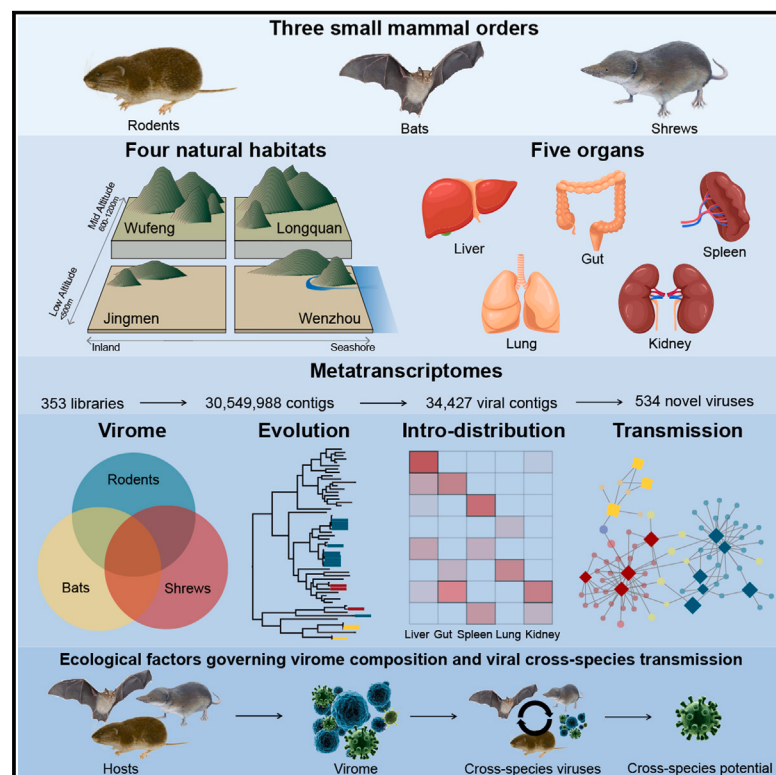


Host traits shape virome composition and virus transmission in wild small mammals

Graphical abstract



Authors

Yan-Mei Chen, Shu-Jian Hu, Xian-Dan Lin, ..., Zhi-Gang Song, Edward C. Holmes, Yong-Zhen Zhang

Correspondence

zhangyongzhen@fudan.edu.cn

In brief

A large study of the mammalian virome identifies differences across rodents, bats, and shrews, revealing insights into host range, transmission, and potential pathogens with the ability to jump across species.

Highlights

- A large number of novel viruses were identified in wild small mammals from China
- Some were of evolutionary significance or had the ability to jump species boundaries
- Shrews carried the most viruses in total and in a single animal species
- Rodents harbored most viruses with the potential be transmitted to new host species

Article

Host traits shape virome composition and virus transmission in wild small mammals

Yan-Mei Chen,^{1,7} Shu-Jian Hu,^{1,7} Xian-Dan Lin,^{2,7} Jun-Hua Tian,^{3,7} Jia-Xin Lv,¹ Miao-Ruo Wang,⁴ Xiu-Qi Luo,¹ Yuan-Yuan Pei,¹ Rui-Xue Hu,¹ Zhi-Gang Song,¹ Edward C. Holmes,^{5,6} and Yong-Zhen Zhang^{1,8,*}

¹State Key Laboratory of Genetic Engineering, Greater Bay Area Institute of Precision Medicine (Guangzhou), School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai 200438, China

²Wenzhou Center for Disease Control and Prevention, Wenzhou, Zhejiang 325002, China

³Wuhan Center for Disease Control and Prevention, Wuhan, Hubei 430022, China

⁴Longquan Center for Disease Control and Prevention, Longquan, Zhejiang 323799, China

⁵Sydney Institute for Infectious Diseases, School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia

⁶Laboratory of Data Discovery for Health Limited, Hong Kong SAR, China

⁷These authors contributed equally

⁸Lead contact

*Correspondence: zhangyongzhen@fudan.edu.cn

<https://doi.org/10.1016/j.cell.2023.08.029>

SUMMARY

Bats, rodents, and shrews are the most important animal sources of human infectious diseases. However, the evolution and transmission of viruses among them remain largely unexplored. Through the meta-transcriptomic sequencing of internal organ and fecal samples from 2,443 wild bats, rodents, and shrews sampled from four Chinese habitats, we identified 669 viruses, including 534 novel viruses, thereby greatly expanding the mammalian virome. Our analysis revealed high levels of phylogenetic diversity, identified cross-species virus transmission events, elucidated virus origins, and identified cases of invertebrate viruses in mammalian hosts. Host order and sample size were the most important factors impacting virome composition and patterns of virus spillover. Shrews harbored a high richness of viruses, including many invertebrate-associated viruses with multi-organ distributions, whereas rodents carried viruses with a greater capacity for host jumping. These data highlight the remarkable diversity of mammalian viruses in local habitats and their ability to emerge in new hosts.

INTRODUCTION

Most human pathogens have animal origins and emerge through cross-species transmission events.^{1,2} The early identification of zoonotic viruses and their hosts is therefore of great importance in efforts to mitigate and prevent future infectious disease outbreaks. Research over the past decade has uncovered a myriad of viruses in animals, particularly following the application of total RNA (meta-transcriptomic) sequencing,^{3–8} in turn revealing the evolutionary origins of a variety of zoonotic pathogens. Importantly, viruses that were later found to be associated with human diseases were first identified in evolutionary and ecological studies of animal viruses (e.g., Jingmen virus and Wenzhou arenavirus).^{3,9–13} Attempts have also been made to identify which mammalian species are most likely to carry zoonotic viruses or which groups of viruses are most able to jump species boundaries and emerge in humans.^{14–17} Despite this, the factors that drive the cross-species transmission of viruses among animals, as well as whether they have the potential to infect and cause outbreaks in humans, remain largely unknown. The emergence of a previously unknown coronavirus, later termed SARS-CoV-

2¹⁸ and the cause of the COVID-19 pandemic, highlights the diversity of zoonotic viruses in mammalian wildlife species as well as the urgent need to understand their potential role in transmitting these viruses to humans.

Bats, rodents, and shrews are the most speciose mammals on Earth.¹⁹ They are distributed worldwide, often live at high densities, and sometimes in close proximity to humans. Historically, rodents have been the most important animal source of human infectious diseases (e.g., plague and viral hemorrhagic fever),^{14,20} whereas bats are currently considered to be natural reservoir hosts of a number of high-impact emerging viruses (e.g., coronaviruses and paramyxoviruses).²¹ A multitude of novel and genetically diverse viruses have been identified and characterized in bats and rodents over the past decade.^{8,22–25} Comparative studies suggest that bats might harbor more viruses and play a more important role in zoonotic transmission than rodents, likely reflecting a number of biological and ecological features,^{14,26,27} including specific immunological traits.²⁸ Although field-based investigations are increasingly revealing the impact of ecological factors on virus diversity,^{23,29,30} there is still no consensus on the role played by mammalian species

that differ in such traits as population size and density that have a fundamental impact on infectious disease dynamics. Indeed, little is known about the interactions between viruses and their mammalian hosts and among viruses, as well as the evolution and spread of viruses within and among allopatric and sympatric mammalian populations. These knowledge gaps hinder our ability to understand the drivers of disease emergence.

To better understand the diversity, ecology, evolution, and transmission of viruses in wild small mammals in local environments, we captured bats, rodents, and shrews from four distinct ecological habitats. Total viromes were obtained from these wild mammals, from which we revealed key aspects of virus evolution and ecology, as well as the interactions among viruses and between viruses and their mammalian hosts.

RESULTS

The richness and ecology of wild small mammals

To better understand the diversity, evolution, and spread of viruses within and among bats, rodents, and shrews, four locations were selected for the field survey of these mammals in Hubei and Zhejiang provinces, China (Figures 1A and S1). These four sampling locations represent natural habitats in subtropical regions, comprising two agricultural areas with woodland and mountain components (Wenzhou and Jingmen) and two mountain forested areas (Longquan and Wufeng). These locations were also located at or close to coastal (Wenzhou and Longquan) or inland (Jingmen and Wufeng) areas. A total of 4,336 individual animals belonging to 44 species from three mammalian orders, Chiroptera ($n = 23$), Eulipotyphla ($n = 6$), and Rodentia ($n = 15$), were captured (Table S1). The small mammals sampled ranged from 13 to 18 species, exhibiting high species diversity and different species compositions at each location (Figures 1B–1D).

The abundance of each species varied considerably (Figure S1). In the case of bats and rodents, there were two to five species with sample sizes of >30 individuals in each of four habitats. Compared with bats and rodents, the Eulipotyphla (i.e., shrews) had relatively lower species richness at all four locations, with none captured in Longquan and only a single species in Jingmen and Wenzhou. However, they exhibited greater species richness and abundance in Wufeng, whereas the Asian house shrew (*Suncus murinus*) had a higher abundance in Wenzhou (Figure S1). The number and composition of all three orders of mammals varied substantially among four locations (Figure 1D), reflecting local biogeographic factors. In particular, the species composition of bats and rodents was significantly affected by habitat (adonis test, $p < 0.001$, explained 14.6% of mammal community variation), with bats and rodents from Wufeng and Longquan (mountain forests) distinct from those sampled in Wenzhou and Jingmen (agricultural areas with woodlands and mountains). Finally, 15 mammalian species were present at two or more locations, whereas the remaining species were only found at one location (Figure 1F).

