Global patterns of phylogenetic diversity and transmission of bat coronavirus

Zhilin Wang1,2†, Guangping Huang1†, Mingpan Huang1,2, Qiang Dai3, Yibo Hu1,2, Jiang Zhou4 & Fuwen Wei1,2,5*

1CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; 2University of Chinese Academy of Sciences, Beijing 100049, China; 3Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China; 4School of Karst Science, Guizhou Normal University, Guiyang 550000, China; 5Center for Evolution and Conservation Biology, Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

Received September 2, 2022; accepted October 12, 2022; published online November 11, 2022

INTRODUCTION

Chiroptera is diverse and consists of over 1,400 bat species that inhabit all continents except Antarctica (Wilson and Mittermeier, 2019). Bats harbor the most coronavirus (CoV) diversity among all mammals, and two major CoV genera, α- and β-CoVs, have been widely detected in bats worldwide (Mollentze and Streicker, 2020). Bats are also potential ancestral hosts for several known highly pathogenic CoVs, which adversely affect global public health (Cui et al., 2019). For example, bat SARSr-CoV has been detected in horseshoe bats and it has high genomic similarity with SARS-CoV, which was first reported to infect humans in 2002 (Li et al., 2005). The newly emerged SARS-CoV-2, the causative agent of the global outbreak of coronavirus disease 2019, also shares relatively high genomic similarities with the bat CoVs BANAL-52 (Temmam et al., 2022), RaTG13 (Li et al.,...
RESULTS

Ancestral host reconstruction within continents

We collected the RdRp gene (440 bp) sequences of 3,594 bat CoVs (2,025 α-CoV and 1,569 β-CoV sequences) worldwide, together with information about the host and geographic factors (Figure 1A; Tables S1 and S2 in Supporting Information). The host dataset included bats from 11 families and 49 genera for α-CoVs and 11 families and 40 genera for β-CoVs across five continents, including Africa (n=783 sequences), Americas (n=154), Asia (n=2,068), Europe (n=301), and Oceania (n=25) (Figure 1B and C). A larger number of CoV sequences have been reported from Asia and Africa than from other continents, and very few sequences have been reported from Americas, where the global bat species richness is relatively high (Figure 1A).

To identify the ancestral hosts of bat CoVs on each continent, a Bayesian phylogeographic analysis was performed to construct the ancestral states of each node in the phylogenetic tree using a discrete state with bat families based on the host dataset. The results showed that bat α- and β-CoVs on different continents have diverse ancestral families, and Rhinolophidae and Vespertilionidae were the most common. Phylogenetic reconstruction of α-CoVs in Africa showed that Molossidae was the most likely ancestral host family (Figure 2A), but β-CoVs likely originated in Vespertilionidae and Pteropodidae (Figure 2B). The phylogenetic reconstruction of α-CoVs in Asia showed that Rhinolophidae was the most likely ancestral host family (Figure 2C), but β-CoVs likely originated in Vespertilionidae and Rhinolophidae (Figure 2D). Vespertilionidae and Rhinolophidae are the most likely ancestral host families for α-CoVs in Europe (Figure 2E), and β-CoVs likely originated in Vespertilionidae (Figure 2F). A random sampling of 30 sequences from each bat family yielded the same tree topologies for α-CoVs and β-CoVs in Africa (Figure S1A and B in Supporting Information) and Europe (Figure S1E and F in Supporting Information), respectively, suggesting the reliability and robustness of this method. Rhinolophidae remained one of the ancestral families associated with bat α- and β-CoVs in the random sampling test for Asia (Figure S1C and D in Supporting Information). Notably, samples from the Americas and Oceania were excluded because of the small number of bat CoV sequences.

Cross-species transmission across bat taxa

To assess whether bat CoVs have undergone cross-species transmission among continents, a discrete Bayesian phylogeographic analysis was performed using Bayesian stochastic search variable selection (BSSVS) to identify the best-supported CoV transmission rates among bat families, and a Bayesian factor (BF) was calculated to estimate the significance of these transmissions. We observed that interfamily-switching events for α- and β-CoVs tended to occur more frequently within continents than between continents, and the switching frequencies of α-CoVs within and between continents were significantly higher than those of β-CoVs (Figure S2 in Supporting Information). The switching of bat CoVs tended to occur within the same bat family (BF value>10) between continents, especially for β-CoVs (Figure S2 in Supporting Information). In addition, Vespertilionidae, Miniopteridae, Rhinolophidae, and Pteropodidae were frequently involved in cross-species transmission events between continents (Figure S2 in Supporting Information), suggesting that these bat taxa should be given more attention in zoonotic investigations.

