Individual bat viromes reveal the co-infection, spillover and emergence risk of potential zoonotic viruses

Jing Wang[1,a], Yuan-fei Pan[2,a], Li-fen Yang[3,a], Wei-hong Yang[3], Chu-ming Luo[4], Juan Wang[3], Guo-peng Kuang[3], Wei-chen Wu[1], Qin-yu Gou[1], Gen-yang Xin[1], Bo Li[5], Huan-le Luo[4], Yao-qing Chen[4], Yue-long Shu[4], Deyin Guo[1,6], Zi-Hou Gao[3], Guodong Liang[7], Jun Li[8], Edward C. Holmes[9,*], Yun Feng[3,*], Mang Shi[1,*]

[1]The Centre for Infection and Immunity Studies, School of Medicine, Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University, Shenzhen, China.

[2]Ministry of Education Key Laboratory of Biodiversity Science and Ecological Engineering, National Observations and Research Station for Wetland Ecosystems of the Yangtze Estuary, Institute of Biodiversity Science and Institute of Eco-Chongming, School of Life Sciences, Fudan University, Shanghai, China

[3]Department of Viral and Rickettsial Disease Control, Yunnan Provincial Key Laboratory for Zoonosis Control and Prevention, Yunnan Institute of Endemic Disease Control and Prevention, Dali, Yunnan, China.

[4]School of Public Health (Shenzhen), Shenzhen campus of Sun Yat-sen University, Sun Yat-sen University, Shenzhen, China.

[5]Yunnan Key Laboratory of Plant Reproductive Adaptation and Evolutionary Ecology and Centre for Invasion Biology, Institute of Biodiversity, School of Ecology and Environmental Science, Yunnan University, Kunming, Yunnan, China

[6]Guangzhou National Laboratory, Guangzhou International Bio-Island, Guangzhou, Guangdong Province, China.

[7]State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China.

[8]Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China.

[9]Sydney Institute for Infectious Diseases, School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia.

[a]These authors contributed equally

*Corresponding authors: Edward C. Holmes (edward.holmes@sydney.edu.au); Yun Feng (ynfy428@163.com); and Mang Shi (shim23@mail.sysu.edu.cn)

Running Title: Comparisons of individual bat virome

ABSTRACT

Bats are reservoir hosts for many zoonotic viruses. Despite this, relatively little is known about the diversity and abundance of viruses within bats at the level of individual animals, and hence the frequency of virus co-infection and inter-species transmission. Using an unbiased meta-transcriptomics approach we characterised the mammalian associated viruses present in 149 individual bats sampled from Yunnan province, China. This revealed a high frequency of virus co-infection and species spillover among the animals studied, with 12 viruses shared among different bat species, which in turn facilitates virus recombination and reassortment. Of note, we identified five viral species that are likely to be pathogenic to humans or livestock, including a novel recombinant SARS-like coronavirus that is closely related to both SARS-CoV-2 and SARS-CoV, with only five amino acid differences between its receptor-binding domain sequence and that of the earliest sequences of SARS-CoV-2. Functional analysis predicts that this recombinant coronavirus can utilize the human ACE2 receptor such that it is likely to be of high zoonotic risk. Our study highlights the common occurrence of inter-species transmission and co-infection of bat viruses, as well as their implications for virus emergence.

Keywords

Bats, virome, SARS-CoV-2, meta-transcriptomics, emergence, Yunnan

INTRODUCTION

Bats (order Chiroptera) are hosts for a larger number of virus species than most mammalian orders[1], and are the natural reservoirs for several emerging viruses that cause infectious disease in human[2]. Recently, there has been considerable research effort directed toward exploring viral diversity in bats as a means to identifying potential zoonotic infections[3]. These studies have greatly expanded the diversity of known bat-borne viruses and identified an array of potentially emerging viruses. However, despite the growing body of work on bat viruses, little is known about the underlying drivers of virus diversity within these animals, nor of the extent patterns of viral co-infection and the frequency of viral spillover among bat species[4,5].

Current virus discovery studies typically pool individual bats by species or by sampling location (e.g., ref [6,7]). Although of great utility, this hinders mechanistic insights due to insufficient resolution. As such, studying the bat virome at the scale of individual animals can help us better understand the diversity and emergence of bat-borne viruses[4]. For example, the co-infection of phylogenetically related viruses within an individual host facilitates the occurrence of recombination[8], which may in turn have contributed to the emergence of a number of zoonotic viruses (e.g. SARS-CoV[9]). Importantly, the frequency of virus co-infection in bats can be resolved through the study of the viromes of individual animals. Resolution at the scale of individual animals is also required to better understand the frequency and determinants of virus spillover among bats[4,10]. For example, correlating host traits at the individual level with the probability of cross-species transmission is an important way to reduce confounding effects.

Many previous studies of bat viruses have preferentially targeted relatives of known human pathogens[3]. Although time and cost-effective, this necessarily limits our ability to discover novel zoonotic viruses. In contrast, other studies have utilized metagenomics approaches to explore the total bat virome (e.g., ref [7]), with meta-transcriptomic sequencing demonstrating great utility as a means to characterise the total diversity of viruses without *a priori* knowledge of which viruses are present[11,12].

Yunnan province in southwestern China has been identified as a hotspot for the diversity of bat species and bat-borne viruses. A number of highly pathogenic viruses have been detected there, including close relatives of SARS-CoV-2, such as bat viruses RaTG13[13], RpYN06[14], and RmYN02[15], as well as relatives of SARS-CoV, such as WIV1[16] and Rs4231[17]. It has been hypothesized that the presence of mixed roosts of bats in Yunnan (i.e., multiple bat species occupying the same roost) contributes to the frequent cross-species transmission of viruses, promoting their recombination and ultimately leading to transient spillovers or successful cross-species transmissions[17]. Thus, wild bat populations in Yunnan provide a unique opportunity to study the diversity, spillover and emergence risk of bat-borne viruses.

We performed intensive field sampling of individual wild bats in Yunnan province, China. In particular, we characterised the total mammal-associated virome of wild bats at the scale of individual animals using unbiased meta-transcriptomic sequencing. We then explored the

cross-species transmission of viruses among individual animals from different species and quantitatively tested how host phylogeny and geographic (i.e. sampling) location may impact the probability of cross-species transmission. Finally, we identified viruses of potentially high emergence risk and evaluated their pathogenic potential using a combination of phylogenetic analysis and *in sillico* molecular dynamics simulations.

RESULTS

Characterization of the bat viromes

Between 2015 and 2019, rectum samples were collected from 149 individuals of bat in six counties/cities, Yunnan province, China, representing six genera and 15 species (Fig. 1; Table S1). Total RNA was extracted and sequenced individually for each individual bat. Meta-transcriptomic sequencing yielded an average of 41,789,834 clean non-rRNA reads for each animal. In total, 1,048,576 contigs were *de novo* assembled from non-rRNA reads, among which the 36,029 contigs annotated as viruses were used for virome characterization.

We focused on characterizing the mammal-associated viromes of bats (Fig. 2), which include the RNA and DNA viral families or genera that are known to infect mammalian hosts (rather than those viruses more likely associated with bat diet or microbiome). In total, we identified 41 mammal-associated virus species belonging to 11 families (Table S2). Most of the viruses detected were RNA viruses, comprising 32 of the 41 viral species. The *Reoviridae* was the most prevalent viral family, present in 27.5% of individuals sampled, followed by the *Picornaviridae* (12.1%) and the *Coronaviridae* (8.7%) (Fig. 3C). The prevalence of the remaining viral families was relatively low (≤ 4%).

We next quantified the virus load and number of virus species for each individual bat (Fig. 3). Of the 149 individual bats analysed, 70 were positive for at least one virus species (positive rate 47.0%). We consider those with relatively high viral load (reads per million total reads > 1) as true positives. Among positive individuals, one-third were infected by more than one viral species (1.5 viral species per individual on average). The number of virus species per individual was uneven among host genera (Fig. 3B). We used Poisson regression to estimate the effect of host genera on the number of virus species. This revealed that *Rhinolophus* bats carried significantly more virus species per individual than average, whereas *Aselliscus* bats carried significantly less.