To test whether the mammals sampled reflect their true richness at each of the study locations, we performed a rarefaction analysis. A diminishing trend was observed in all four locations (Figure 1E). Although the rarefaction curve in Longquan did not

reach a plateau, with one new species added for nearly every 40 samples, this analysis suggests that most (70%–90%) of the small mammal species in all four locations could be sampled with roughly 400 samples. Hence, the mammals captured in this study generally reflect their true abundance in all four locations.

The virome of wild small mammals

The viromes in these wild small mammals were determined using the meta-transcriptomic protocol established in our laboratory.^{3–6} RNA was extracted from lung, liver, kidney, spleen tissue samples, and intestinal feces, and RNA sequencing libraries were prepared according to tissue type, host species, and sampling location. In total, 2,443 individual mammals were analyzed, yielding over 14.5 billion sequence reads across 353 libraries that could be used for the assembly and identification of viruses (Table S2). Overall, 30,549,988 contigs were assembled, including 34,427 viral contigs. Viral contigs were subsequently classified into >96 viral families and grouped according to the known host associations of the closest viruses assigned by Blastx and phylogenetic analyses (Figure S2). Although bacterial, fungal, and plant-associated viruses (13,913 contigs) were also detected, they were excluded from subsequent analyses.

Based on analysis of the RNA-dependent RNA polymerase (RdRp) genes for RNA viruses and replicase protein genes for DNA viruses, 669 vertebrate- and invertebrate-associated viruses from >42 families were identified and characterized, comprising >40 families of RNA viruses and two families of DNA viruses (Figure 2A; Table S3). Of these, 534 viruses from >40 families were defined as novel species according to the current species demarcation criteria by the International Committee on Taxonomy of Viruses³¹ (ICTV; Table S4).

Diverse and abundant viruses were present and highly prevalent in the mammals sampled (Figure 2A). With the exception of the woolly horseshoe bat (*Rhinolophus luctus*), from which only one individual was captured, 1–150 viruses were identified in all remaining mammalian species, with the greatest number in Smith's shrews (*Chodsigoa smithii*). However, virome composition varied largely at the level of virus family among the three mammalian orders (adonis test, $p < 0.001$; Figures 2B and 2C). Viruses from the families *Paramyxoviridae*, *Picornaviridae*, *Astroviridae*, *Flaviviridae*, *Rhabdoviridae*, *Sedoreoviridae*, and unclassified viruses from the order *Picornavirales* were most commonly detected (>20 mammalian species from the three orders), accounting for 48.0% of the total viruses identified herein (Figures 2A and 2C). Other viruses were more dispersed. For example, members of the *Arteriviridae* and *Hantaviridae* were more frequently detected in rodents and shrews: arteriviruses were detected in 4 of 15 rodent species (4/15), 0/23 bats, and 2/6 shrews, and hantaviruses were detected in 7/15 rodents, 3/23 bats, and 3/6 shrews (Chi-squared test, both $p < 0.05$). By contrast, viruses from the families *Coronaviridae* and *Hepadnaviridae* seemed to be more prevalent in bats: coronaviruses, 5/15 rodent species, 8/23 bats, 1/6 shrews (Chi-squared test, $p = 0.69$); hepadnaviruses, 0/15 rodents, 6/23 bats, 1/6 shrews (Chi-squared test, $p = 0.10$). Notably, shrews harbored myriad other viruses, including henipaviruses and those associated with invertebrates (e.g., *Iflaviridae*, *Dicistroviridae*, *Permutotetraviridae*, and *Spinareoviridae*), and had a similar diversity of

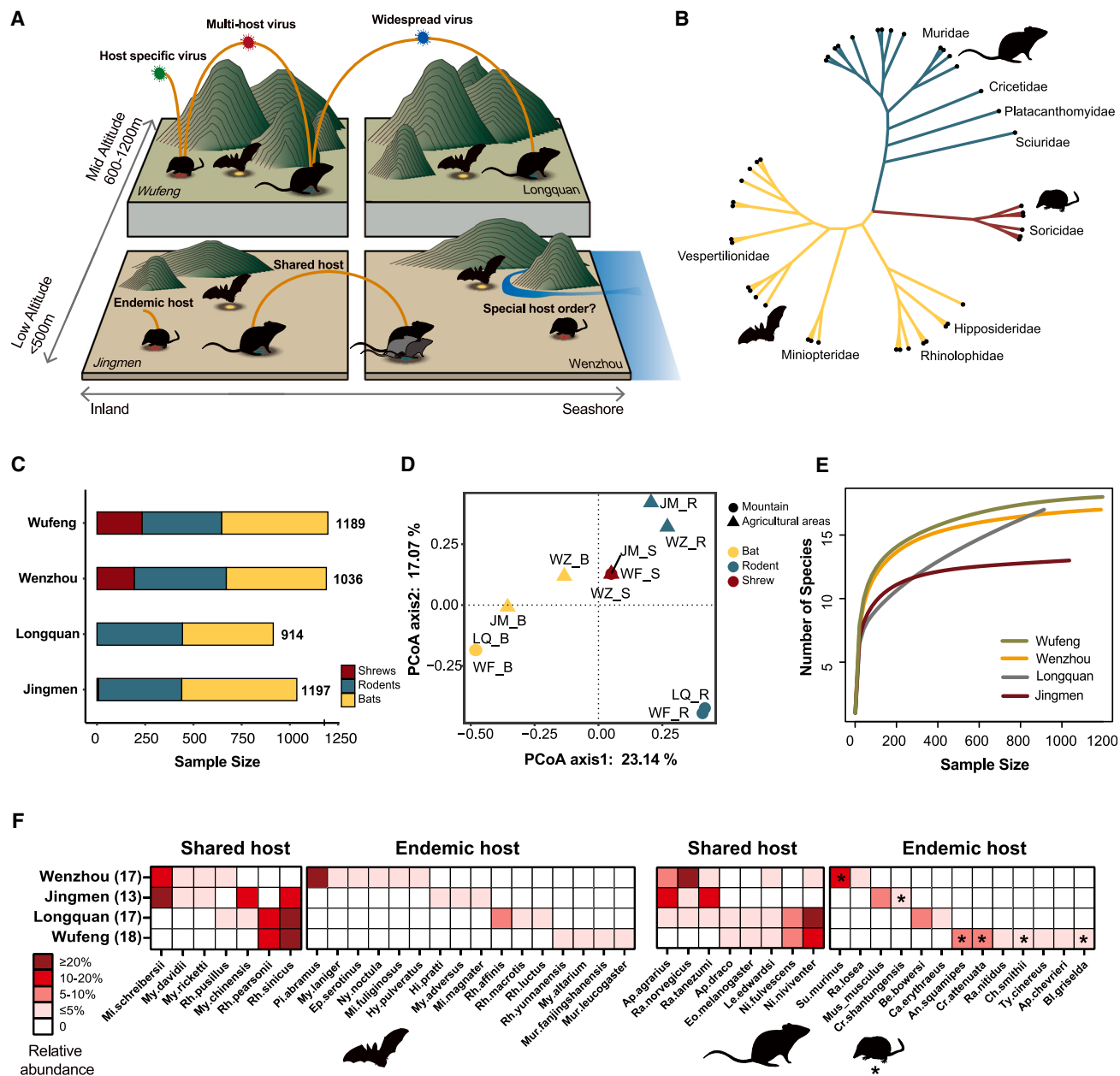


Figure 1. Sample locations, landscapes, and wild small mammal compositions

(A) Geographical locations and landscapes of this study. Shared hosts, mammalian species sampled from more than one location; endemic host, mammalian species sampled in only one location.

(B) Phylogeny of the three mammalian orders, with the nine mammalian families included in this study being labeled.

(C) Mammalian order compositions at each location.

(D) Principal coordinate analysis showing the variation in mammalian compositions between mountainous and agricultural areas.