A similar model was used to determine cross-species transmission among families and genera within each continent. More than three interfamily-switching events (BF>10) were observed for α-CoVs in Africa (n=9 times), Asia (n=15), Europe (n=3), and the Americas (n=3) and β-CoVs in Africa (n=3), Asia (n=10), and Europe (n=2) (Figure 3A–G), indicating that the switching frequency of α-
CoVs among bat families was much higher than that of β-CoVs within each continent; this was further confirmed by the observation of intergenus-switching events for α- and β-CoVs (Figure S3 in Supporting Information). Interhost-switching events of α-CoVs from Miniopteridae, Molossidae, Rhinolophidae, and Vespertilionidae, and the genera Chaerephon, Molossus, Rhinolophus, Miniopterus, and Myotis, were greater than those of other bat families and genera. Moreover, the genera Epomops, Rousettus, Cynopterus, Rhinolophus, and Pipistrellus and the families Pteropodidae, Rhinolophidae, and Vespertilionidae were the main β-CoV donors (Figure 3; Figure S4 in Supporting Information). The genera Tadarida, Rousettus, and Rhinolophus, as well as the families Pteropodidae, Rhinolophidae, and Vespertilionidae and the genera Eonycteris, Hipposideros, and Rousettus as well as the families Hipposideridae and Vespertilionidae were the main receivers for α-CoVs and β-CoVs, respectively (Figure 3; Figure S4 in Supporting Information). Notably, Pteropodidae acted only as a receiver in the α-CoV interfamily-switching events (Figure 3) but as the major donor in β-CoV transmission events in Africa and Asia, and Miniopterus was only involved in α-CoV cross-species transmission events (Figure S4 in Supporting Information).

We further observed that 14 of 18 (77.8%) cross-genus α-CoV switching events were between bat genera in different families, and the proportion of β-CoV dropped to 4 of 13 (30.8%) in Africa (Figure S3A and B in Supporting Information), whereas 22 of 35 (62.9%) cross-genus α-CoV and 14 of 32 (34.3%) cross-genus β-CoV switching events occurred in Asia (Figure S3C and D in Supporting Information); 3 of 6 (50%) cross-genus α-CoV and no cross-genus β-CoV switching events occurred in Europe (Figure S3E and F in Supporting Information), and 5 of 13 (38.5%) cross-genus α-CoV switching events occurred in the Americas (Figure S3G in Supporting Information). These results
suggest that interhost-switching events for α-CoV occurred in more distantly related host taxa within continents than for β-CoV. Furthermore, *Rhinolophus, Myotis, Miniopterus,* and *Hipposideros* for α-CoV (Figure S3A, C, and E in Supporting Information) and *Rhinolophus* and *Hipposideros* for β-CoV were involved in the most cross-genus switching events.
Figure 3 Interfamily bat CoV switches within continents. The left panels display the strongly supported host switches between bat families for the α-CoVs (A, C, E, and G) and β-CoVs (B, D, and F) in Africa, Asia, Europe, and the Americas, respectively. Arrows indicate the direction of the host switch, and arrow thickness is positively correlated with the host switch significance level. All exact BF values are available in Tables S5 and S6 in Supporting Information. The right panels display the histograms of the total number of host-switching events (state changes counts using Markov jumps) from/to each bat family along the significant interfamily switches for the α-CoVs (A, C, E, and G) and β-CoVs (B, D, and F) in Africa, Asia, Europe, and the Americas, respectively.
between genera among the different bat families (Figure S3B and D in Supporting Information).

To further quantify and test whether the phylogenetic distance (PD) between bat genera affected the transmission of CoVs among bat genera, we used a Bayesian phylogeographic inference approach with a generalized linear model (GLM) to quantify the contributions of the PDs between bat genera to the intergenic transition rates within continents. We also observed that the α- and β-CoV transmission rates among bat genera were restricted by the bat phylogeny, except in Europe, and particularly for β-CoVs (Figure S5 in Supporting Information). Similar results were observed for α- and β-CoVs after analyzing random sampling datasets (Tables S3–S8 in Supporting Information).

**Spatial dissemination of CoVs among geographic regions**

We also used the Bayesian discrete phylogeographic model with geographic regions as a character state to reconstruct the spatial dissemination of bat CoVs within each continent. Bat α- and β-CoVs exhibited frequent virus lineage dissemination among geographic regions (BF>10) in Africa, Asia, Europe, and the Americas (Figure 4). We further observed that lineage dissemination of α- and β-CoVs occurred between pairs of either close or distant geographical regions (Figure 4). These regions were distributed in eastern and southern Africa, southern South America, Western Europe, and Southeast Asia were involved in the most frequent α- and β-CoV spatial dissemination events (Figure 4; Figures S6 and S7 in Supporting Information). The same analysis of a random subset yielded similar trends (Tables S9, S10 in Supporting Information).

Testing the contributions of ecological factors to bat CoV lineage dissemination within each continent will likely inform the general principles of bat CoV lineage dissemination between geographic regions. The same Bayesian phylogeographic inference approach using a GLM was employed, with ten factors, including geographic distance (grid cell distance (GCD)), mean annual temperature (MAT), mean annual precipitation (MAP), human population density (HPD), bat species richness (BSR), mammalian species richness (MSR), mean fractional area of agricultural expansion (AE), mean probability of urban expansion (UE), global agricultural cropland coverage (GAL), and urban/built-up area (UBU). Among all of the factors, geographic distance was the strongest factor limiting the spatial dissemination of α-CoVs and β-CoVs within continents, except in Europe (Figure 4). In addition, MAT promoted the spread of β-CoVs in Asia (Figure 4B). These results can be used to guide the development of possible ecological interventions to block the transmission of bat CoVs.

**Phylogenetic diversity**

To identify the host taxa and geographic regions defined as hotspots of global bat CoV evolution, two phylogenetic structural indicators, namely, standardized effect size (SES) for mean phylogenetic distance index (MPD) and nearest taxon distance index (MNTD) values were used to evaluate the phylogenetic diversity of the bat CoVs.