Cross-species transmission of viruses among bats

We identified 12 virus species that are shared among different bat species, accounting for the possibility of index-hopping (Fig. 4A). The 12 species identified belong to the *Coronaviridae* (4 species), *Reoviridae* (3), *Picornaviridae* (3), *Parvoviridae* (1) and *Polyomaviridae* (1). Rotavirus A type 1 (RVA1) and Mammalian orthoreovirus (MRV) had the broadest host range, being detected in five and four bat species, respectively, with both these viruses shared among the bat families Hipposideridae and Rhinolophidae. The remaining 10 viral species were only found in two bat species, and most were only shared among animals from the same host

genus, with the exception of bat coronavirus HKU10 (BtCoV-HKU10, present in *Hi. pomona* and *Rh. thomasi*) and Bat RVJ-like rotavirus BtSY1 (BtRV1, present in *Hi. larvatus* and *Rh. macrotis*).

The number of viral species found in two bat species ranged from zero to three. Partial Mantel tests showed that more closely phylogenetically related or closely geographically located bat individuals had more similar virome compositions and had more virus species in common (Table S4). For example, the viromes of *Rhinolophus* or *Hipposideros* bats form two network modules, in which individuals within the same genus are more inter-connected (i.e., shared more viruses) than individuals from different genera (Fig. 4A). A Poisson regression analysis showed that the number of virus species shared between pairs of bats was significantly associated with both the phylogenetic and the geographic distance of hosts, after controlling for the confounding effect of date of sampling (Fig. 4B, C).

Phylogenetic analysis identifies viruses of potentially high emergence risk

Phylogenetic analysis identified five viral species what were closely related to known human or livestock pathogens, which we termed “viruses of concern” (Fig. 5 and Table S5). The five viruses of concern belong to two viral families – the *Coronaviridae* (three species) and the *Reoviridae* (two species). Notably, all the five viruses were detected in more than one bat species, and their prevalence was relatively high, especially Mammalian orthoreovirus and Rotavirus A type 1 (Table S5).

The three coronaviruses were closely related to highly pathogenic viruses that infect humans or swine. A phylogenetic analysis using the conserved RdRp protein revealed that both Bat SARS-like virus BtSY1 and BtSY2 belong to the subgenus *Sarbecovirus* of betacoronaviruses and are closely related to human SARS-CoV (>90% nucleotide identity). Notably, other key functional genes (e.g., NTD, RBD, N) of Bat SARS-like virus BtSY2 were more closely related to SARS-CoV-2 (i.e., the early Wuhan-Hu-1 reference strain), indicative of a past history of recombination. We present further analysis of the evolutionary history and zoonotic potential of this virus below. The other coronavirus, which was Rhinolophus bat coronavirus HKU2-like, belonged to the genus *Alphacoronavirus* and was closely related to SADS-CoV of pigs in the RdRp gene.

The remaining two viruses of concern from the *Reoviridae* are known species -Mammalian orthoreovirus and Rotavirus A type 1. In total, three different types of Rotavirus A were identified according to the RdRp phylogeny. The nucleotide identity between their RdRp was less than 80%, so we demarcated them as three different types. Of these, Rotavirus A type 1 was most closely related to human Rotavirus A, while Rotavirus Type 2 and 3 belong to viral lineages associated with *Eonycteris* and *Rousettus* bats, respectively (Fig. S1).

As well as viruses of concern, 28 viral species can be classed as newly discovered viruses of mammals. The *Picornaviridae* (n = 8) contained the highest number of the newly discovered

172 viral species, followed by the *Astroviridae* (n = 5), *Parvoviridae* (n = 4), and *Caliciviridae* (n =
173 3).

174 **The evolutionary history and zoonotic potential of two SARS-related coronavirus**

175 We next evaluated the evolutionary history and zoonotic potential of the two SARS-related
176 coronaviruses detected in our samples: Bat SARS-like virus BtSY1 and BtSY2 (for simplicity
177 referred to as BtSY1 and BtSY2, respectively, in the following text) (Fig. 6). Phylogenetic trees
178 were estimated using the nucleotide sequences of key genes: the RdRp (RNA-dependent RNA
179 polymerase), N-terminal domain (NTD) and receptor-binding domain (RBD) of spike protein,
180 and the nucleoprotein (N). This analysis revealed that in the NTD, RBD and N gene trees,
181 BtSY1 clustered with SARS-CoV forming an “S-1” clade, while BtSY2 clustered with SARS-
182 CoV-2 forming the “S-2” clade. Notably, while BtSY1 remained in the S1 clade in the
183 phylogeny of the RdRp gene, BtSY2 also fell into the S-1 clade. Hence, BtSY2 appears to be a
184 recombinant between the S-1 and S-2 lineages.

185 At the scale of the whole genome, BtSY1 generally exhibited the highest genetic identity to
186 human SARS-CoV viruses (93%). Indeed, in comparisons to previously identified SARS-
187 related viruses (i.e., WIV16, Rs4231), BtSY1 shared highest identity with human SARS-CoV
188 viruses in ORF1b (nsp13 and nsp15) and the NTD, although it was relatively more distant in
189 ORF1a and the RBD, as well as in the S2 region of S gene (Fig. 6B). Specifically, it exhibited
190 98.13% similarity with SARS-CoV in the NTD, but only 88.61% identity in the RBD domain.

191 In marked contrast, BtSY2 shared 92% genetic identity with SARS-CoV-2 at the whole
192 genome scale, although with the occurrence of recombination. Indeed, we identified potential
193 recombination at positions 12035-20708bp, which encodes ORF1a (nsp7~nsp11) and ORF1b
194 (nsp12~nsp14), with this region instead showing strong sequence similarity to SARS-CoV
195 (92.3%). The remainder of its genome is very similar to SARS-CoV-2, particularly in the region
196 encoding the NTD and RBD (95.15% and 93.70%, respectively), although no furin cleavage
197 site was detected in the spike protein (Fig. 6B).

198 To evaluate the human-receptor-binding potential of BtSY2, we inferred the structure of its
199 RBD using a homology-modelling approach and performed molecular dynamics simulations
200 (Fig. 6). This revealed that there are only five amino-acid substitutions in the RBD in
201 comparison to the SARS-CoV-2 strain Wuhan-Hu-1, with three of these located at the interface
202 of RBD-hACE2 complex (i.e., the receptor-binding motif) (Fig. 6C). Molecular dynamics
203 simulations further revealed that the binding stability and energy of the RBD-hACE2 complex
204 are very similar between BtSY2 and SARS-CoV-2 Wuhan-Hu-1 (Fig. 6D), suggesting that
205 BtSY2 may be able to utilize human ACE2 receptor for cell entry.

206 **DISCUSSION**

207 We have characterised the mammalian-associated virome of individual bats. This revealed an
208 unexpectedly high frequency of co-infection, with approximately one-third of the virus-positive
209 individuals simultaneously infected by two or more viruses. The frequency of co-infection in

individual bats has seldom been investigated, and only a few studies have explored the co-infection of specific viral species using consensus PCR methods (e.g., paramyxovirus, ref [18]). As such, this study provides the first empirical evidence for co-infection using an unbiased omics approach. Co-infection is prerequisite for virus recombination or reassortment[8], and the gut microbiome can facilitate the recombination of enteric viruses[19]. Hence, the high frequency of co-infection observed here suggests that recombination and reassortment are very likely to happen within individual bats, which in turn may facilitate the emergence of zoonotic viruses[9].

Our results also revealed frequent virus spillover among different bat species, identifying 12 different viral species from different families that infect multiple host species. The ability of viruses to jump host species boundaries appears to be a near universal trait among viruses[20]. Our results are of note because they show that the probability of virus spillover among pairs of host individuals is negatively associated with host phylogenetic and geographic distance, thereby supporting the hypothesis that phylogenetically related or spatially closely located hosts share more viruses[21,22]. The frequent virus spillover among phylogenetically related or spatially co-located bats provides an opportunity for viromes of different bat species to exchange, further expanding genetic diversity of circulating viruses.