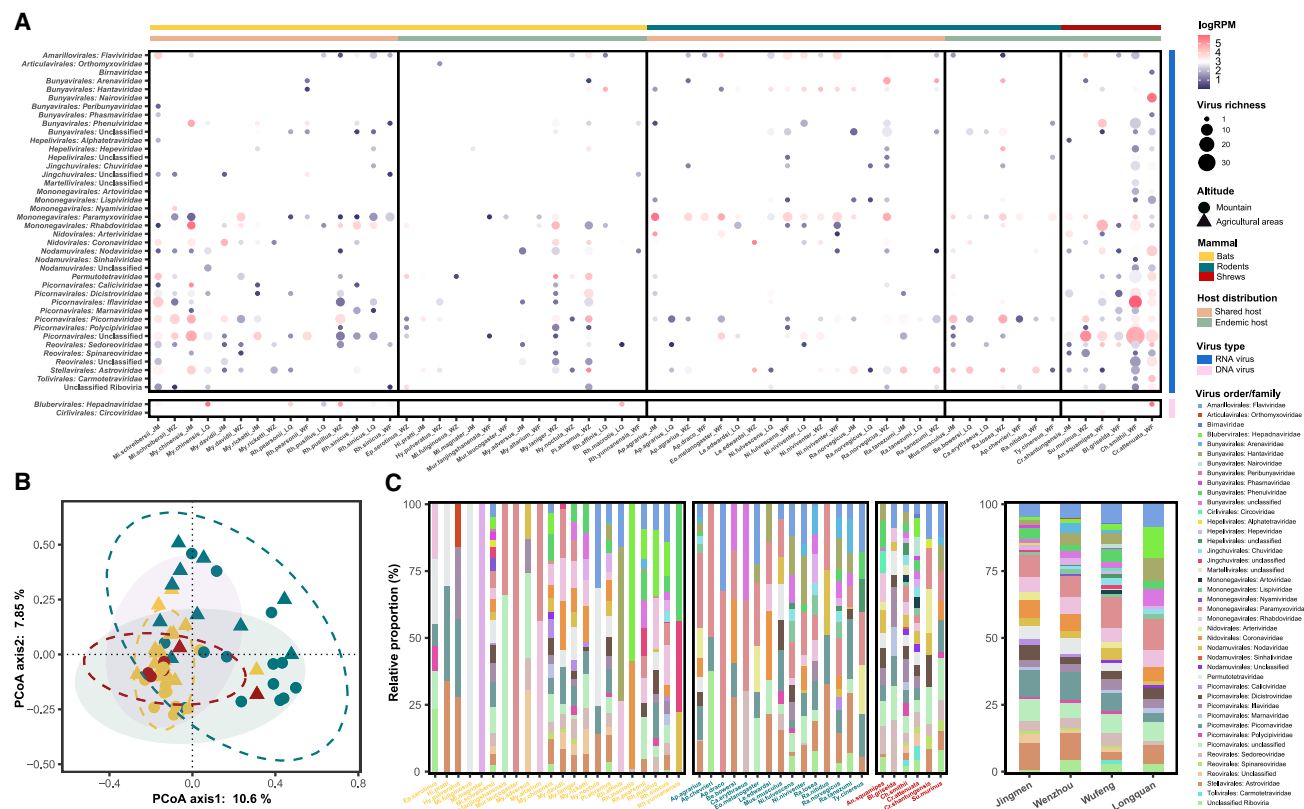
(E) Rarefaction curves of mammal sample size in each location.

(F) Mammalian species compositions at each sampling location. Shrews are denoted with an asterisk.

See also [Figure S1](#) and [Table S1](#).

hantaviruses as rodents (Figures 2A and 2C). Finally, virome composition also varied among sampling locations, with mammals from Wufeng and Longquan (mountain forests) different to those from Wenzhou and Jingmen (agricultural areas with

woodlands and mountains) in the principal coordinate analysis (PCoA) plot (adonis test, $p < 0.01$; Figure 2B). In particular, mammals living in Wufeng harbored more viruses (Kruskal test, $p < 0.01$; Figure 2C).



Organ distribution and intra-organ abundance of viruses in wild small mammals

We next sought to determine the distribution and abundance of the viruses within the internal organs. Viruses from different families exhibited considerable variation in organ distribution and intra-organ abundance in these wild mammals (Figures 3A and S3; Table S5). Almost all viruses from the *Astroviridae* (44/49), *Caliciviridae* (5/6), *Coronaviridae* (22/26), and *Picornaviridae* (59/78) were found in high abundance in feces. The *Paramyxoviridae* were most commonly detected and at high abundance in kidney, with 40.7% found in the kidney only. Notably, a high abundance of viruses from the *Arenaviridae* (8/10), *Arteriviridae* (9/12), *Flaviviridae* (38/51), and *Hantaviridae* (19/21) were detected in lung, liver, kidney, and spleen but not or rarely in feces, whereas viruses from the *Hepeviridae* and *Nairoviridae* were distributed in all five organs. Interestingly, a number of viruses from invertebrate-associated families, such as the *Rhabdoviridae* (26/57), *Iflaviridae* (17/42), *Nodaviridae* (11/35), *Phenuiviridae* (7/37), were also present in two or more organs at considerable abundance, although their abundance was generally much higher in feces (Figure 3A). Finally, viruses that were commonly detected (e.g., astroviruses, coronaviruses, paramyxoviruses, and picornaviruses) tended to have a multi-organ distribution

compared with those rarely found (e.g., caliciviruses, spinareoviruses, and sinhaliviruses) (Figure 3B).

Notably, even viruses from the same family had a distinct intra-host distribution in different mammals (Figures 3B and S3; Table S6). For example, although the abundance of paramyxoviruses in kidney was the highest in all three mammalian groups, a higher proportion of viruses exhibited a multi-organ distribution in shrews (66.7%) and rodents (65.2%) than in bats (21.9%) (Fisher's exact test, $p = 0.09$ [shrews vs. bats], $p < 0.001$ [rodents vs. bats]). For astroviruses, a multi-organ distribution was more common in rodents (50.0%) than in bats (26.1%) and shrews (16.7%), but the difference was not significant (Fisher's exact test, $p = 0.127$ [rodents vs. bats], $p = 0.197$ [rodents vs. shrews]). Remarkably, shrews and bats not only harbored more invertebrate-associated viruses (84.3% and 52.3%) than rodents (28.1%) (Fisher's exact test, both $p < 0.001$) but also a high proportion of which exhibited a multi-organ distribution (shrews, 35.3%; bats, 29.6%). In particular, more than half of rhabdoviruses and permutotetraviruses in bats had multi-organ distributions, whereas nearly half of dicistroviruses, iflaviruses, and unclassified viruses from the *Picornavirales* exhibited multi-organ distribution in shrews. Although a bat chuvirus and five shrew chuviruses were mainly detected in feces, a rodent chuvirus (Wufeng rodent

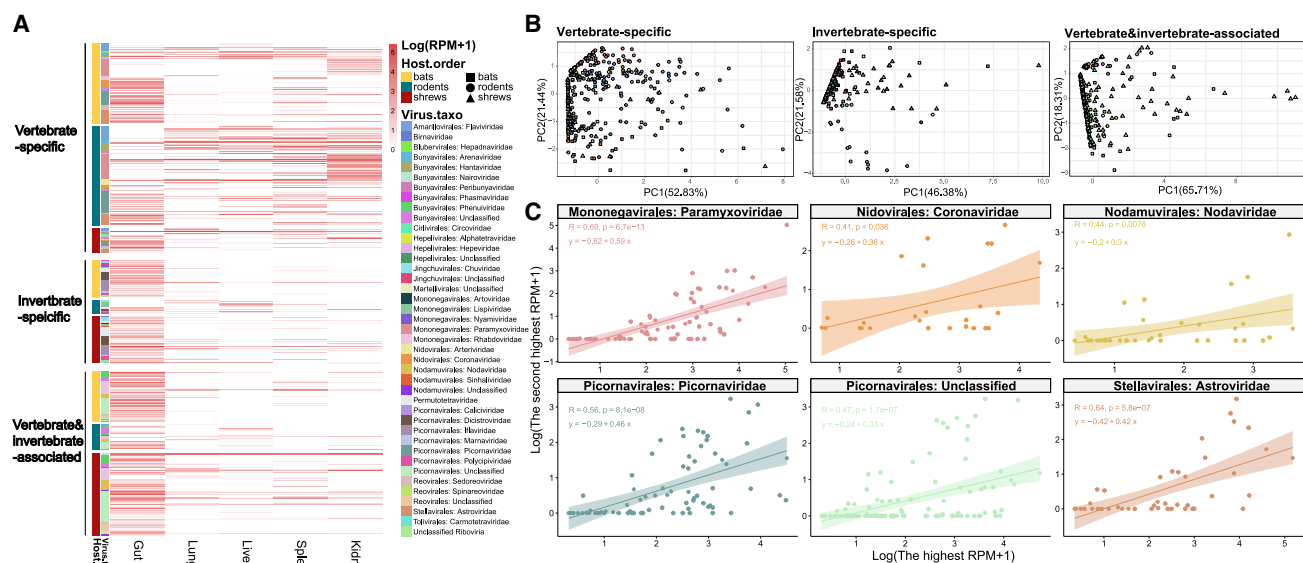


Figure 3. Organ distribution of vertebrate- and invertebrate-associated viruses in wild small mammals

(A) Heatmap showing virus abundance in five types of internal organs.

(B) Principal component analysis (PCA) plot showing the multi-organ distribution of viruses in bats, rodents, and shrews. Each point represents one virus detected in one organ.

(C) Relationships of virus abundance between the primary target and the secondary organs. Each point represents one virus in the family in question. $p < 0.05$ indicates significant relation.