Significant and negative SES MPD and SES MNTD values were observed in Africa, Asia, and Europe, indicating clustered phylogenetic relationships within most of the bat families and genera for α- and β-CoVs (Figure 5; Figure S8 in Supporting Information). Notably, only the genus *Pteropus* in the family Pteropodidae in Asia had a significant positive SES MPD value (*P*>0.95) for α-CoVs, indicating an overdispersed phylogenetic structure. The α-CoVs of Pteropodidae and *Tadarida* in Africa, Hipposideridae in Asia, Miniopteridae and *Pipistrellus* in Europe, and Phyllostomidae and *Glossophaga* in Americas, along with β-CoVs of Nycteridae and *Myonycteris* in Africa, Hipposideridae, and *Macroglossus* in Asia, Vespertilionidae, and *Eptesicus* in Europe had the highest SES MPD values within each continent (*P*<0.05, Figure 5; Figure S8 in Supporting Information), suggesting the host taxa hotspots for bat CoV evolution. Overall, the phylogenetic diversity of α-CoVs was lower than that of β-CoVs in most bat families and genera in Africa (Figure 5; Figure S8A and B in Supporting Information), suggesting a stronger host structure effect and greater phylogenetic conservatism for α-CoVs than β-CoVs. However, contrasting results were observed for Asia (Figure 5; Figures S8C and D in Supporting Information). Analysis of random α- and β-CoV subsets supported these results (Tables S11–S18 in Supporting Information).

Negative and mostly significant SES MPD and SES MNTD values (*P*<0.05) were observed within most geographic regions (grid cells), indicating significant phylogenetic clustering of CoV lineages in these regions (Figure 6; Figure S9 in Supporting Information). The α-CoVs in eastern Africa (grid1783) and the β-CoVs in Western Europe (grid3050) had significantly positive SES MPD and SES MNTD values, suggesting an overdispersed phylogenetic structure of bat CoV lineages in these regions (Figure 6A and C; Figure S9A and C in Supporting Information). In addition, the α-CoVs in East Asia (grid2808) and Southeast Asia (grid2437) had significantly positive SES MPD values (*P*>0.95) (Figure 6A). The phylogenetic β-diversity based on interregional MPD values also indicated that CoVs isolated from geographically close regions were more closely related, suggesting that regions adjacent to hotspot regions should be investigated (Figure S10 in Supporting Information). Analysis of the random subsets yielded mostly similar results (Tables S19–S23 in Supporting Information).

To explore the associations between ecological factors and
the phylogenetic diversity of CoVs across global geographic regions, we used the residuals from the spaGLMM analysis to individually evaluate the relationships between global bat CoV phylogenetic diversity (SES-MPD and -MNTD values) and the ecological factors (MAT, MAP, HPD, BSR, MSR, AE, UE, GAL, and UBU). BSR, MAP, and GAL were significantly correlated with α-CoV phylogenetic diversity (SES MPD), and HPD was significantly correlated with β-CoV phylogenetic diversity (SES MPD) (Figure 6B and D; Table S25 in Supporting Information), whereas MAT, UE, and UBU were significantly correlated with α-CoV phylogenetic diversity (SES MNTD), and MAP, HPD, and UBU

Figure 4  CoV spatial dissemination by geographic region. The map displays the level of the Bayes factor (BF) value support for each of the transmission routes considered for α-CoVs (A) and β-CoVs (B) in Africa, Asia, Europe, and the Americas, respectively. All exact BF values are available in Tables S9 and S10 in Supporting Information. Contributions of the predictor variables to the dissemination of the bat CoVs lineages among geographic regions in Africa, Asia, Europe, and the Americas, respectively. (Upper) Black lines show the estimated conditional effect sizes for the predictor coefficients (>0, positive association, <0, negative association). Error bars represent the 95% highest posterior density (HPD) credible interval for the estimates. (Lower) Bars show the posterior probability of inclusion of each predictor in the model. IP and IC denote inclusion probability and coefficients, respectively.
Figure 5  Phylogenetic diversity of the bat α- and β-CoVs within each bat family in Africa, Asia, and Europe. A, C, E, Phylogenetic diversity as measured by the standardized effect size of the mean phylogenetic distance (SES MPD). C, D, F, Phylogenetic diversity as measured by the standardized effect size of the mean nearest taxon distance (SES MNTD). Values departing significantly from the null model (P-value < 0.05 or >0.95) are indicated with an asterisk. All precise SES MPD and SES MNTD values are available in Tables S11–S14 in Supporting Information.
Figure 6  Phylogenetic diversity and underlying drivers. The phylogenetic structure (SES MPD) of global bat α-CoVs (A) and β-CoVs (C) using the complete geographic dataset and the underlying drivers (B and D). Red circle indicates significant coefficients, while gray circle shows nonsignificant effects. All precise SES MPD values and correlation results are available in Tables S19, S20, and S25 in Supporting Information.
were significantly correlated with β-CoV phylogenetic diversity (SES MNTD) (Figure S9B and D in Supporting Information; Table S25 in Supporting Information).

These results identify the bat taxa and geographic regions with high phylogenetic diversity of CoVs and associated factors, suggesting hotspots for the emergence of novel, zoonotic bat-borne CoVs.

