami, M.A., Alharbi, N.S., Hajrah, N.H., Sabir, M.J., Mutwakil, M.H.Z., Kabli, S.A., Alsulaimany, F.A.S., Obaid, A.Y., Zhou, B., Smith, D.K., Holmes, E.C., Zhu, H., Guan, Y., 2016. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science 351, 81–84. http://dx.doi.org/10.1126/science.aac8608.

Smith, C.S., De Jong, C.E., Meers, J., Henning, J., Wang, L.-F., Field, H.E., 2016. Coronavirus infection and diversity in bats in the Australasian region. EcoHealth. http://dx.doi.org/10.1007/s10393-016-1116-x.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. http://dx.doi.org/10.1093/molbev/mst197.

Tang, X.C., Zhang, J.X., Zhang, S.Y., Wang, P., Fan, X.H., Li, L.F., Li, G., Dong, B.Q., Liu, W., Cheung, C.L., Xu, K.M., Song, W.J., Vijaykrishna, D., Poon, L.L.M., Peiris, J.S.M., Smith, G.J.D., Chen, H., Guan, Y., 2006. Prevalence and genetic diversity of coronaviruses in bats from China. J. Virol. 80, 7481–7490. http://dx.doi.org/10.1128/JVI.00697-06.

Tao, Y., Shi, M., Chommanard, C., Queen, K., Zhang, J., Markotter, W., Kuzmin, I.V., Holmes, E.C., Tong, S., 2017. Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history. J. Virol. http://dx.doi.org/10.1128/JVI.01953-16. (JVI.01953-16).

Vijaykrishna, D., Smith, G.J.D., Zhang, J.X., Peiris, J.S.M., Chen, H., Guan, Y., 2007. Evolutionary insights into the ecology of coronaviruses. J. Virol. 81, 4012–4020. http://dx.doi.org/10.1128/JVI.02605-06.

Wertheim, J.O., Chu, D.K.W., Peiris, J.S.M., Kosakovsky Pond, S.L., Poon, L.L.M., 2013. A case for the ancient origin of coronaviruses. J. Virol. 87, 7039–7045. http://dx.doi.org/10.1128/JVI.03273-12.

WHO, 2017. Middle East Respiratory Syndrome Coronavirus (MERS-CoV) WHO MERS-CoV Global Summary and Assessment of Risk.

Woo, P.C.Y., Lau, S.K.P., Huang, Y., Yuen, K.-Y., 2009. Coronavirus diversity, phylogeny and interspecies jumping. Exp. Biol. Med. 234, 1117–1127. http://dx.doi.org/10.3181/0903-MR-94.

Woo, P.C.Y., Lau, S.K.P., Lam, C.S.F., Lau, C.C.Y., Tsang, A.K.L., Lau, J.H.N., Bai, R., Teng, J.L.L., Tsang, C.C.C., Wang, M., Zheng, B.-J., Chan, K.-H., Yuen, K.-Y., 2012. Discovery of seven novel mammalian and avian coronaviruses in the genus *Deltacoronavirus* supports bat coronaviruses as the gene source of *Alphacoronavirus* and *Betacoronavirus* and avian coronaviruses as the gene source of *Gammacoronavirus* and *Deltacoronavirus*. J. Virol. 86, 3995–4008. http://dx.doi.org/10.1128/JVI.06540-11.

Woolhouse, M.E.J., Gowtage-Sequeria, S., 2005. Host range and emerging and reemerging infectious diseases. Emerg. Infect. Dis. 11, 1842–1847. http://dx.doi.org/10.3201/eid1112.050997.

Wright, E.S., 2016. Using DECIPHER v2.0 to analyze big biological sequence data in R. R J. 8, 352–359.

Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., Du, J., Yang, F., Zhang, S., Jin, Q., 2015. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. ISME J. 1–12. http://dx.doi.org/10.1038/ismej.2015.138.

Yang, Y., Du, L., Liu, C., Wang, L., Ma, C., Tang, J., Baric, R.S., Jiang, S., Li, F., 2014. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-

human transmission of MERS coronavirus. Proc. Natl. Acad. Sci. 111, 12516–12521. http://dx.doi.org/10.1073/pnas.1405889111.

Yang, X.L., Hu, B., Wang, B., Wang, M.N., Zhang, Q., Zhang, W., Wu, L.J., Ge, X.Y., Zhang, Y.Z., Daszak, P., Wang, L.F., Shi, Z.L., 2016. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of Severe Acute Respiratory Syndrome coronavirus. J. Virol. 90, 3253–3256. http://dx.doi.org/10.1128/JVI.02582-15.

Young, C.C.W., Olival, K.J., 2016. Optimizing viral discovery in bats. PLoS One 11, 1–18. http://dx.doi.org/10.1371/journal.pone.0149237.

Zeng, L., Gao, Y., Ge, X., Zhang, Q., Peng, C., Yang, X., Tan, B., Chen, J., 2016. Bat Severe Acute Respiratory Syndrome-like coronavirus WIV1 encodes an extra accessory protein, ORFX, involved in modulation of the host immune response. J. Virol. 90, 6573–6582. http://dx.doi.org/10.1128/JVI.03079-15.