See also [Figure S3](#) and [Tables S5](#) and [S6](#).

chuvirus 1) had a higher abundance in both liver and spleen than in feces, and the other two shrew chuviruses (Wufeng shrew chuvirus 1 and Wufeng shrew chuvirus 6) were present in both lung and spleen in addition to feces ([Table S7](#)), suggesting that these three chuviruses produced systemic infections.

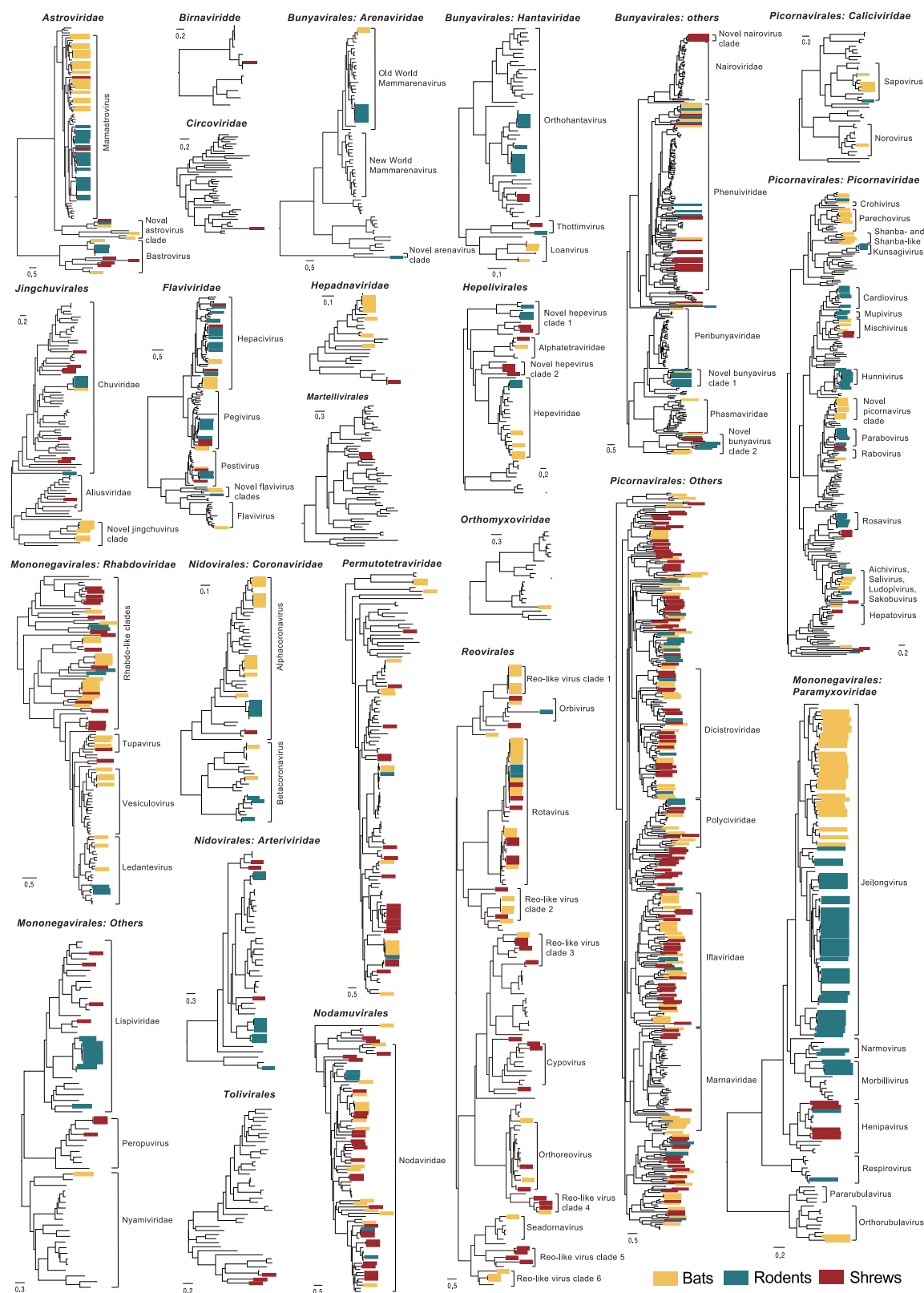
Of most note, viruses were detected in other internal organs when their abundance in the primary target organs reached a certain level ([Figures 3C](#) and [S3](#)). For example, viruses from the *Astroviridae*, *Coronaviridae*, and *Picornaviridae* were detected in the lung, liver, kidney, and spleen when their abundance in feces reached to $10^{1.0}$, $10^{0.72}$, and $10^{0.63}$ reads per million (RPM), respectively ([Figure 3C](#); [Table S5](#)). A similar observation was made for viruses from the *Paramyxoviridae*, with a threshold of virus “spillover” from kidney into other internal organs of approximately $10^{1.05}$ RPM. For viruses from the invertebrate-associated *Nodaviridae* and unclassified viruses from the order *Picornavirales*, the threshold values were $10^{0.67}$ RPM and $10^{0.73}$ RPM, respectively. Finally, in addition to the primary target organs, spleens had the second highest detection rate (41.5%) of the spillover viruses.

Diversification and evolution of viruses in wild small mammals

Phylogenetic analyses of the viruses identified here revealed that bats, rodents, and shrews harbored a high diversity of mammalian viruses, including those of evolutionary significance ([Figures 4](#), [S4](#), [S5](#), and [S6](#)). For example, a novel hantavirus (Wufeng rodent hantavirus), identified in *Eothenomys melanogaster* voles from Wufeng, differed from known hantaviruses (>30% amino acid [aa] difference) but formed a sister taxon to shrew thottimviruses

([Figure S5](#)). A bat arenavirus (Wufeng bat mammarenavirus 1), identified in both *R. pearsonii* bats from Wufeng and *Pipistrellus abramus* bats from Wenzhou, is the first Old World arenavirus discovered in bats ([Figure S4](#)) and showed a close evolutionary relationship to *Ryukyu mammarenavirus* (>93.3% aa identity) identified in a *Mus caroli* mouse from Yunnan province >2,000 kilometers distant. Novel astroviruses identified here clustered with those previously identified in reptiles, amphibians, and soil in China, and formed a sister lineage to the established mamastroviruses and avastroviruses ([Figure S4](#)). Within the *Picornaviridae*, two novel shrew-associated picornaviruses, Wufeng shrew picornavirus 5 and Wenzhou shrew picornavirus 1, formed distinct and sister lineages to the established paraboviruses and dicapiviruses, respectively ([Figure S4](#)). In the *Paramyxoviridae*, two novel rodent-associated morbilliviruses formed a separate cluster with >40% aa difference from known morbilliviruses, including human measles virus, whereas three novel jeilongviruses, present in multiple rodent species and locations, clustered with *Jeilongvirus rungwense* (MW579602) and formed a distinct lineage within the Jeilongvirus phylogeny ([Figure S4](#)).

Also of note was the discovery of viruses related to those already known to infect human or domestic animals ([Figures S4](#) and [S5](#)). In particular, close relatives of viruses associated with human disease were identified in multiple mammal species and sampling locations with considerable abundance, including *Wenzhou mammarenavirus* (in six mammal species; 89.4%–96.9% aa identity to known viruses), *Hantaan orthohantavirus* (one mammal species; 97.4% aa identity), *Seoul orthohantavirus* (three mammal species; 98.0%–99.4% aa identity), *Cardiovirus*



(legend on next page)

A (three mammal species; 83.4%–90.0% aa identity), *Aichivirus* A (one mammal species; 94.3% aa identity), *Rotavirus* A (eight mammal species; 85.6%–95.5% aa identity), and *Mammalian orthoreovirus* (one mammal species; 86.7% aa identity). In addition, relatives of human orthorubulavirus 4 (70% aa identity) were identified in bats from Wenzhou and Wufeng. A novel henipavirus (Wenzhou shrew henipavirus 1), detected in both *S. murinus* shrews and *Apodemus agrarius* mice from Wenzhou, was relatively closely related (82.0% aa identity in RdRp) to Langya virus recently identified in febrile patients and *Crocidura* shrews,³² suggesting an ability to jump species boundaries. Finally, two novel rotaviruses were identified in *C. shantungensis* shrew and multiple bat species, respectively, which showed an evolutionary relationship (67.7% and 69.9% of aa identity in RdRp) to the zoonotic *Rotavirus* A.

Although SARS-CoV-2-related viruses were not identified, we discovered 13 coronaviruses in eight bat species (three families, four genera), five rodent species (one family, five genera), and one shrew species (Figure S5). Of note were two novel bat alpha-coronaviruses detected in *Myotis* and *Miniopterus* species from Jingmen and Wenzhou and were related to porcine epidemic diarrhea virus (PEDV; NC_003436) with 73.2%–74.6% aa identities in the open reading frame 1ab (ORF1ab) polyprotein. In addition, a close relative of swine acute diarrhea syndrome coronavirus (SADS-CoV) was identified in a *R. macrotis* bat for the first time, exhibiting >97% aa identity in ORF1ab and >95% aa identity in spike protein with SADS-CoV.³³ Finally, a single SARS-related coronavirus was identified in a *R. sinicus* bat from Jingmen (Hubei province), exhibiting 96.5% aa identity to the SARS-CoV ORF1ab polyprotein, although it was most closely related to a coronavirus previously reported in the same bat species from Hubei (99.0% aa identity in ORF1ab).²⁴ Interestingly, the virus detected here exhibited only 78.6% and 74.5% aa identity to the SARS-CoV spike protein and receptor binding domain, respectively.