We identified two SARS-related coronaviruses in *Rhinolophus* bats (*Rh. marshalli*, *Rh. pusillus* *Rh. thomasi*, and *Rh. macrotis*) which we suggest are at particular risk for emergence. One of the SARSr-CoV - Bat SARS-like coronavirus BtSY2 (i.e., BtSY2) - is related to both SARS-CoV and SARS-CoV-2 and likely to have a history involving recombination. Notably, there are only five amino acid differences in the receptor-binding domain (RBD) of spike protein of this virus compared with SARS-CoV-2 strain Wuhan-Hu-1[23], which makes it the closest relative to SARS-CoV-2 found in China in this particular genomic region. In contrast, the nsp7~nsp11 proteins of ORF1a and nsp12~nsp14 protein of ORF1b were closely related to SARS-CoV, indicating that these genes were likely to be acquired from another SARSr-CoV. The remainder of its genome was closely related to SARS-CoV-2 and to several bat coronavirus previously found in Yunnan, including RaTG13[13], RmYN02[15], and RpYN06[14] that are all close relatives of SARS-CoV-2. Together, these findings strongly suggest that virus spillover and co-infection in related bat species contribute to the recombination of potentially pathogenic coronavirus and could possibly facilitate virus emergence in other species.

Functional analysis indicated that Bat SARS-like coronavirus BtSY2 likely has the ability to the bind human ACE2 receptor, and even has slightly higher affinity than SARS-CoV-2 Wuhan-Hu-1. Three of the five substitutions in the RBD - Q498H, N501Y and H519N - have been reported to increase affinity to human ACE2[24], and notably, the N501Y substitution is present in the Alpha, Beta, Gamma and Omicron variants of SARS-CoV-2. In addition, we found that the nsp7-nsp14 proteins (in which nsp12 is the replicase, i.e., RdRp) of BtSY2 were closely related to those of SARS-CoV. A comparative study showed that SARS-CoV can replicate more rapidly than SARS-CoV-2 *in vitro*[25], while another suggested that nsp14 is likely associated with virulence[26]. Hence, these data tentatively suggest that BtSY2 may be able to replicate

rapidly with similar virulence as SARS-CoV. Although this issue merits further consideration, this virus is potentially of high risk of emergence and so should be monitored carefully.

We identified another four viruses of concern, likely to be pathogenic in humans or livestock. Bat SARS-like virus BtSY1 is closely related to SARS-CoV[27,28]. Rhinolophus bat coronavirus HKU2-like is closely related to SADS-CoV, which causes severe diarrhea and death in swine[29,30], Rotavirus A causes diarrhea in humans[31,32], while Mammalian orthoreovirus is known to have a broad host range and cause diarrhea in swine[33,34]. Interestingly, all these viruses of concern were found in more than one bat species in our samples, suggesting that these potentially zoonotic viruses may have a broader host range or have a higher rate of spillover than other viruses.

METHODS

Ethics Statement

This research, including the procedures and protocols of specimen collection and processing, was reviewed and approved by the Medical Ethics Committee of the Yunnan Institute of Endemic Diseases Control and Prevention. (No. 20160002).

Sample collection

A total of 149 rectum samples from bats were collected from six counties/cities in Yunnan province, China between 2015 and 2019. Bats were trapped using net traps and were primarily identified according to morphological criteria and confirmed by a barcode gene (COI) in the meta-transcriptomics analysis. The bats collected belonged to 15 species. The majority were from the genus *Rhinolophus* (n=54) and comprised *Rhinolophus pusillus* (n=16), *Rhinolophus thomasi* (n=14), *Rhinolophus stheno* (n=12), *Rhinolophus marshalli* (n=7), *Rhinolophus pearsonii* (n=2), *Rhinolophus macrotis* (n=2), and *Rhinolophus affinis* (n=1). The genus *Hipposideros* (n=26) animals comprised *Hipposideros larvatus* (13), *Hipposideros armiger* (11), and *Hipposideros pomona* (2). The genus *Rousettus* (n=23) animals comprised *Rousettus leschenaultia* (n=18) and *Rousettus amplexicaudatus* (n=5). The *Aselliscus* (n=35), *Cynopterus* (n=9) and *Eonycteris (n=2)* genera animals only contained *Aselliscus stoliczkanus* (n=35), *Cynopterus sphinx* (n=9)*, and Eonycteris spelaea* (n=2), respectively. All rectum samples were collected from each individual bat and then stored at -80°C until use.

RNA extraction, library preparation and sequencing

All samples from each individual bat were homogenized using grinding bowls and rods in MEM medium. The homogenized samples were then centrifuged at 12,000 rpm for 30 min at 4°C to obtain supernatant. Total RNA extraction and purification were performed using the RNeasy Plus universal mini kit (QIAGEN) according to the manufacturer's instructions. RNA sequencing library construction and ribosomal RNA depletion were performed using the Zymo-Seq RiboFree™ Total RNA Library Kit (Zymo Research). Paired-end (150 bp) sequencing of the 149 dual-indexed libraries was performed on an Illumina NovaSeq platform.

Virus discovery pipeline

Viral genomes assembly and annotation

Raw paired-end sequence reads were first quality controlled, and rRNA reads were removed by mapping against the rRNA database downloaded from the SILVA website (https://www.arb-silva.de/) using Bowtie2. The clean reads were then *de novo* assembled into contigs using MEGAHIT (version 1.2.8)[35]. We performed a blastx search of contigs against the NCBI nr database using Diamond (version 0.9.25)[36] to roughly classify the sequences by kingdom. The e-value was set at 0.001 to achieve high sensitivity while reducing false positives. Those contigs classified as viruses were used for later analysis. Some viral contigs with unassembled overlaps were merged using SeqMan in the Lasergene software package (version 7.1)[37]. We searched for ORFs in each viral genome using the NCBI ORFfinder (https://ftp.ncbi.nlm.nih.gov/genomes/TOOLS/ORFfinder/), with the genetic code set to standard and with ATG as the only start codon. Then we performed a blastp search against the nr database and manually annotated the viral contigs according to the results.

Quantification of virus abundance

We quantified the abundance of each virus in each library as the number of viral reads per million non-rRNA reads (i.e., RPM) by mapping clean non-rRNA reads of each library to the corresponding contigs. To reduce false positives, we applied an abundance threshold of 1 RPM. We further reduced the number of possible false positives from index hopping using the same criterion as described in Shi, et al. [38].

PCR confirmation of virus genomes

The genome sequence of Bat SARS-like virus BtSY2 was obtained and confirmed by PCR amplification and Sanger sequencing. The WTA product was performed using the Complete Whole Transcriptome Amplification Kit (WTA2)[39] (Sigma-Aldrich, St. Louis, MO), with the PCR reaction then undertaken using a set of self-designed primer pairs based on the obtained reads. To confirm the recombination breakpoints, long fragments were obtained using the SuperScript IV Reverse Transcriptase and Expand Long Template PCR System.

Viral species demarcation and phylogenetic analysis

Viral species were identified based using sequences of the conserved replicase proteins (RNA viruses: RdRp, *Polyomaviridae*: LTAg, *Anelloviridae*: ORF1 protein, *Parvoviridae*: NS1, and other DNA viruses: DNA pol). We applied a 90% cut-off of amino acid sequence similarity to demarcate different virus species. The viruses were aligned using MAFFT (version 7.48)[40] and ambiguously aligned regions were removed using TrimAl[41]. Phylogenetic trees were then estimated by the maximum likelihood (ML) approach implemented in PhyML version 3.0[42], employing the LG model of amino acid substitution and the Subtree Pruning and Regrafting (SPR) branch-swapping algorithm. For SARS-related viruses, nucleotide sequences of RdRp,

NTD, RBD and N genes were used for phylogenetic analysis, employing the GTR substitution model.

Recombination analysis of SARS-related viruses

The recombination analysis of SARS-related viruses was performed using similarity plots as implemented in Simplot 3.5.1[43]. The nucleotide sequences of the SARS-related viruses were analyzed with reference strains obtained from GenBank, comprising SARS-CoV Tor2, SARS-CoV-2 Wuhan-Hu-1, as well as some of the closest bat SARSr-CoVs identified so far: Rs4231, WIV16, RaTG13, and BANAL-20-52.

Protein structure modelling and molecular dynamics (MD) simulations

Homology modelling

We built homology models of the Bat SARS-like virus BtSY2 RBD-hACE2 protein complex with MODELLER (version 10.3)[44], using the known structure of a SARS-CoV-2 RBD-hACE2 complex (PDB ID: 6M0J, resolution 2.45 Å)[45] as a template. The similarity between Bat SARS-like virus BtSY2 RBD and the template was 97.4%. We removed all NAG and water molecules in the template, and kept the zinc and chloride atoms. We built 100 homology models and selected the top three models based on normalized DOPE score[46] for the later MD simulations.