DISCUSSION

CoV diversity in bats has attracted attention in epidemiological surveys. Here, we present a global pattern of the phylogenetic diversity and transmission of bat CoVs using 3,594 bat CoV RdRp sequences from 13 bat families and 64 genera (including 206 known species) across five continents. This study revealed that bat α- and β-CoVs exhibited significantly frequent interhost-switching events between and within continents but with the latter more frequent, probably due to their limited migration. Our results also indicate the ancestral α- and β-CoV bat families in Africa, Asia, and Europe, and Rhinolophidae and Vespertilionidae were the most common. Finally, our study also identified the global hotspots of bat taxonomic and geographical regions and assessed the effects of ecological factors on the transmission and phylogenetic diversity of bat CoVs, which were not determined in previous studies. These findings will help guide the prevention and control of bat-borne CoV zoonoses.

Virus host-switching is an important mechanism driving virus evolution (Anthony et al., 2017; Cui et al., 2007; Latinne et al., 2020; Streicker et al., 2012), so it is significant to identify important bat taxa to predict future bat CoV emergence and cross-species transmission. Rhinolophidae and Vespertilionidae (especially the genera *Rhinolophus* and *Myotis*) have played key roles in the cross-species transmission of bat α- and β-CoVs in Africa, Asia, and Europe, further supporting the results of a previous study (Latinne et al., 2020). Therefore, these two families in these regions may carry the oldest CoV lineages, perhaps serving as the origin of cross-transmission events. Rhinolophidae and Vespertilionidae, which are distributed worldwide (Wilson and Mittermeier, 2019), share the same habitats as other bat families, which may facilitate mutual transmission (Cui et al., 2007; Luo et al., 2013). Higher rates of viral transmission among bat species that co-roost in caves and abandoned mineshaft support this finding (Ge et al., 2016; Willoughby et al., 2017). Rhinolophidae may have originated in Asia before rapidly spreading to Africa and Europe (Stoffberg et al., 2010). The probable ancestral origin, species diversity, and the number of species in these two bat families (Wilson and Mittermeier, 2019) in these regions, combined with the high CoV diversity (Letko et al., 2020), may explain their central role in the cross-species transmission of bat CoVs. *Miniopterus*, the only bat genus in the family Miniopteridae, as well as some species in the families Vespertilionidae and Pteropodidae, have long-distance migration abilities (Breed et al., 2010; Wilson and Mittermeier, 2019), which may account for their predominant roles in CoV cross-species transmission between continents. *Miniopterus* has been implicated only in α-CoV cross-species transmission; however, this genus shares habitat with other bat species that carry β-CoVs (Luo et al., 2013); this finding requires further investigation. In contrast with the results of a previous study (Latinne et al., 2020), our results suggest that Pteropodidae plays important roles in the cross-species transmission of bat α- and β-CoVs and has more phylogenetically diverse CoV lineages in Africa and Asia, particularly concerning some important bat genera, such as *Pteropus*, *Myonycteris*, and *Macroglossus*, probably because of abundant food resources in the two regions (Salah Uddin Khan et al., 2010). Foraging promotes contact between CoV-carrying Pteropodidae and other bats, driving cross-species transmission and the evolution of CoVs (Islam et al., 2016). These findings highlight the need to increase our focus on CoV evolution and cross-species transmission in these bat taxa hotspots.

Our study revealed that eastern and southern Africa, Western Europe, and Southeast Asia were foci of spatial dissemination events and phylogenetic diversity for bat α- and β-CoVs. In particular, the regions of eastern Africa (grid1783), East Asia (grid2808), Southeast Asia (grid2437), and Western Europe (grid3050) were determined to be hotspots of bat CoV phylogenetic diversity. Southeast Asia and Western Europe have been hypothesized to act as refugia during the Quaternary glacial period for many bat species, such as *Rhinolophus affinis* (Mao et al., 2010), as have eastern Africa during the African Pleistocene forest refugia hypothesis for several bats in the family Pteropodidae (Hassanin et al., 2015). The long-term and stable presence of these bat species may partly explain the establishment of these regions as foci of bat CoV phylogenetic diversity and cross-species transmission. In addition, the abundance of food resources and habitat adaptation may be important factors in the establishment of these foci. Due to various factors, these areas have high MSR that may act as the hotspots for emerging infectious diseases originating in wildlife (Jones et al., 2008). These findings suggest that these regions should be paid more attention in the prevention of emerging bat-borne CoV diseases.

The roles of host characteristics and ecological factors in virus cross-species transmission and evolution have been identified (Allen et al., 2017; Carlson et al., 2022; Plowright et al., 2017; Streicker et al., 2010). Our GLM results strongly suggest that interhost PD constrains the cross-species transmission of CoVs among bat hosts, further supporting the hypothesis that host phylogeny is a biotic factor driving the occurrence of RNA virus cross-species transmission between...
closely related host species (Albery et al., 2021). In particular, the transmission of β-CoVs is more strongly restricted than that of α-CoVs, supporting the results of our Bayesian phylogenetic analysis that interhost-switching events for α-CoVs tend to occur between more distant relatives than those of β-CoVs, and the latter was more highly constrained by interhost PD. Geographic distance was the dominant limiting factor for the spatial dissemination of bat CoVs, which corresponds with the conclusions about the spatial spread of other viruses (Lemey et al., 2014; Yang et al., 2020). In addition, a high MAP volume promotes the abundance of food resources in Asia, which may contribute to the spread of β-CoVs by foraging bats. An example supporting this view is that in Malaysia and Bangladesh, which have abundant rainfall, the consumption of food resources contaminated by bat waste products is responsible for the spillover of the NiVthat from bats to pigs and humans (Luby et al., 2006). These results remind us that closely related and sympatric species deserve continued attention and that climatic factors should not be ignored in the context of the spread of bat CoVs. The correlation analysis also showed that BSR was a positive driver of bat α-CoV phylogenetic diversity but not β-CoV phylogenetic diversity, possibly because α-CoVs are spread across a wider range of bat host types than β-CoVs, allowing them to adapt to new hosts and driving their evolution. MAT, agricultural land use (cropland), and HPD had significant negative effects on CoV phylogenetic diversity, possibly because these factors weakened the linkages between different bat taxa. Previous research has shown that changes in climate and human activities alter connectivity (e.g., shared habitats and food resources) among different hosts and affect species dispersal (Beyer et al., 2021; Carlson et al., 2022; Newbold et al., 2018); thus, influencing the evolution of viruses, as jumping between hosts drives viral evolution. These findings suggest that human- and climate-related factors may also have profoundly affected the evolution of bat-borne CoVs in addition to BSR.