We also identified a large number of viruses previously thought to be invertebrate-specific or associated (Figures 4 and S6). These viruses can be broadly divided into four categories based on their positions in phylogenetic trees and their distribution within and among mammals. Of particular note is the first category comprising viruses identified in mammals for the first time and present in multiple species and organs at relatively high abundance (Figure S6; Table S5)—this suggests that these mammals are bona fide hosts for these viruses rather than being a component of host diet. For example, Wenzhou rodent chuvirus 1 was identified in lung tissue of both *Rattus* rats and *Apodemus* mice with high prevalence rates (5.56%–63.2%) and abundance (17.98–74.81 RPM) (Figure S6; Table S7). The rodent chuvirus clustered with Lishi spider virus 1 identified in a spider from Wuhan,³ but with only 32.9% aa identity. In addition, they have different genome structures: linear segmented for Wenzhou rodent chuvirus 1 but circular segmented for Lishi spider virus 1.

Another rodent chuvirus, Wufeng rodent chuvirus 1, was detected in both liver and spleen of a *Niviventer* rat and formed a sister taxon to the known viruses of the family *Chuviridae*.

The second category comprised viruses that clustered together and formed distinct clades within established invertebrate-associated virus groups, such as the novel jingchuvirus clade identified in multiple bat species from different locations that formed a divergent lineage with <40% aa identity to the remaining *Jingchuvirales*. A similar pattern was seen for novel virus clades in the *Bunyavirales* and *Picornavirales*, as well as the *Nodaviridae* and *Rhabdoviridae*. Viruses in the third category had very close relationships to those previously identified in invertebrates. For example, the tick-associated *Ledantevirus yongjia* virus was identified in rodents from Wufeng with 98.7% aa identity to the reference sequence. Jingmen rodent permutotetravirus 1 found in both rodents and shrews showed >96% aa similarity to Guiyang permutotetra-like virus 1 previously identified in spiders. The remaining viruses fell into the final category, which were present in sporadic, but sometimes notable, topological positions. The most striking example was Longquan bat nodamuvirus 1, which was distinct from all known nodamuviruses (Figure S6; Table S3). Overall, 54.5% of these invertebrate-associated viruses were sampled from shrews, and >35% had multi-organ distribution (Table S6).

Transmission of viruses in wild small mammals

The majority of the newly identified viruses clustered together or with those identified previously according to their animal hosts in their respective phylogenetic trees, regardless of their geographic origins (Figures 4, S4, S5, and S6). Although 531 newly identified viruses were found in one host species, likely cross-species transmission events occurred between mammals at the species, genus, family, and even order levels. In total, 138 viruses were identified in $\geq$ two species of mammals, 106 viruses in $\geq$ two genera of mammals, 49 viruses in $\geq$ two families of mammals, and 18 viruses in two or all three orders of mammals (Figures 5A and 5B).

The cross-species transmitted viruses identified here comprised 29 viral clades (families or orders), especially those from the *Paramyxoviridae*, *Picornaviridae*, *Flaviviridae*, *Astroviridae*, and *Iflaviridae* (Figure 5A). Notably, the 18 viruses capable of transmission between mammalian orders not only included vertebrate-associated viruses (e.g., hepacivirus, arenavirus, henipavirus, and astrovirus) but also some invertebrate-associated viruses (e.g., dicistrovirus and permutotetravirus) (Figure 5A). Strikingly, *Rotavirus* A had the widest range of mammal hosts (8 species), co-circulating in bats, rodents, and shrews. Additionally, three newly identified hepaciviruses were present in both rodents and shrews. However, most cross-species transmission events (e.g., coronaviruses, hantaviruses) occurred at the species or genus level.

To better characterize the transmission of viruses in these wild mammals, we constructed host-virus correlation networks

Figure 4. Phylogenetic diversity of vertebrate- and invertebrate-associated viruses

Phylogenetic trees were estimated based on amino acid sequences of the RdRp protein for RNA viruses and the replicase protein for DNA viruses. Viruses newly identified in this study are color marked by mammal hosts. All trees are midpoint rooted for clarity, with branch lengths scaled to the number of amino acid substitutions per site. The scale bars represent the number of substitutions per site.

See also Figures S4, S5, and S6.

Table 1. Summary of best-fit models

Term	p value	Deviance explained
Model 1: total virus richness model		39.65%
Host order	–	32.12%
Rodents	0.9095	–
Shrews	<0.001 ^a	–
Host sample size	<0.001 ^a	6.75%
Model 2: total virus abundance model		32.35%
Host order	–	11.26%
Rodents	0.1337	–
Shrews	<0.01 ^a	–
Host sample size	<0.01 ^a	14.40%
Altitude	–	6.27%
Agricultural areas	<0.05 ^a	–
Model 3: number of cross-species transmitted virus model		72.09%
Total virus richness	<0.001 ^a	24.51%
Host order	–	6.07%
Rodents	<0.01 ^a	–
Shrews	0.3224	–
Host sample size	<0.05 ^a	3.49%
Altitude	–	1.50%
Agricultural areas	0.0943	–
Total virus abundance	0.1572	1.06%
Model 4: potential for cross-species transmission model		20.80%
Host order	–	7.53%
Rodents	<0.001 ^a	–
Shrews	<0.001 ^a	–
Virus family-level taxonomy	<0.01 ^a	6.44%
Total virus abundance	<0.05 ^a	0.56%
Is multiorgan-distributed	0.0776	0.24%
Model 5: potential for cross-species transmission model (virus family nested within host family)		34.40%
Host family	<0.001 ^a	1.59%
Virus family, host family	0.1980	11.30%
Virus family-level taxonomy	0.1020	1.25%
Is multiorgan-distributed	<0.05 ^a	0.19%
Is invertebrate-associated	0.1189	0.50%
Model 6: potential for cross-species transmission model (virus family nested within host order)		25.00%
Host order	<0.001 ^a	2.31%
Virus family, host order	0.7323	4.22%
Virus family-level taxonomy	<0.05 ^a	2.57%
Total virus abundance	0.0572	0.26%
Is multiorgan-distributed	0.0621	0.29%

^aSignificant.

the most important factor associated with virus richness and the second most important for virus abundance, explaining 32.1% and 11.3% of total deviance in the two models (Table 1). Notably, shrews had a significantly positive effect on virus richness ($p < 0.001$) and abundance ($p < 0.01$) (Figures 6A and 6C) and harbored higher virus richness (48.2 viruses per host species) and abundance (328,333 RPM per host species) than bats (11.9 viruses; 58,446 RPM) and rodents (9.1 viruses; 50,866 RPM) (Figures 6B and 6D). This highlights the importance of shrews as virus reservoirs. Host sample size was also significantly associated with total virus richness (explaining 6.75% of total deviance) and abundance (14.4%), and mammals with greater sample sizes tended to carry more viruses at higher abundance (Figures 6A and 6C; Table 1). However, significant correlations were observed in bats (species richness, $r = 0.69$, $p < 0.001$; abundance, $r = 0.48$, $p < 0.01$) and rodents (species richness, $r = 0.85$, $p < 0.001$; abundance, $r = 0.59$, $p < 0.01$), but not in shrews (Figures 6B and 6D). In addition, host habitat impacted virus abundance (explaining 7.1%, $p < 0.05$, Figure 6C), although the effects varied. Rodents inhabiting agricultural areas (Wenzhou and Jingmen) harbored significantly higher virus abundance than those from mountainous regions (Longquan and Wufeng) (Wilcoxon test, $p < 0.01$, Figure 6D). Shrews, however, showed the opposite trend, with those living in mountainous regions tending to harbor more abundant viruses, although the effect was not statistically significant (Wilcoxon test, $p = 0.533$, Figure 6D).

The best model to predict the number of cross-species transmitted viruses explained 72.1% of the variance and included total virus richness, host order, host sample size, host habitat (altitude), and total virus abundance (Figure 6E; Table 1). Not surprisingly, total virus richness explained the largest fraction (24.5% of total deviance), supporting the null hypothesis that higher total virus diversity is associated with more frequent virus spillover.^{14,15} Of note was that rodents had the highest frequency of virus spillover, followed by bats, compatible with a host effect independent of total virus diversity (Figure 6F). In line with this, host order (explaining 6.07% of the deviance) and host sample size (explaining 3.49%) were also significant predictors of the number of cross-species transmitted viruses (Figure 6E). Of note, rodents had a significantly positive association with the number of cross-species transmitted viruses ($p < 0.01$, Figure 6E). In addition, host habitat (agricultural or mountainous areas) and total virus abundance per host species also contributed to the overall predictive power of the best model but were not statistically significant (Table 1).