MD simulation

We used the CHARMM-GUI webservice[47] to prepare inputs for MD simulations. The three homology models described above and a SARS-CoV-2 RBD-hACE2 complex with known structure (PDB ID: 6M0J) were input to the CHARMM-GUI solution builder pipeline. The four systems were solvated in a water box of 13.5 nm × 9.2 nm × 8.3 nm, with KCl at the concentration of 0.15M. We used CHARMM36m force field[48] for protein and ions, and TIP3P model[49] for water.

The models processed by CHARMM-GUI were then used as inputs to GROMACS (version 2022.3)[50] for MD simulations. The following steps were performed sequentially for each model: (1) energy minimization, (2) 1-ns-long equilibration in NPT ensemble, and (3) 1-ns-long equilibration in NVT ensemble. The temperature and pressure were set to 300K and 1 atm, respectively. We then performed production simulations in NVT ensemble. Production simulation for the top homology model was 1000 ns long, and we performed another two 500-ns-long simulations for the rest two homology models as replications. Similarly, we performed one 1000-ns-long production simulation for the SARS-CoV-2 RBD-hACE2 complex (PDB ID: 6M0J) and two 500-ns-long replications. The settings of these simulations were the same as those described in ref [51].

Analysis of the MD data

We performed two sets of analyses on the data retrieved from MD. First, we evaluated the stability of RBD-hACE2 binding by measuring deviation of the protein backbones (measured as RMSD) in the duration of simulations, using PLUMED (version 2.7.4)[52]. The backbone

RMSD were calculated with respect to energy-minimized structure of each model. We also calculated RMSD separately for RBD, hACE2 and the RBD-hACE2 interface (residues within 0.8 nm to the other subunit in the 6M0J model). Second, we estimated and compared the binding energy of RBD-hACE2 complex using FoldX (version 4)[53]. We visualized the structure of RBD-hACE2 complex using PyMOL (version 2.4.2)[54].

Statistical analysis

All statistical analyses were performed in R (version 4.2.0) [55]

Comparing the number of virus species per individual between bat genera

To determine whether specific bat genera harbor more virus species per individual than others, we used a Poisson regression to fit the effect of host genera on number of virus species per individual. We extracted and visualized the estimated effect size of each host genus and its 95% CI. An effect is considered to be significant if its 95% does not include zero.

Analysis of cross-species transmission of viruses

To show possible cross-species virus transmission, we visualized the virus-sharing pattern among different bat species using a bipartite network. In this network, a node is either a host or a virus species, and an edge linking a host node and a virus node indicates the presence of that virus in that host. We performed edge betweenness clustering on such network to find network modules, which are subset of nodes such that connections between these nodes are denser than outside of the subset, using the igraph package[56] in R. A biological interpretation of a network module is that host species within the same module shared more viruses than outside that module.

We performed two sets of statistical tests to further quantitatively evaluate cross-species transmission of viruses among bats. First, we assessed the strength of the correlation of virome composition with both host phylogeny and geographic location using partial Mantel tests implemented in the ecodist package[57]. Differences between virome compositions were represented by Bray-Curtis distance, and phylogenetic distance between hosts was measured as the sum of branch length of pairs of hosts in the COI gene tree. We then used Poisson regression to estimate the effects of (1) phylogenetic distance between hosts, and (2) geographic distance between sample locations on the number of shared virus species between pairs of hosts, including the time intervals between sampling dates to control for its confounding effect.

DATA AND CODE AVAILABILITY

All meta-transcriptomic sequencing reads have been deposited in SRA database under the project accession (accession id here). The viral genome sequences determined here have been deposited at NCBI GenBank (accession id here). Sample metadata and other materials required to reproduce our results have been deposited together with code and scripts in a GitHub repository (https://github.com/Augustpan/Individual-Bat-Virome). Viral genome sequences are temporarily stored in the above GitHub repository for reviewing, until the NCBI GenBank accessions become available.

REFERENCES

1 Mollentze, N. & Streicker, D. G. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proc Natl Acad Sci U S A* **117**, 9423-9430, doi:10.1073/pnas.1919176117 (2020).

2 Hayman, D. T. S. Bats as Viral Reservoirs. *Annual Review of Virology* **3**, 77-99, doi:10.1146/annurev-virology-110615-042203 (2016).

3 Van Brussel, K. & Holmes, E. C. Zoonotic disease and virome diversity in bats. *Current Opinion in Virology* **52**, 192-202, doi:https://doi.org/10.1016/j.coviro.2021.12.008 (2022).

4 Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus diversity, spillover and emergence. *Nat Rev Microbiol* **18**, 461-471, doi:10.1038/s41579-020-0394-z (2020).

5 Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nat Rev Microbiol*, doi:10.1038/s41579-021-00652-2 (2021).

6 Li, Y. *et al.* Virome of Bat Guano from Nine Northern California Roosts. *Journal of Virology* **95**, e01713-01720, doi:10.1128/JVI.01713-20 (2021).

7 Wu, Z. *et al.* Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *The ISME Journal* **10**, 609-620, doi:10.1038/ismej.2015.138 (2016).

8 Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? *Nature Reviews Microbiology* **9**, 617-626, doi:10.1038/nrmicro2614 (2011).

9 Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* **17**, 181-192, doi:10.1038/s41579-018-0118-9 (2019).

10 Plowright, R. K. *et al.* Ecological dynamics of emerging bat virus spillover. *Proc Biol Sci* **282**, 20142124, doi:10.1098/rspb.2014.2124 (2015).

11 Shi, Y. *et al.* Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China Plain. *Microbiome* **6**, doi:ARTN 27 10.1186/s40168-018-0409-4 (2018).

12 Shi, M. *et al.* Redefining the invertebrate RNA virosphere. *Nature* **540**, 539-543, doi:10.1038/nature20167 (2016).

13 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

14 Zhou, H. *et al.* Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* **184**, 4380-4391.e4314, doi:https://doi.org/10.1016/j.cell.2021.06.008 (2021).

15 Zhou, H. *et al.* A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Current Biology* **30**, 2196-2203.e2193, doi:https://doi.org/10.1016/j.cub.2020.05.023 (2020).

16 Ge, X.-Y. *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535-538, doi:10.1038/nature12711 (2013).

17 Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLOS Pathogens* **13**, e1006698, doi:10.1371/journal.ppat.1006698 (2017).

18 Peel, A. J. *et al.* Synchronous shedding of multiple bat paramyxoviruses coincides with peak periods of Hendra virus spillover. *Emerging Microbes & Infections* **8**, 1314-1323, doi:10.1080/22221751.2019.1661217 (2019).

19 Erickson, A. K. *et al.* Bacteria Facilitate Enteric Virus Co-infection of Mammalian Cells and Promote Genetic Recombination. *Cell Host Microbe* **23**, 77-88 e75, doi:10.1016/j.chom.2017.11.007 (2018).

20 Geoghegan, J. L., Duchene, S. & Holmes, E. C. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog* **13**, e1006215, doi:10.1371/journal.ppat.1006215 (2017).

21 Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nat Rev Microbiol* **15**, 502-510, doi:10.1038/nrmicro.2017.45 (2017).

22 Streicker, D. G. *et al.* Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* **329**, 676-679, doi:10.1126/science.1188836 (2010).

23 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269, doi:10.1038/s41586-020-2008-3 (2020).

24 Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e1220, doi:10.1016/j.cell.2020.08.012 (2020).

25 Ogando, N. S. *et al.* SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *Journal of General Virology* **101**, 925-940, doi:https://doi.org/10.1099/jgv.0.001453 (2020).

26 Gussow, A. B. *et al.* Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. *Proceedings of the National Academy of Sciences* **117**, 15193-15199, doi:10.1073/pnas.2008176117 (2020).

27 Drosten, C. *et al.* Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *New England Journal of Medicine* **348**, 1967-1976, doi:10.1056/NEJMoa030747 (2003).

28 Guan, Y. *et al.* Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science* **302**, 276-278, doi:10.1126/science.1087139 (2003).

29 Gong, L. *et al.* A New Bat-HKU2–like Coronavirus in Swine, China, 2017. *Emerging Infectious Disease journal* **23**, 1607, doi:10.3201/eid2309.170915 (2017).

30 Zhou, P. *et al.* Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* **556**, 255-258, doi:10.1038/s41586-018-0010-9 (2018).