Although the available RdRp sequences of the global bat CoVs were collected, the RdRp sequences were amplified using conserved primers (Watanabe et al., 2010), which may have reduced the detection of variant strains, and whole genomes or other gene sequences (such as the spike gene) should be considered in a future study. The RdRp gene sequence is relatively conserved with less gene recombination than other CoV genes (Yin et al., 2020), which would better reflect the evolutionary history of bat CoVs. The topology of the phylogenetic tree constructed with the RdRp sequence was highly consistent with that of a previous study (Anthony et al., 2017). In addition, a large number of bat species inhabit the Americas, but there is a lack of available data on bat CoVs. Therefore, more investigations are needed to obtain more data from these regions.

In conclusion, this study revealed the global landscape of the macroevolution, ecology, and transmission of bat CoVs across bat taxa and identified bat taxa and geographic regions that define potential hotspots for the emergence of novel zoonotic bat-borne CoVs. These findings will help guide future prevention and control of bat-borne CoV zoonoses.

METHODS

Sequence data

We collected the sequence data of the highly conserved RdRp gene (440 bp) of global bat α- and β-CoVs from the GenBank and GISAID databases (Tables S1 and S2 in Supporting Information). As few gamma (γ) and delta (δ) CoV sequences are publicly available, we excluded γ- and δ-CoVs from the following analysis. Our sequence dataset contained the host family, the host genus, and the location information available in GenBank metadata or in original publications (as of May 5, 2021) for all of the CoV sequences. We referred to Volume 9 of the World Mammal Handbook (HMW) for taxonomic information on bat hosts (Wilson and Mittermeier, 2019). We used the API tool provided by GeoNames.org (http://api.geonames.org) to assign geographic coordinates to all bat CoV RdRp sequences that had not been annotated with latitude and longitude coordinates. To improve the quality of the location data for the CoV samples in our dataset, we excluded sequences that had only the country name as the location information. To increase the consistency of the sequence length of all bat CoV RdRp sequences and to distinguish the classifications of these CoVs at the genus level, all bat CoV RdRp sequences were aligned using MUSCLE and trimmed to 360 base pairs. Then, a phylogenetic tree was constructed based on all bat CoV RdRp sequences using the maximum likelihood method implemented in RAxML v8.2. X to separate the α- and β-CoV datasets, referring to the classification of the International Committee on the Taxonomy of Viruses (ICTV). To ensure the inclusion of as many CoVs as possible in the analysis, each CoV genus was included in a host dataset with a bat family and genus and a dataset with only the geographic region. Next, to avoid bias due to CoV sequence sampling, we randomly selected 30 sequences per host family, 15 sequences per host genus, or 10 sequences per grid cell to form random α- and β-CoV subsets. All subsequent analyses were based on complete datasets and the random subsets of the bat α- and β-CoVs.

Species distribution, climate, and human activity data

We first obtained a world map from Bzdt website (http://bzdt.ch.mnr.gov.cn/). To obtain more fine-scaled ecological maps with worldwide species distribution, climate, and human activity data, we established an equal-area grid system (Behrmann cylindrical equal-area projection) with a cell size...
of 385.9 km × 385.9 km, representing 148,953 km² grid areas, and numbered all grid cells individually (total of 3,485 grid cells). The global distribution ranges of the bat species were derived from the vectorized shapefile from the 2020 IUCN spatial database (www.iucnredlist.org/resources/spatial-data-download), which was rasterized into the grid system to obtain a global BSR map. MSR data were obtained from the NASA Shuttle Radar Topography Mission in 2013 (https://sedac.ciesin.columbia.edu/data/set/species-global-mammal-richness-2015). We also collected MAT and MAP data, with a spatial resolution of 2.5′, from the WorldClim database (https://worldclim.org/), and the UBU was available from consensus land-cover datasets (http://www.earthenv.org/landcover). Other human activity datasets were collected from the NASA Shuttle Radar Topography Mission data (http://srtm.csi.cgiar.org/); these factors were HPD, with a 30 arc-sec resolution in 2015 (in persons km⁻²); GAL in the year 2000 (dataset represents the proportion of land areas); the AE; and the UE. Based on the bat CoV sampling coordinate information, we plotted the coordinates using QGIS software and matched the locations to grid cells to calculate bat CoV phylogenetic diversity and bat CoV spatial dissemination among the geographic regions (grid cells). After all ecological data were mapped to each grid cell, we used QGIS software to extract the average of all of the values within each grid cell to represent the corresponding variable value for that grid cell (Table S24 in Supporting Information); these values were predictors of the spatial patterns of global bat CoV dissemination and phylogenetic diversity.