Finally, we used mixed GAMs to identify factors associated with whether a virus was capable of cross-species transmission (although such events may not have occurred recently). The best-fit model explained 20.8% of total deviance, and included effects of host order, virus taxonomy, total virus abundance, and whether or not a virus was multi-organ distributed (Figure 6G; Table 1). Notably, viruses in rodents had a significantly higher potential to jump across host species than those in bats and shrews ($p < 0.001$; Figure 6G). Although the virus family-level taxonomy was the second most important and significant factor in the best-fit model ($p < 0.01$, Table 1), it did not identify any virus family with a higher potential for cross-species transmission (Figure 6G). To

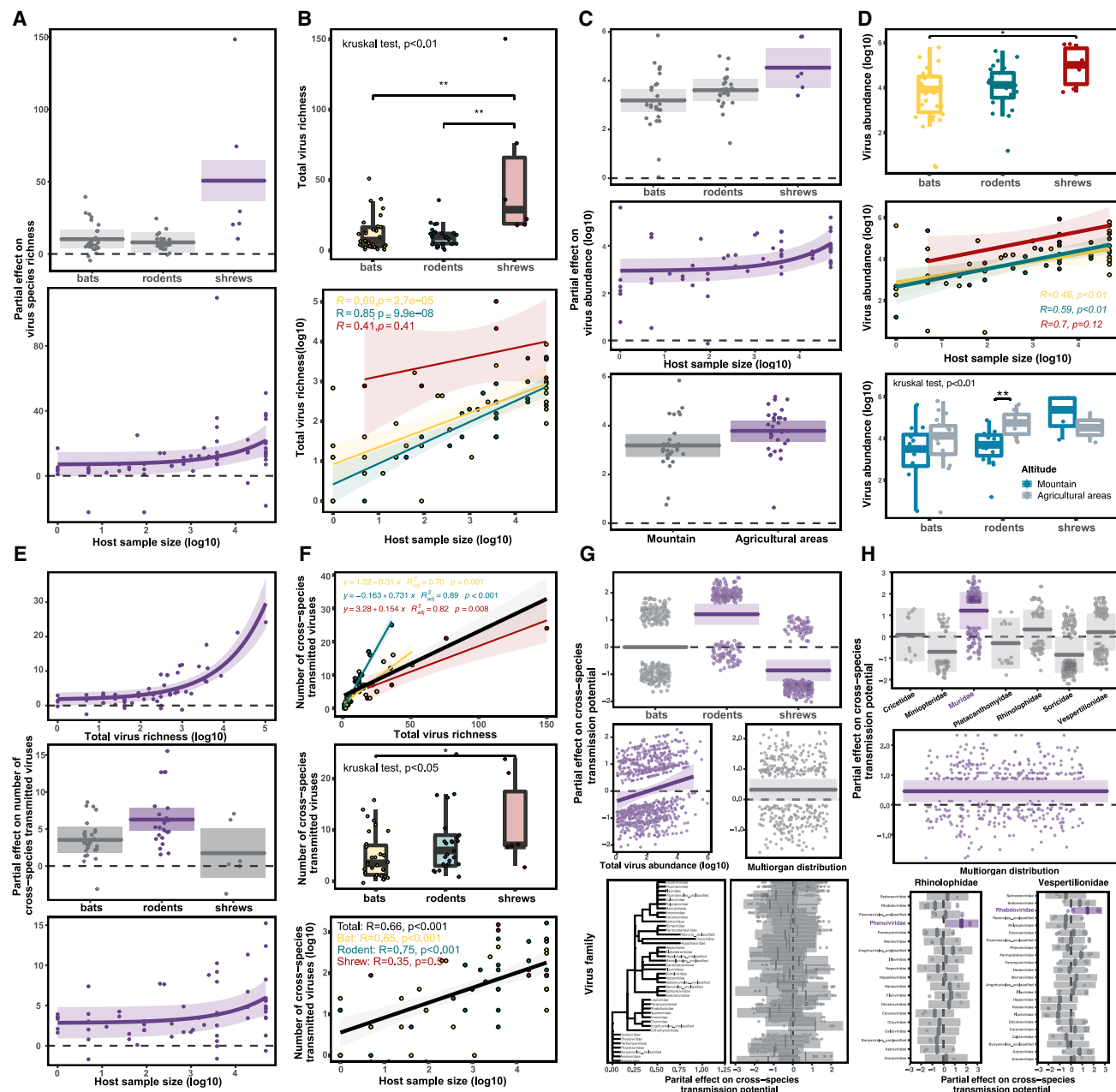


Figure 6. Ecological factors governing virome composition and cross-species virus transmission

(A) Partial effect plots of the best-fit model for total virus richness.
 (B) Effect of host order and host sample size on total virus richness per host species.
 (C) Partial effect plots of the best-fit model for total virus abundance.
 (D) Effect of host order, host sample size, and host habitat on total virus abundance.
 (E) Partial effect plots of the best-fit model for the number of cross-species transmitted viruses.
 (F) Effect of total virus richness, host order, and host sample size on the number of cross-species transmitted viruses.
 (G) Partial effect plots of the best-fit model for cross-species virus transmission potential.
 (H) Partial effect plots of the best-fit model for cross-species virus transmission potential including a random effect of virus taxonomy nested within host family. Partial effect plots show the relative effect for the predictor variables included in the corresponding best-fit model. Points in partial effect plots represent partial residuals, lines indicate the effect of each variable, and shaded regions show the 95% confidence intervals of the partial effect. Factors that had a significant effect are colored purple. Only the subset of virus families that include significant effects are shown in (H). Box plots and scatter plots show the difference of the response variables among groups or the variation trend along with continuous variables. Points in box and scatter plots indicate virus richness/abundance in one mammal species from one location, while lines show a linear regression fit for all mammals (dark) and for each mammal order, with its 95% confidence interval (CI) indicated by the shading. $p < 0.05$ indicates significant relation. Asterisks in box plots indicate significant difference between groups with $*p < 0.05$ or $*p < 0.01$.

further investigate whether there was a host-specific virus family effect (i.e., the effect of a specific association between hosts and viruses), two independent sets of GAMs were used, including a random effect of a virus family nested within host order or host family (Table 1). A host-specific virus family effect contributed to the best models of both sets, although the overall effect was not statistically significant (model 5, $p = 0.198$; model 6, $p = 0.732$). In the best-fit model at the host family level (model 5), phenuiviruses from *Rhinolophidae* bats and rhabdoviruses from *Vespertilionidae* bats had a significantly higher potential to jump among host species (Figure 6G). Additionally, the best models also suggested that viruses at greater abundance (model 4, $p < 0.05$; model 6, $p = 0.058$) and those distributed in multiple organs (model 4, $p = 0.078$; model 5, $p < 0.05$; model 6, $p = 0.062$) were more likely to spill over and infect more hosts (Figures 6G and 6H).

DISCUSSION

We performed a large-scale investigation of the diversification, evolution, and ecology of viruses in bats, rodents, and shrews from four locations representing different habitats in the Chinese subtropics. The richness and abundance of the mammalian species studied were greater than reported previously,^{22,34} with a large number of novel viruses identified. Viruses were identified in all mammals, with the exception of the woolly horseshoe bat.

We identified viruses in multiple species of the wild small mammals sampled, some at high prevalence, indicative of relatively frequent cross-species virus transmission. Notably, some of these newly identified viruses were related to known human or domestic animal pathogens, including newly identified viruses from the orthorubulaviruses as well as PEDV-, SADS-, and SARS-related coronaviruses. In addition, a novel henipavirus sampled from shrews was related to Langya virus—a recently reported etiologic agent of human fever.³² Finally, known human or domestic animal pathogens (e.g., *Rotavirus A*, *Seoul virus*, *Wenzhou mammarynavirus*) were also found to be highly prevalent in these wild small mammals. Together with the identification of SARS-CoV-2-related coronaviruses in multiple bat species,^{35–40} these data highlight the need to strengthen the surveillance of human populations that interact with these animal species.

We also identified viruses that fell into important phylogenetic positions. Within the *Hantaviridae*, Wufeng rodent hantavirus identified in an *Eothenomys* vole formed a sister taxon to the genus *Thottimvirus*, which occupies the important evolutionary position within the subfamily *Mammantavirinae*.⁴¹ However, all currently known thottimviruses are from shrews, adding complexity to our understanding of the origin of mammal hantaviruses. In the case of the *Arenaviridae*, Wufeng bat mammarynavirus 1 is the first bat arenavirus identified in the Old World. Combined with the description of New World bat arenavirus in Trinidad,⁴² these data indicate that bats are hosts of arenaviruses and suggest that there have been host jumps of arenaviruses from rodents into bats. In addition, the viruses identified in bats, rodents, and shrews from the *Astroviridae*, *Paramyxoviridae*, and *Picornaviridae*, respectively, were relatively distinct from all recognized viruses and formed sister taxa to other vi-

ruses from their respective families. The discovery of these viruses therefore provides more information on the evolutionary origins of these viruses.