31 Parashar, U., Gibson, C., Bresee, J. & Glass, R. Rotavirus and Severe Childhood Diarrhea. *Emerging Infectious Disease journal* **12**, 304, doi:10.3201/eid1202.050006 (2006).

32 Anderson, E. J. & Weber, S. G. Rotavirus infection in adults. *The Lancet Infectious Diseases* **4**, 91-99, doi:https://doi.org/10.1016/S1473-3099(04)00928-4 (2004).

33 Thimmasandra Narayanappa, A. *et al.* A Novel Pathogenic Mammalian Orthoreovirus from Diarrheic Pigs and Swine Blood Meal in the United States. *mBio* **6**, e00593-00515, doi:10.1128/mBio.00593-15 (2015).

34 Jiang, R.-D. *et al.* Bat mammalian orthoreoviruses cause severe pneumonia in mice. *Virology* **551**, 84-92, doi:https://doi.org/10.1016/j.virol.2020.05.014 (2020).

35 Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676, doi:10.1093/bioinformatics/btv033 (2015).

36 Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* **18**, 366-368, doi:10.1038/s41592-021-01101-x (2021).

37 Clewley, J. P. Macintosh sequence analysis software. *Molecular Biotechnology* **3**, 221-224, doi:10.1007/BF02789332 (1995).

38 Shi, M. *et al.* Total infectome characterization of respiratory infections in pre-COVID-19 Wuhan, China. *PLOS Pathogens* **18**, e1010259, doi:10.1371/journal.ppat.1010259 (2022).

39 Zhu, Y. *et al.* Clinical utility of a novel urine-based gene fusion TTTY15-USP9Y in predicting prostate biopsy outcome. *Urologic Oncology: Seminars and Original Investigations* **33**, 384.e389-384.e320, doi:https://doi.org/10.1016/j.urolonc.2015.01.019 (2015).

40 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

41 Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973, doi:10.1093/bioinformatics/btp348 (2009).

42 Guindon, S. & Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* **52**, 696-704, doi:10.1080/10635150390235520 (2003).

43 Lole Kavita, S. *et al.* Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *Journal of Virology* **73**, 152-160, doi:10.1128/JVI.73.1.152-160.1999 (1999).

44 Sali, A. Comparative protein modeling by satisfaction of spatial restraints. *Molecular Medicine Today* **1**, 270-277, doi:https://doi.org/10.1016/S1357-4310(95)91170-7 (1995).

45 Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215-220, doi:10.1038/s41586-020-2180-5 (2020).

46 Shen, M.-y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Science* **15**, 2507-2524, doi:https://doi.org/10.1110/ps.062416606 (2006).

47 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859-1865, doi:https://doi.org/10.1002/jcc.20945 (2008).

48 Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**, 71-73, doi:10.1038/nmeth.4067 (2017).

49 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935, doi:10.1063/1.445869 (1983).

50 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25, doi:https://doi.org/10.1016/j.softx.2015.06.001 (2015).

51 Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330-336, doi:10.1038/s41586-022-04532-4 (2022).

52 Bonomi, M. *et al.* Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* **16**, 670-673, doi:10.1038/s41592-019-0506-8 (2019).

53 Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Research* **33**, W382-W388, doi:10.1093/nar/gki387 (2005).

54 Schrödinger, L. L. C.. The PyMOL Molecular Graphics System, Version 1.8 (2015).

55 R. C. Team. R: A Language and Environment for Statistical Computing (2022).

56 Csárdi, G. & Nepusz, T. The igraph software package for complex network research (2006).

57 Goslee, S. C. & Urban, D. L. The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *Journal of Statistical Software* **22**, 1 - 19, doi:10.18637/jss.v022.i07 (2007).

ACKNOWLEDGMENTS

We thank the Alibaba Cloud Computing Co. Ltd. for providing the computational resources for rapid data processing, and the local Centers for Disease Control and Prevention in six trapping sites for their assistance in specimen collection.

MS was supported by Shenzhen Science and Technology Program (KQTD20200820145822023), Guangdong Province “Pearl River Talent Plan” Innovation and Entrepreneurship Team Project (2019ZT08Y464), and Health and Medical Research Fund (COVID190206). YF was supported by the National Natural Science Foundation of China (3156004), Yunnan Reserve Talents for Academic and Technical Leaders of Middle-Aged and Young People (2019HB052), Jianguo Xu Academician Workstation (NO. 2018IC155). ECH was supported by an Australian Research Council Australia Laureate Fellowship (FL170100022) and by InnoHK funding from the Innovation and Technology Commission of Hong Kong. GDL was supported by United States National Institutes of Health U01 AI151810.

AUTHOR CONTRIBUTIONS

Conceptualization, E.C.H., Y.F., and M.S.; Methodology, J.W., Y.-F.P., W.-H.Y., C.-M.L., W.-C.W., B.L., Y.-L.S., D.-Y.G., G.-D.L., J.L., Y.F., E.C.H., and M.S.; Investigation, J.W., Y.-F.P., L.-F.Y., W.-H.Y, C.-M.L., J.W., G.-P.K., W.-C.W., Q.-Y.G., G.-Y.X., H.-L.L., Y.-Q.C. E.C.H., Y.F., and M.S.; Writing – Original Draft, Y.-F.P.; Writing – Review and Editing, J.W., Y.-F.P., L.-F.Y., W.-H.Y, C.-M.L., J.W., G.-P.K., W.-C.W., Q.-Y.G., G.-Y.X., H.-L.L., Y.-Q.C., Y.-L.S., D.-Y.G., Z.-H.G., G.-D.L., J.L., E.C.H. and M.S. Funding Acquisition, J.L., Y.F. and M.S.;

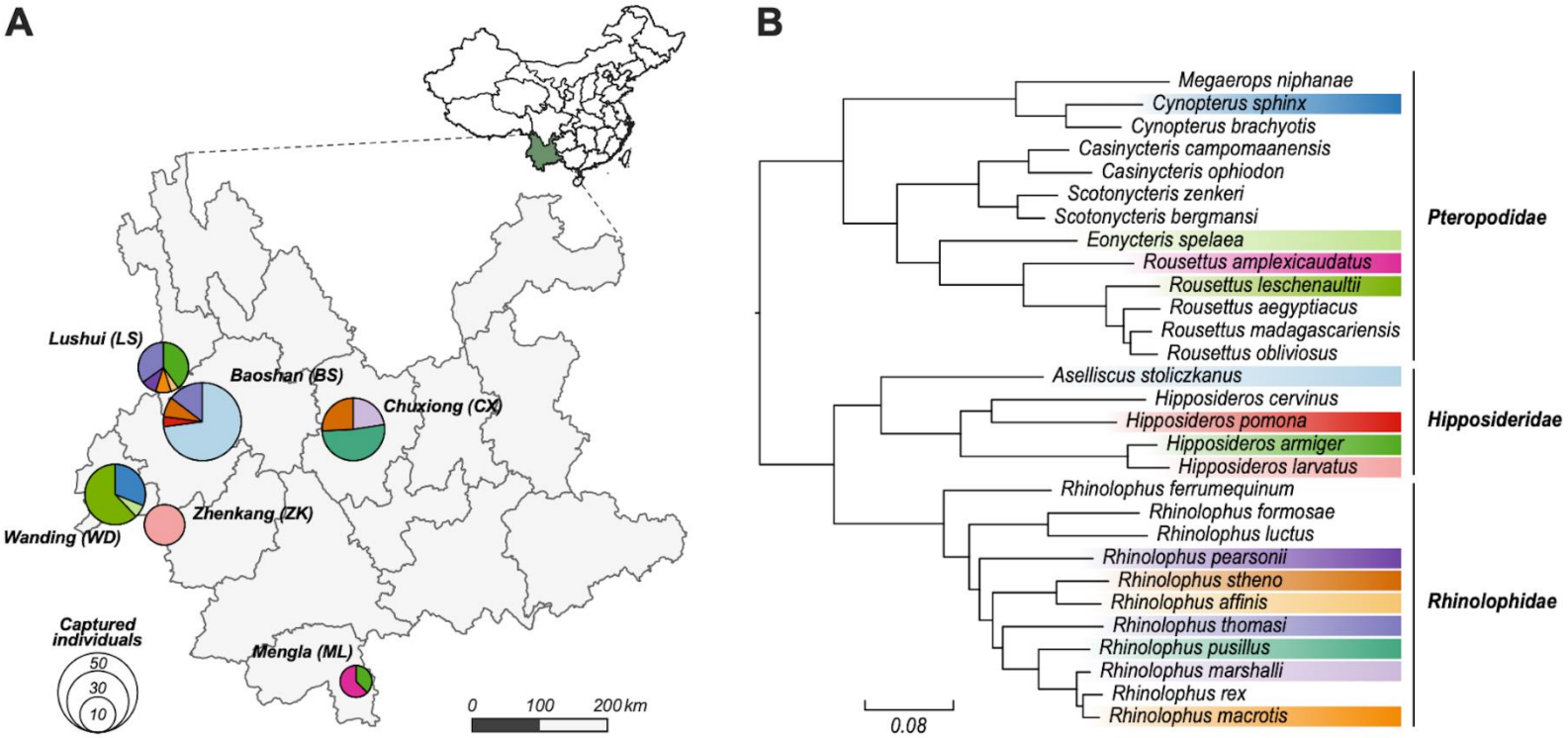
Resources (sampling), L.-F.Y., W.-H.Y., Z.-H.G., G.-D.L., and Y.F.; Resources (computational), J.L., E.C.H., and M.S.; Supervision, Z.-H.G., G.-D.L., E.C.H., Y.F., and M.S..