**Model selection and phylogenetic analysis**

Bayesian phylogenetic analysis using the Markov chain Monte Carlo (MCMC) approach was performed in BEAST 1.10.4 with the BEAGLE library (Drummond et al., 2012). The models were preliminarily tested using the following combinations to select the best-fit model: substitution model (HKY/GTR), codon partition scheme, molecular clock type (strict/lognormal uncorrelated relaxed clock/random local clock/ fixed local clock), and coalescent models (constant population size/exponential growth/Bayesian SkyGrid/ GMRF Bayesian Skyride/Bayesian Skyline). The best-fitting model was selected using the modified Akaike information criterion implemented in Tracer v1.7.1 (Rambaut et al., 2018). Model fit was assessed through log marginal likelihoods obtained by path sampling and stepping stone sampling analysis between the prior and the posterior (Xie et al., 2011). We also used TempEst v1.5.3 (Rambaut et al., 2016) to evaluate the temporal phylogenetic structure of the remaining α- and β-CoV datasets after removing the CoV sequences with unknown sampling years. The TEMPEST results revealed that the datasets did not contain sufficient time information to accurately estimate the time of divergence from the nearest common ancestor. Finally, we constructed best-fit models for all subsequent BEAST analyses, including an HKY substitution model with two codon partitions ((1+2), 3), a strict molecular clock, and a constant population size coalescent model with a fixed substitution rate of 1.0. Each MCMC analysis dataset was run sufficiently long to ensure convergence of the MCMC results as diagnosed using TreeAnnotator v1.8.2. The diagnostic criterion was an effective sample size of all parameters >200 after 10% burn-in.

**Ancestral state reconstruction and transition rates of global bat CoVs**

We used the Bayesian discrete phylogeographic approach in BEAST v1.10.4, with an asymmetric discrete trait substitution model, to construct the ancestral states of each node in the phylogenetic tree using two discrete host states, with the bat family, genus, and a discrete geographic region state for each grid cell. We used the TreeAnnotator v1.8.2 tool to summarize the trees in the form of maximum clade credibility trees annotated with discrete traits, and SpreaD3 software (Bielejec et al., 2016) was used to estimate interactive visualizations of the dispersal process and to compute the BF test to assess support for significant pairwise switches between discrete trait states using the BSSVS procedure (Lemey et al., 2014). We interpreted the BF results according to Lemey et al (Lemey et al., 2014) (BF>3: substantial support, BF>10: strong support, BF>30: very strong support, BF>100: decisive support). We also estimated the count of state-switching events (Markov jumps) along the branches of the phylogenetic tree globally (for the three discrete traits) and for each strongly supported (BF>10) transition between character states (for only bat families and regions) (Minin and Suchard, 2008). The convergence of the MCMC runs was confirmed using Tracer v1.7.1. These analyses were performed for each trait using data from the complete dataset and the random subsets.

**Quantifying the contributions of ecological factors on global bat CoV spatial dissemination**

To infer the potential contributions of BSR, MSR, GCD, and two climatic and five human-related factors that have been associated with bat CoV spatial dissemination in different geographic regions within each continent, we applied the GLM extension of the Bayesian phylogeographic model (Lemey et al., 2014) to the complete datasets of the bat CoV geographic region. We employed a BSSVS approach with a binomial prior probability distribution model for each predictor’s inclusion that reflected a 50% prior probability of no predictors being included. Using the same method, we calculated the correlation between the PD of the bat genera from the nearest common ancestor.
Timetree (http://www.timetree.org/) and the relationship of CoV transmission among the bat genera. Each predictor was log-transformed, normalized, and incorporated into these models. We determined the mean posterior probability, the BF support value, and the contribution of each predictor to the log-linear ratio matrix for each predictor.

Phylogenetic diversity of global bat CoVs

We used MPD and MNTD as indicators and their SES (Webb et al., 2002) with the R package picante (Kembel et al., 2010) to calculate the phylogenetic structures of the CoV lineages within each family, genus, and geographic region (grid cell). The MPD and MNTD represent the mean pairwise distances of all pairs of CoVs within a host or region and the mean pairwise distance of each CoV to its nearest phylogenetic neighbor within a host and a region, respectively. The MPD reflects the phylogenetic structures of the CoV lineages in the entire phylogenetic tree and indicates the overall degree of differentiation of the CoV lineages in the community, while the MNTD reflects the phylogenetic structure closer to the branch tips and indicates locally clustered taxa within the community. We also constructed a null model involving the random shuffling of tip labels 1,000 times across the tips of the entire phylogenetic tree to calculate SES MPD and SES MNTD values. SES MPD and SES MNTD values represent the difference between the phylogenetic distances of the observed community and the null community. Low and negative values indicate phylogenetic clustering, suggesting that the community was composed of CoVs with similar phylogeny. Positive and high values indicate overdispersion of the phylogenetic structure, implying that the community is composed of CoVs with distant phylogeny. Values close to 0 represent random dispersion, indicating that the CoVs in the community were random. All phylogenetic diversity analyses were carried out using complete dataset and random sampling for each bat taxon and geographic region.