It is likely that many families of vertebrate RNA viruses have their ultimate origins from invertebrates³ and arthropod-borne viruses can infect and replicate in both arthropod vectors and vertebrate hosts.⁴³ We identified a large number of viruses previously thought to be invertebrate-specific or invertebrate-associated. Many were found in mammals for the first time, had a multi-organ distribution, and were presented in multiple species, indicating that these wild mammals are likely to be natural hosts. Of these viruses, the chuviruses were of most note due to their high prevalence in mammals, wide distribution within the host, and divergent positions in the phylogenetic tree, suggesting that they may have a long evolutionary history in mammals. Remarkably, compared with bats and rodents, a much greater proportion of invertebrate-associated viruses were identified in shrews. In particular, some of these shrew viruses were related to those previously identified in invertebrates or mammals, whereas others were quite distinct from known invertebrate viruses (e.g., viruses from *Bunyavirales*). The fact that shrews feed on arthropods may in part explain why shrews harbor much more viruses than bats and rodents.

It is widely believed that viral diversity, in part, reflects the richness and abundance of their hosts.^{15,21} Although the species richness of rodents is approximately twice of that of bats,¹⁹ bats are often considered to harbor more viruses rather than rodents due to their respective biological and ecological features.^{14,26,27} Of the animals sampled here, bats had the highest richness, followed by rodents and shrews. However, although the total number of viruses identified in bats was more than that in rodents, the average number of viruses identified per species of bats and rodents was similar. In addition, shrews had the most viruses in total and in a single species, with up to 150 viruses from 29 viral clades (orders or families) identified in Smith's shrews. Our best model to predict total virus richness also identified host order as the most important determinant, explaining more than half of the total deviance. Hence, compared with other biological factors, host organisms may have the biggest impact on the richness of viruses they carry.

Since most human infectious diseases are caused by viruses originating in nonhuman animals, studies of viral cross-species transmission have largely focused on zoonotic spillover.^{14,44} However, zoonotic spillover represents only a small fraction of all the possible cross-species virus transmission events. As they possess larger population sizes with greater opportunities for host contact, most cross-species transmission events likely occur among wildlife, occasionally posing a threat to particular animal species.^{4,45,46} Here, cross-species virus transmission was apparent at host species, genus, family, and even order levels, with 20.7% of viruses found in more than two mammalian species. Our best models for both the number of cross-species transmitted viruses and viral spillover potential revealed significant effects at the level of mammalian order. This was confirmed by the different slope of the fit curves for total virus richness and the number of cross-species transmitted viruses in bats, rodents, and shrews, as well as the absence of host sample size

effect in shrews. Strikingly, viruses in rodents had a significantly higher potential of being found in other host species than those viruses in bats and shrews, even though shrews harbored the highest number of both total and cross-species transmitted viruses, and bats have been reported to harbor a greater proportion of zoonotic viruses than rodents.^{14,26,27} This might in part be explained by the wide geographic range, large population sizes, and high population densities of some rodent species.^{19,20,47} Notably, despite the widely held notion that bats are special reservoirs due to their association with a large number of high-profile zoonotic viruses (e.g., SARS-CoV, SARS-CoV-2, rabies viruses, and Nipah virus), viruses from these families in bats did not exhibit a greater potential for cross-species transmission. Finally, viruses with multi-organ distributions within individual hosts were more likely to spread to other host species.

Also of note was that the data generated here indicated that viruses were detected in multiple internal organs when their abundance in the primary target organ reached a particular level. This pattern was observed in both vertebrate and invertebrate-associated viruses. Although variation in organ distribution and the intra-organ abundance of viruses may reflect differences in the receptors used and their tissue distributions,²¹ the spillover of viruses from their primary target organs into other internal organs may assist in the overall process of host adaptation.

Limitations of the study

This study faced two major limitations. First, the analysis was performed based on pooled animal samples, which may reduce the sensitivity of virus discovery. Although biological replicates were used for predominant species to identify as many viruses as possible, we cannot exclude that viruses at low abundance went undetected. Second, the best model to identify the potential for cross-species virus transmission potential only explained ~21% of the total deviance, indicating that there are additional explanatory factors that remain to be discovered.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **METHOD DETAILS**
 - Study design and sample collection
 - RNA extraction, library construction, and sequencing
 - Data processing
 - Virus identification, quantification, and intra-host distribution
 - Virus confirmation and genome extension
 - Phylogenetic analysis
 - Ecological factors governing virome composition and viral cross-species transmission in small mammals
 - Mammal diversity

- Validation of the method for sample mixing
- Host-virus correlation network
- Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2023.08.029>.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (grants 32130002, 31930001, 32041004, and U22A20526). Y.-M.C. is supported by the Shanghai Rising-Star Program (grant no. 21QA1407800). E.C.H. is supported by the NHMRC Investigator Grant (GNT2017197) and AIR@InnoHK, administered by the Innovation and Technology Commission, Hong Kong, China.

AUTHOR CONTRIBUTIONS

Y.-Z.Z. conceived, designed, and led the study. X.-D.L., J.-H.T., and M.-R.W. performed sample collection and field surveys. Y.-M.C. and S.-J.H. analyzed the data. S.-J.H., J.-X.L., X.-Q.L., Y.-Y.P., R.-X.H., and Y.-M.C. performed the experiments. Y.-M.C., S.-J.H., E.C.H., and Y.-Z.Z. wrote the paper with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 25, 2023

Revised: July 13, 2023

Accepted: August 22, 2023

Published: September 20, 2023

REFERENCES

1. Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007). Origins of major human infectious diseases. *Nature* 447, 279–283. <https://doi.org/10.1038/nature05775>.
2. Holmes, E.C., Goldstein, S.A., Rasmussen, A.L., Robertson, D.L., Crits-Christoph, A., Wertheim, J.O., Anthony, S.J., Barclay, W.S., Boni, M.F., Doherty, P.C., et al. (2021). The origins of SARS-CoV-2: A critical review. *Cell* 184, 4848–4856. <https://doi.org/10.1016/j.cell.2021.08.017>.
3. Li, C.X., Shi, M., Tian, J.H., Lin, X.D., Kang, Y.J., Chen, L.J., Qin, X.C., Xu, J., Holmes, E.C., and Zhang, Y.Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* 4, e05378. <https://doi.org/10.7554/eLife.05378>.
4. Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S., et al. (2016). Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543. <https://doi.org/10.1038/nature20167>.
5. Shi, M., Lin, X.D., Chen, X., Tian, J.H., Chen, L.J., Li, K., Wang, W., Eden, J.S., Shen, J.J., Liu, L., et al. (2018). The evolutionary history of vertebrate RNA viruses. *Nature* 556, 197–202. <https://doi.org/10.1038/s41586-018-0012-7>.
6. Chen, Y.M., Sadiq, S., Tian, J.H., Chen, X., Lin, X.D., Shen, J.J., Chen, H., Hao, Z.Y., Wille, M., Zhou, Z.C., et al. (2022). RNA viromes from terrestrial sites across China expand environmental viral diversity. *Nat. Microbiol.* 7, 1312–1323. <https://doi.org/10.1038/s41564-022-01180-2>.
7. Edgar, R.C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D., Novakovsky, G., Buchfink, B., Al-Shayeb, B., et al. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature* 602, 142–147. <https://doi.org/10.1038/s41586-021-04332-2>.