COMPETING INTERESTS

The authors declare no competing interests.

562 **FIGURES**

563



564

565

566 **Fig. 1 | Overview of the samples analysed in this study.** (**A**) Locations in Yunnan province
567 China where bat samples were taken. Pie charts indicate the composition of bat species
568 sampled at each location, while the total area of the pies are proportional to number of
569 captured individuals. Colours indicate different bat species, which are consistent with the
570 colouring scheme in plot B. (**B**) Phylogeny of bats, including those sampled as part of this
571 study. The tree was estimated using nucleotide sequences of bat COI gene utilising a
572 maximum likelihood (ML) method. Coloured strips indicate the bat species sampled in this
573 study.

574

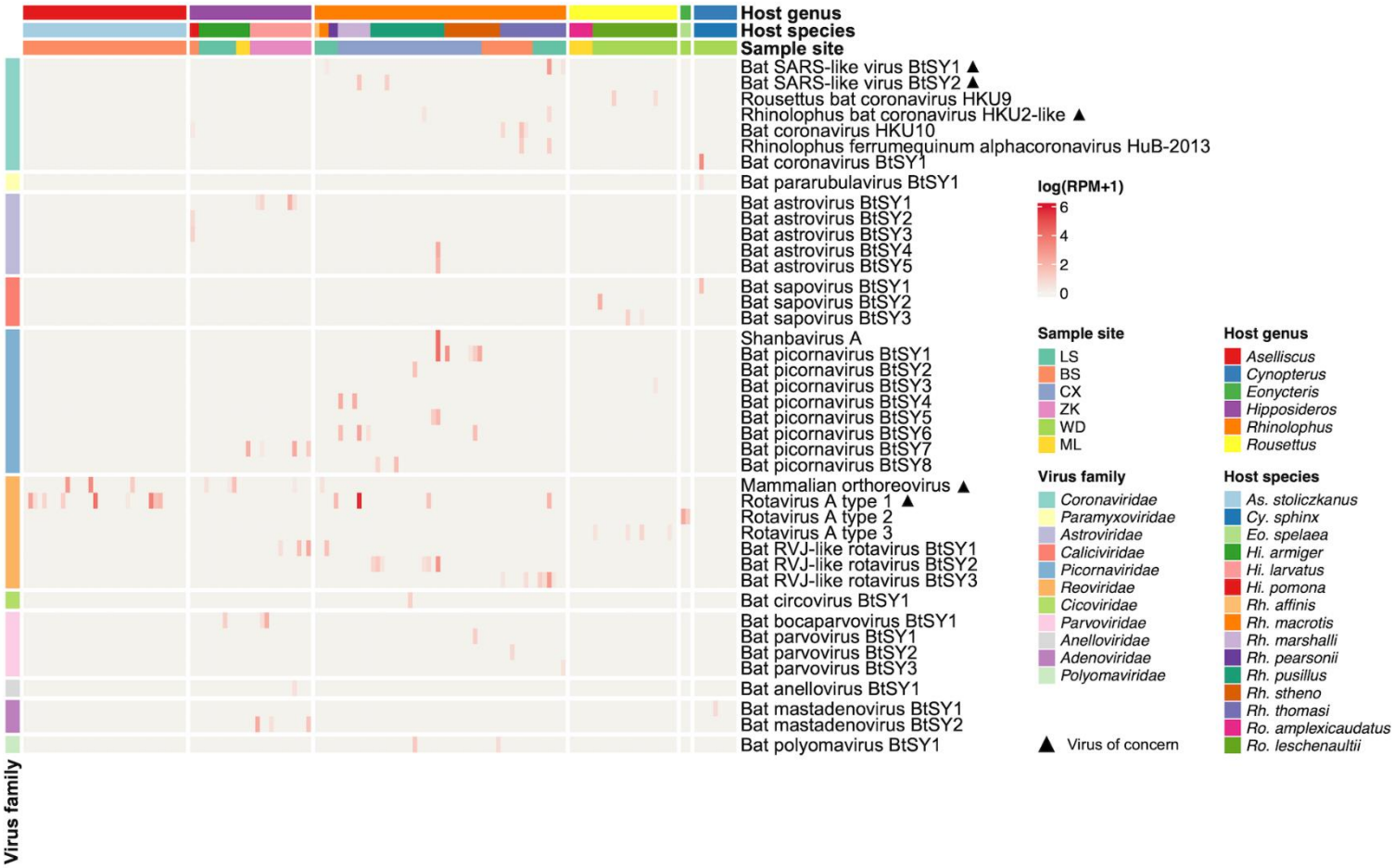
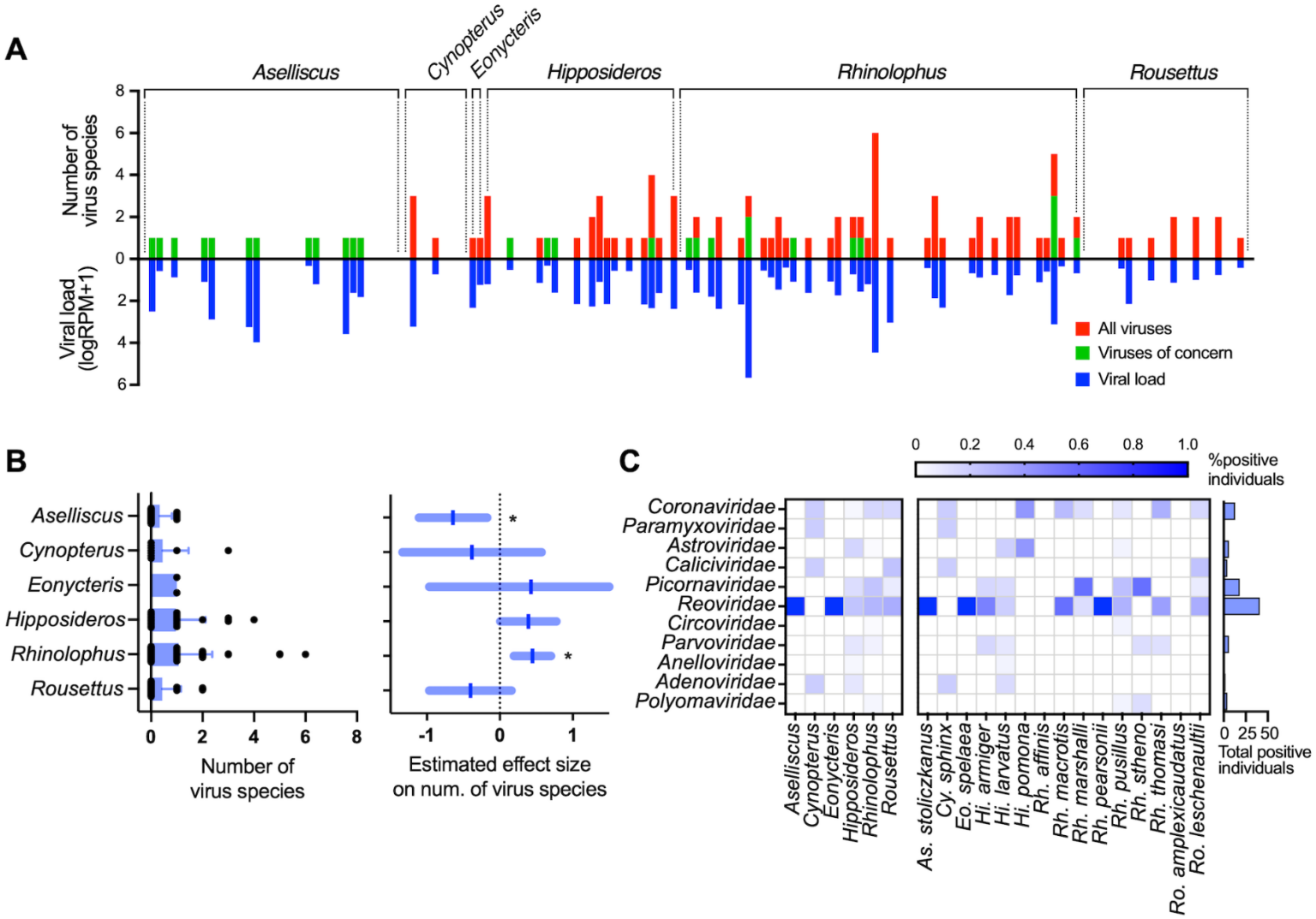