The interregional MPD values (equivalent to phylogenetic β-diversity) were calculated using the comdist function in the picante package, corresponding to the MPD between all pairs of CoVs drawn from the regions. These were used to cluster regions in a dendrogram based on their phylogenetic similarity using the hclust function and the complete linkage method.

Quantifying the contributions of influencing factors on global bat CoV phylogenetic diversity

To explore the effects of factors affecting the spatial phylogenetic diversity of global bat α- and β-CoVs, we calculated the residuals from the semipart spatially explicit generalized linear mixed model (spaGLMM) implemented in the spaMM package (Rousset and Ferdy, 2014) in the R environment to assess the contributions of the same nine ecological factors to phylogenetic diversity using SES MPD and MNTD. For all of the above spaGLMM analyses, a correlation matrix according to the Matérn correlation function was assumed and fitted based on the longitude and latitude information of the center point of each grid cell when fitting the mixed model. The Matérn correlation function, which contains a scale parameter and a smoothness parameter, is widely applied to spatial correlation models by including exponential and squared exponential models as special cases. In the results of the semipart spaMM analyses, when the confidence interval of the estimated coefficient of an explanatory variable significantly deviated from zero, the variable was considered to have a significant effect on the level of phylogenetic diversity.

Compliance and ethics  The author(s) declare that they have no conflict of interest.

Acknowledgements This work was supported by the National Natural Science Foundation of China (32192420) and the Special Foundation for National Science and Technology Basic Research Program of China (2021FY1000304). We thank Dr. Ying Liu, Dr. Biao He and Dr. Yang Li for their suggestions on the revision of the manuscript.

References

Albery, G.F., Becker, D.J., Brierley, L., Brook, C.E., Christofferson, R.C., Cohen, L.E., Dallas, T.A., Eskew, E.A., Fagre, A., Farrell, M.J., et al. (2021). The science of the host-virus network. Nat Microbiol 6, 1483–1492.

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., et al. (2017). Global patterns in coronavirus diversity. Virus Evol 3, vex012.

Allen, T., Murray, K.A., Zambrana-Torrelio, C., Morse, S.S., Rondinini, C., Di Marco, M., Brett, N., Olival, K.J., and Daszak, P. (2017). Global hotspots and correlates of emerging zoonotic diseases. Nat Commun 8, 1124.

Beyer, R.M., Manica, A., and Mora, C. (2021). Shifts in global bat diversity suggest a possible role of climate change in the emergence of SARS-CoV-1 and SARS-CoV-2. Sci Total Environ 767, 145413.

Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., and Lemey, P. (2016). SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. Mol Biol Evol 33, 2167–2169.

Breed, A.C., Field, H.E., Smith, C.S., Edmonston, J., and Meers, J. (2010). Bats without borders: long-distance movements and implications for disease risk management. EcoHealth 7, 204–212.

Carlson, C.J., Albery, G.F., Merow, C., Trisos, C.H., Zipfel, C.M., Eskew, E.A., Olival, K.J., Ross, N., and Bansal, S. (2022). Climate change increases cross-species viral transmission risk. Nature 607, 555–562.

Chen, L., Liu, B., Yang, J., and Jin, Q. (2014). DBarVir: the database of bat-associated viruses. Database 2014, bau021.

Cui, J., Han, N., Streicker, D., Li, G., Tang, X., Shi, Z., Hu, Z., Zhao, G., Fontanet, A., Guan, Y., et al. (2007). Evolutionary relationships between bat coronaviruses and their hosts. Emerg Infect Dis 13, 1526–1532.

Cui, J., Li, F., and Shi, Z.L. (2019). Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 17, 181–192.

Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29, 1969–1973.

Ge, X.Y., Wang, Z., Zhang, W., Hu, B., Li, B., Zhang, Y.Z., Zhou, J.H., Luo, C.M., Yang, X.L., Wu, L.J., et al. (2016). Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. Virol
Sin 31, 31–40.
Hassaini, A., Khouider, S., Gembu, G.C., M. Goodman, S., Kadjio, B., Nesi, N., Pourrut, X., Nakouené, E., and Bonillo, C. (2015). The comparative phylogeography of fruit bats of the tribe Secontycterini (Chiroptera, Pteropodidae) reveals cryptic species diversity related to African Pleistocene forest refugia. Comptes Rendus Biologies 338, 197–211.
Islam, M.S., Sazzad, H.M.S., Satter, S.M., Sultana, S., Hossain, M.J., Hasan, M., Rahman, M., Campbell, S., Cannon, D.L., Ströther, U., et al. (2016). Nipah virus transmission from bats to humans associated with drinking traditional liquor made from date Palm Sap, Bangladesh, 2011–2014. Emerg Infect Dis 22, 664–670.
Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., and Daszak, P. (2008). Global trends in emerging infectious diseases. Nature 451, 990–993.
Kembel, S.W., Cowan, P.D., Helms, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., and Webb, C.O. (2010). Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464.
Latinne, A., Hu, B., Olival, K.J., Zhu, G., Zhang, L., Li, H., Chmura, A.A., Field, H.E., Zambrana-Torrelio, C., Epstein, J.H., et al. (2020). Origin and cross-species transmission of bat coronaviruses in China. Nat Commun 11, 4235.
Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C.A., Smith, D.J., Pybus, O.G., Brockmann, D., et al. (2014). Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLoS Pathog 10, e1003932.
Letko, M., Seifert, S.N., Olival, K.J., Plowright, R.K., and Munster, V.J. (2020). Bat-borne virus diversity, spillover and emergence. Nat Rev Microbiol 18, 461–471.
Li, T., Tang, X., Wu, C., Yao, X., Wang, Y., Lu, X., and Lu, J. (2020). The use of SARS-CoV-2-related coronaviruses from bats and pangolins to polarize mutations in SARS-CoV-2. Sci China Life Sci 63, 1608–1611.
Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H., et al. (2005). Bats are natural reservoirs of SARS-like coronaviruses. Science 310, 676–679.
Luby, S.P., Rahman, M., Hossain, M.J., Blum, L.S., Husain, M.M., Gurley, E., Khan, R., Ahmed, B.N., Rahman, S., Nahar, N., et al. (2006). Foodborne transmission of Nipah virus, Bangladesh. Emerg Infect Dis 12, 1888–1894.
Luo, J., Jiang, T., Lu, G., Wang, L., Wang, J., and Feng, J. (2013). Bat conservation in China: should protection of subterranean habitats be a priority? Orxys 47, 526–531.
Mao, X.G., Zhu, G.J., Zhang, S., and Rossiter, S.J. (2010). Pleistocene climatic cycling drives intra-specific diversification in the intermediate horseshoe bat (Rhinolophus affinis) in Southern China. Mol Ecol 19, 2754–2769.
Minin, V.N., and Suchard, M.A. (2008). Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol 56, 391–412.
Mollentze, N., and Streicker, D.G. (2020). Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. Proc Natl Acad Sci USA 117, 9423–9430.
Newbold, T. (2018). Future effects of climate and land-use change on terrestrial vertebrate community diversity under different scenarios. Proc R Soc B 285, 20180792.
Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L., and Lloyd-Smith, J.O. (2017). Pathways to zoonotic spillover. Nat Rev Microbiol 15, 502–510.
Rambaut, A., Drummond, A.J., Xie, D., Balee, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systat Bio 67, 901–904.
Rambaut, A., Lam, T.T., Max Carvalho, L., and Pybus, O.G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2, vew007.
Rouset, F., and Ferdy, J.B. (2014). Testing environmental and genetic effects in the presence of spatial autocorrelation. Ecography 37, 781–790.
Salah Uddin Khan, M., Hossain, J., Gurley, E.S., Nahar, N., Sultana, R., and Luby, S.P. (2010). Use of infrared camera to understand bats’ access to date palm sap: implications for preventing Nipah virus transmission. EcoHealth 7, 517–525.
Stoffberg, S., Jacobs, D.S., Mackie, I.J., and Matthee, C.A. (2010). Molecular phylogenetics and historical biogeography of Rhinolophus bats. Mol Phylogenet Evol 54, 1–9.
Streicker, D.G., Lemey, P., Velasco-Villa, A., and Rupprecht, C.E. (2012). Rates of viral evolution are linked to host geography in bat rabies. PLoS Pathog 8, e1002720.
Streicker, D.G., Turnelle, A.S., Vonhof, M.J., Kuzmin, I.V., McCracken, G.F., and Rupprecht, C.E. (2010). Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science 329, 676–679.
Temmam, S., Vonghayloith, K., Baquero, E., Munier, S., Bonomi, M., Regnault, B., Douangboubpha, B., Karami, Y., Christien, D., Sanamxay, D., et al. (2022). Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature 504, 330–336.
Watanabe, S., Masangkay, J.S., Nagata, N., Morikawa, S., Mizutani, T., Fukushima, S., Alviola, P., Omatu, T., Ueda, N., Iha, K., et al. (2010). Bat coronaviruses and experimental infection of bats, the Philippines. Emerg Infect Dis 16, 1217–1223.
Webb, C.O., Ackerly, D.D., McPeek, M.A., and Donoghue, M.J. (2002). Phylogenies and community ecology. Annu Rev Ecol Syst 33, 475–505.
Willoughby, A., Phelps, K., and Olival, K. (2017). A comparative analysis of viral richness and viral sharing in cave-roosting bats. Diversity 9, 35.
Wilson, D. E, Mittermeier, R.A. (2019). Handbook of the Mammals of the World. Vol. 9. Bats. Lynx Edicions, Barcelona.
Xie, W., Lewis, P.O., Fan, Y., Kuo, L., and Chen, M.H. (2011). Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst Biol 60, 150–160.
Yang, Q., Zhao, X., Lemey, P., Suchard, M.A., Bi, Y., Shi, W., Liu, D., Qi, W., Zhang, G., Stenseth, N.C., et al. (2020). Assessing the role of live poultry trade in community-structured transmission of avian influenza in China. Proc Natl Acad Sci USA 117, 5949–5954.
Yin, W., Mao, C., Luan, X., Shen, D.D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., et al. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science 368, 676–679.
Zhou, P., Fan, H., Lan, T., Yang, X.L., Shi, W.F., Zhang, W., Zhu, Y., Yin, W., Mao, C., Luan, X., Shen, D.D., Song, H., Zhao, R., Chen, Y., et al. (2021). Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. Cell 184, 4380–4391.e14.
Zhou, P., Fan, H., Lan, T., Yang, X.L., Shi, W.F., Zhang, W., Zhu, Y., Zhang, Y.W., Xie, Q.M., Mani, S., et al. (2018). Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. Nature 556, 255–258.

SUPPORTING INFORMATION

The supporting information is available online at https://doi.org/10.1007/s11427-022-2221-5. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.