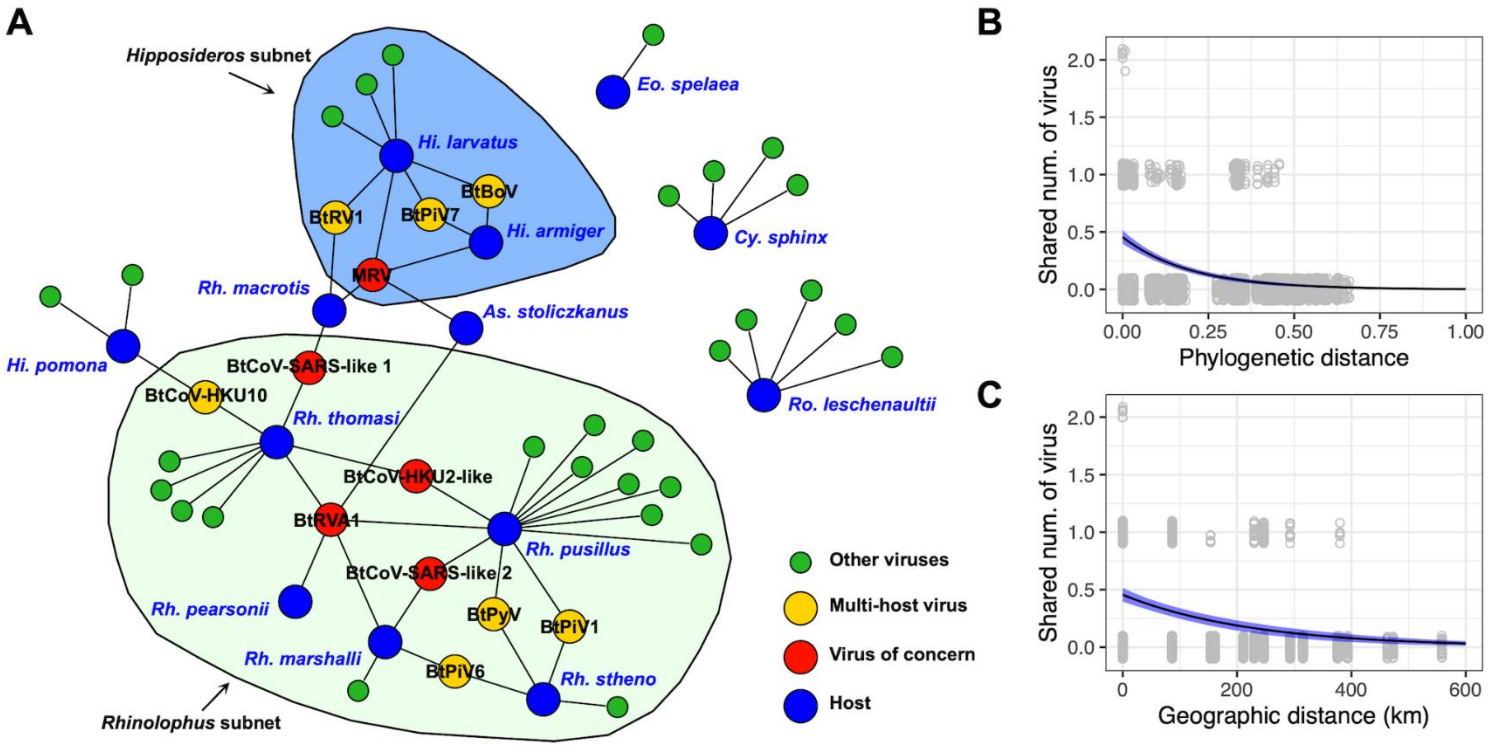
Fig. 2 | Characterization of the mammal-associated virome of bats. The heatmap displays the distribution and abundance of mammal-associated viruses in individual bats. Each column represents an individual bat, while each row represents a virus species. The abundance of viruses in each individual is represented as a logarithm of the number of mapped reads per million total reads (RPM). Sampling site, host taxonomy (species and genus) and virus taxonomy are shown as coloured strips at top and left, respectively. Black triangle marks indicate “viruses of concern”, defined as those that are closely related to known human or livestock pathogens (>90% amino acid similarity in RNA-dependent RNA polymerase).



585

586 **Fig. 3 | Comparison of mammal-associated virus diversity among different bat taxa.** (**A**)
587 Virus load and the number of virus species in individual bats. Red bars, the total number of
588 mammal-associated viruses per host. Green bars, number of viruses of concern per host. Blue
589 bars represent viral load per host as logarithm of the sum of total viral RPM. (**B**) Left,
590 comparison of the number of viruses per individual host among six bat genera (mean+SD).
591 Right, estimated effect size of each bat genus on the number of virus species per individual bat
592 by Poisson regression (mean±95%CI). Stars indicate significant effects (in which zero is not
593 included in the 95%CI). (**C**) Comparison of the prevalence of 11 viral families among different
594 host genera (left block) and species (right block).

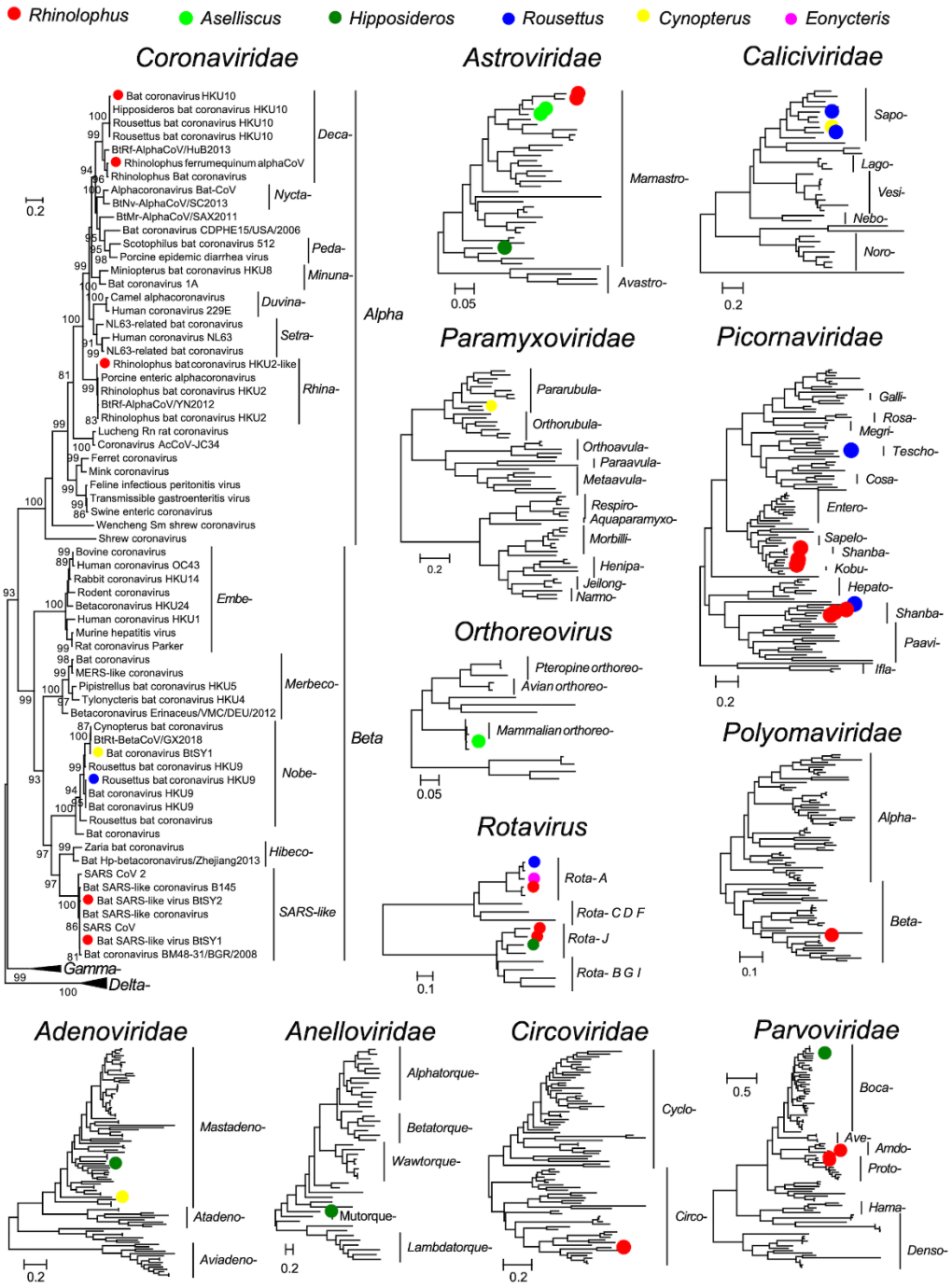
595



596

597 **Fig. 4 | The virus-sharing network of bats.** (**A**) The virus-sharing network reveals
598 connectivity among viromes of different bat taxa. Viruses of concern and putative cross-
599 species transmissions are shown in different colours. Two network modules (subnets) were
600 detected with a network betweenness-based criterion and are visualised by coloured areas.
601 (**B, C**) The relationship between the number of shared viruses with phylogenetic (B) or
602 geographic distance (C) between pairs of host individuals. Phylogenetic distance is calculated
603 as the sum of phylogenetic tree branch length between a pair of hosts, and the tree was
604 estimated with nucleotide sequences of the COI gene employing a maximum likelihood
605 method. The line and blue area is the estimated partial effect and standard error of
606 phylogenetic or geographic distance by Poisson regression.

607



608

609 **Fig. 5 | Evolutionary relationships of 11 viral families detected in our study.** Phylogenetic
610 trees were estimated using a maximum likelihood method based on conserved replicase
611 protein (RNA viruses: RdRp, *Polyomaviridae*: LTAg, *Anelloviridae*: ORF1 protein, *Parvoviridae*:
612 NS1, and other DNA viruses: DNA pol). Trees were midpoint rooted, and the branch length
613 indicates number of nucleotide substitutions per site. Dots indicate viruses detected in our
614 samples, and colours represent host genus.

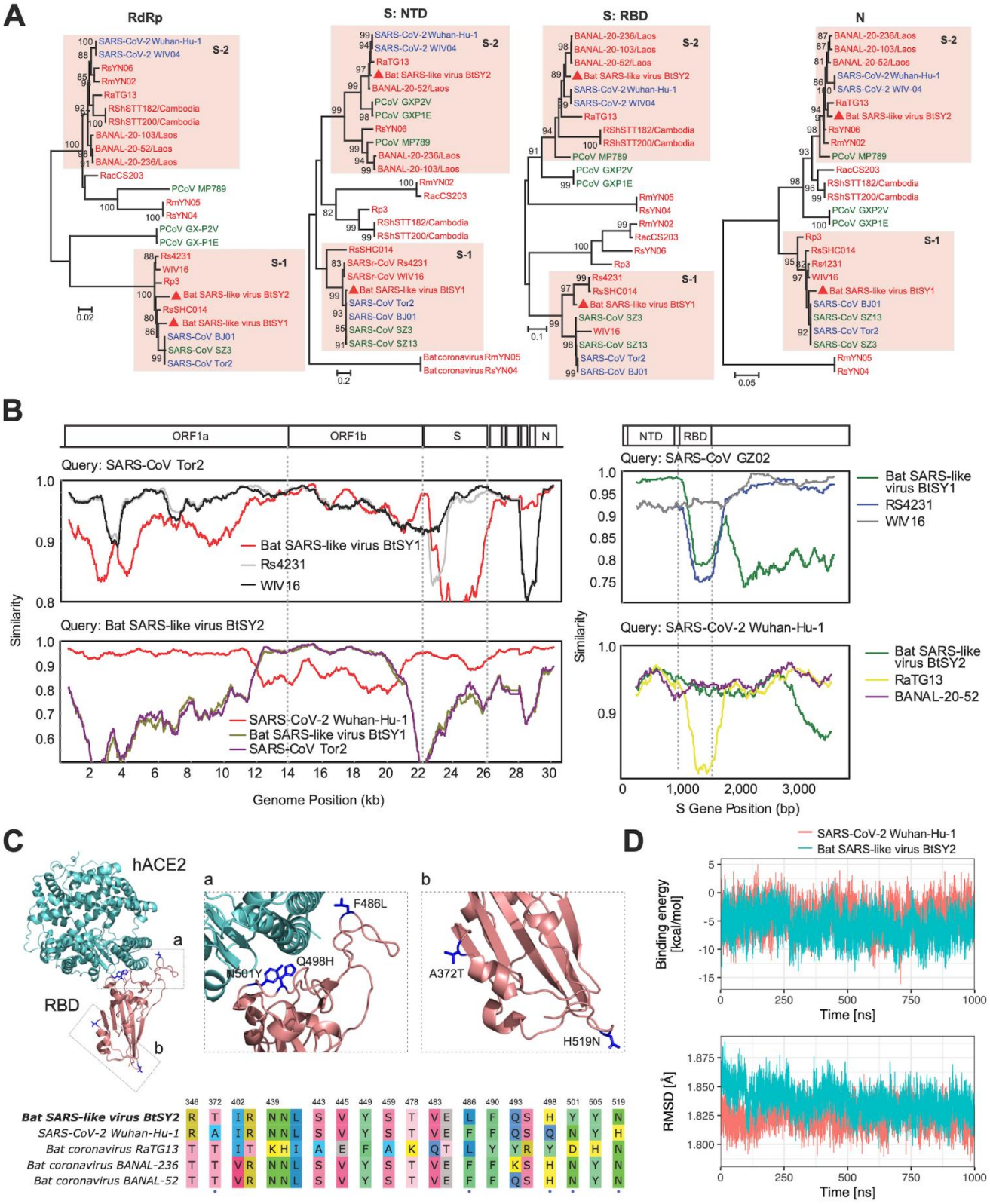


Fig. 6 | Phylogenetic and structural analysis of a potentially zoonotic SARS-related coronavirus detected in our samples. (**A**) Phylogenetic trees of four key functional genes of

SARS-related coronaviruses. Colours of virus strain names indicate the host taxa where the viruses were detected. Red: bats, blue: human, green: others. (**B**) Recombination analysis of SARS-related coronaviruses at the whole genome and spike protein scales. (**C**) Top, homology-modelling structure of the receptor-binding domain (RBD) of Bat SARS-like virus BtSY2 in complex with human angiotensin-converting enzyme 2 (hACE2). Blue-coloured residues on RBD indicate amino-acid substitutions compared with SARS-CoV-2 Wuhan-Hu-1. Bottom, alignment of RBD sequences (residues T333 to G526 of spike protein) of Bat SARS-like virus BtSY2, SARS-CoV-2 Wuhan-Hu-1 and two closely related bat coronavirus. Only polymorphic sites are shown. The five amino acid differences in the RBD of Bat SARS-like virus BtSY2 compared with SARS-CoV-2 Wuhan-Hu-1 are marked with blue dots. (**D**) Molecular dynamics simulation results of stability (top) and binding energy (bottom) of Bat SARS-like virus BtSY2 RBD-hACE2 complex. BtSY1 and BtSY2 are abbreviations for Bat SARS-like virus BtSY1 and BtSY2, respectively.

633 **Supplementary Materials**

634 Supplementary figures S1 ~ S4

635 Supplementary tables S1 ~ S5

636