8. Harvey, E., and Holmes, E.C. (2022). Diversity and evolution of the animal virome. *Nat. Rev. Microbiol.* 20, 321–334. <https://doi.org/10.1038/s41579-021-00665-x>.
9. Qin, X.C., Shi, M., Tian, J.H., Lin, X.D., Gao, D.Y., He, J.R., Wang, J.B., Li, C.X., Kang, Y.J., Yu, B., et al. (2014). A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *Proc. Natl. Acad. Sci. USA* 111, 6744–6749. <https://doi.org/10.1073/pnas.1324194111>.
10. Wang, Z.D., Wang, B., Wei, F., Han, S.Z., Zhang, L., Yang, Z.T., Yan, Y., Lv, X.L., Li, L., Wang, S.C., et al. (2019). A new segmented virus associated with human febrile illness in China. *N. Engl. J. Med.* 380, 2116–2125. <https://doi.org/10.1056/NEJMoa1805068>.
11. Li, K., Lin, X.D., Wang, W., Shi, M., Guo, W.P., Zhang, X.H., Xing, J.G., He, J.R., Wang, K., Li, M.H., et al. (2015). Isolation and characterization of a novel arenavirus harbored by rodents and shrews in Zhejiang Province, China. *Virology* 476, 37–42. <https://doi.org/10.1016/j.virol.2014.11.026>.
12. Blasdel, K.R., Duong, V., Eloit, M., Chretien, F., Ly, S., Hul, V., Deubel, V., Morand, S., and Buchy, P. (2016). Evidence of human infection by a new Mammarenavirus endemic to southeastern Asia. *eLife* 5, e13135. <https://doi.org/10.7554/eLife.13135>.
13. Liu, X., Zhang, X., Wang, Z., Dong, Z., Xie, S., Jiang, M., Song, R., Ma, J., Chen, S., Chen, K., et al. (2020). A tentative tamdy Orthonaivirus related to febrile illness in northwestern China. *Clin. Infect. Dis.* 70, 2155–2160. <https://doi.org/10.1093/cid/ciz602>.
14. Olival, K.J., Hosseini, P.R., Zambrana-Torrel, C., Ross, N., Bogich, T.L., and Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature* 546, 646–650. <https://doi.org/10.1038/nature22975>.
15. Mollentze, N., and Streicker, D.G. (2020). Viral zoonotic risk is homogeneous among taxonomic orders of mammalian and avian reservoir hosts. *Proc. Natl. Acad. Sci. USA* 117, 9423–9430. <https://doi.org/10.1073/pnas.1919176117>.
16. Wille, M., Geoghegan, J.L., and Holmes, E.C. (2021). How accurately can we assess zoonotic risk? *PLoS Biol.* 19, e3001135. <https://doi.org/10.1371/journal.pbio.3001135>.
17. Lipsitch, M., Barclay, W., Raman, R., Russell, C.J., Belser, J.A., Cobey, S., Kasson, P.M., Lloyd-Smith, J.O., Maurer-Stroh, S., Riley, S., et al. (2016). Viral factors in influenza pandemic risk assessment. *eLife* 5, e18491. <https://doi.org/10.7554/eLife.18491>.
18. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. <https://doi.org/10.1038/s41586-020-2008-3>.
19. Wilson, D.E., and Reeder, D.M. (2005). *Mammal Species of the World. A Taxonomic and Geographic Reference, Third Edition* (Johns Hopkins University Press).
20. Meerburg, B.G., Singleton, G.R., and Kijlstra, A. (2009). Rodent-borne diseases and their risks for public health. *Crit. Rev. Microbiol.* 35, 221–270. <https://doi.org/10.1080/10408410902989837>.
21. 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. <https://doi.org/10.1038/s41579-020-0394-z>.
22. Lin, X.D., Wang, W., Hao, Z.Y., Wang, Z.X., Guo, W.P., Guan, X.Q., Wang, M.R., Wang, H.W., Zhou, R.H., Li, M.H., et al. (2017). Extensive diversity of coronaviruses in bats from China. *Virology* 507, 1–10. <https://doi.org/10.1016/j.virol.2017.03.019>.
23. Tirera, S., de Thoisy, B., Donato, D., Bouchier, C., Lacoste, V., Franc, A., and Laverne, A. (2021). The Influence of habitat on viral diversity in Neotropical rodent hosts. *Viruses* 13, 1690. <https://doi.org/10.3390/v13091690>.
24. Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., et al. (2016). Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *ISME J.* 10, 609–620. <https://doi.org/10.1038/ismej.2015.138>.
25. Wu, Z., Lu, L., Du, J., Yang, L., Ren, X., Liu, B., Jiang, J., Yang, J., Dong, J., Sun, L., et al. (2018). Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. *Microbiome* 6, 178. <https://doi.org/10.1186/s40168-018-0554-9>.
26. Luis, A.D., Hayman, D.T., O'Shea, T.J., Cryan, P.M., Gilbert, A.T., Pulliam, J.R., Mills, J.N., Timonin, M.E., Willis, C.K., Cunningham, A.A., et al. (2013). A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc. Biol. Sci.* 280, 20122753. <https://doi.org/10.1098/rspb.2012.2753>.
27. Luis, A.D., O'Shea, T.J., Hayman, D.T.S., Wood, J.L.N., Cunningham, A.A., Gilbert, A.T., Mills, J.N., and Webb, C.T. (2015). Network analysis of host-virus communities in bats and rodents reveals determinants of cross-species transmission. *Ecol. Lett.* 18, 1153–1162. <https://doi.org/10.1111/ele.12491>.
28. Déjosez, M., Marin, A., Hughes, G.M., Morales, A.E., Godoy-Parejo, C., Gray, J.L., Qin, Y., Singh, A.A., Xu, H., Juste, J., et al. (2023). Bat pluripotent stem cells reveal unusual entanglement between host and viruses. *Cell* 186, 957–974.e28. <https://doi.org/10.1016/j.cell.2023.01.011>.
29. Bergner, L.M., Orton, R.J., Benavides, J.A., Becker, D.J., Tello, C., Biek, R., and Streicker, D.G. (2020). Demographic and environmental drivers of metagenomic viral diversity in vampire bats. *Mol. Ecol.* 29, 26–39. <https://doi.org/10.1111/mec.15250>.
30. Rodríguez-Nevado, C., Lam, T.T., Holmes, E.C., and Pagán, I. (2018). The impact of host genetic diversity on virus evolution and emergence. *Ecol. Lett.* 21, 253–263. <https://doi.org/10.1111/ele.12890>.
31. ICTV (2022). The ICTV report on virus classification and taxon nomenclature. <https://ictv.global/report>.
32. Zhang, X.A., Li, H., Jiang, F.C., Zhu, F., Zhang, Y.F., Chen, J.J., Tan, C.W., Anderson, D.E., Fan, H., Dong, L.Y., et al. (2022). A zoonotic Henipavirus in febrile patients in China. *N. Engl. J. Med.* 387, 470–472. <https://doi.org/10.1056/NEJMc2202705>.
33. 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. <https://doi.org/10.1038/s41586-018-0010-9>.
34. Chen, H.X., Wang, Z., and Tang, S.Z. (1992). *Surveillance and Investigation of Epidemiologic Hemorrhagic Fever in China* (Beijing Science and Technology Press).
35. Hu, D., Zhu, C., Ai, L., He, T., Wang, Y., Ye, F., Yang, L., Ding, C., Zhu, X., Lv, R., et al. (2018). Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg. Microbes Infect.* 7, 154. <https://doi.org/10.1038/s41426-018-0155-5>.
36. Murakami, S., Kitamura, T., Suzuki, J., Sato, R., Aoi, T., Fujii, M., Matsugo, H., Kamiki, H., Ishida, H., Takenaka-Uema, A., et al. (2020). Detection and characterization of bat Sarbecovirus phylogenetically related to SARS-CoV-2, Japan. *Emerg. Infect. Dis.* 26, 3025–3029. <https://doi.org/10.3201/eid2612.203386>.
37. Temmam, S., Vongphaylo, K., Baquero, E., Munier, S., Bonomi, M., Regnault, B., Douangboubpha, B., Karami, Y., Chretien, D., Sanamxay, D., et al. (2022). Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* 604, 330–336. <https://doi.org/10.1038/s41586-022-04532-4>.
38. Wacharapluesadee, S., Tan, C.W., Maneeorn, P., Duengkae, P., Zhu, F., Joyjinda, Y., Kaewpom, T., Chia, W.N., Ampoot, W., Lim, B.L., et al. (2021). Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* 12, 972. <https://doi.org/10.1038/s41467-021-21240-1>.
39. Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C., et al. (2020). A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of

- the spike protein. *Curr. Biol.* 30, 3896. <https://doi.org/10.1016/j.cub.2020.09.030>.
40. Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. <https://doi.org/10.1038/s41586-020-2012-7>.
41. Guo, W.P., Lin, X.D., Wang, W., Tian, J.H., Cong, M.L., Zhang, H.L., Wang, M.R., Zhou, R.H., Wang, J.B., Li, M.H., et al. (2013). Phylogeny and origins of hantaviruses harbored by bats, insectivores, and rodents. *PLoS Pathog.* 9, e1003159. <https://doi.org/10.1371/journal.ppat.1003159>.
42. Gonzalez, J.P., Emonet, S., de Lamballerie, X., and Charrel, R. (2007). Arenaviruses. *Curr. Top. Microbiol. Immunol.* 315, 253–288. https://doi.org/10.1007/978-3-540-70962-6_11.
43. Weaver, S.C., Forrester, N.L., Liu, J., and Vasilakis, N. (2021). Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence. *Nat. Rev. Microbiol.* 19, 184–195. <https://doi.org/10.1038/s41579-020-00482-8>.
44. 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. <https://doi.org/10.1038/nrmicro.2017.45>.
45. 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. <https://doi.org/10.1038/s41586-022-04788-w>.
46. Lin, X.D., Wang, W., Guo, W.P., Zhang, X.H., Xing, J.G., Chen, S.Z., Li, M.H., Chen, Y., Xu, J., Plyusnin, A., and Zhang, Y.Z. (2012). Cross-species transmission in the speciation of the currently known Murinae-associated hantaviruses. *J. Virol.* 86, 11171–11182. <https://doi.org/10.1128/JVI.00021-12>.
47. Lin, X.D., Guo, W.P., Wang, W., Zou, Y., Hao, Z.Y., Zhou, D.J., Dong, X., Qu, Y.G., Li, M.H., Tian, H.F., et al. (2012). Migration of Norway rats resulted in the worldwide distribution of Seoul hantavirus today. *J. Virol.* 86, 972–981. <https://doi.org/10.1128/JVI.00725-11>.
48. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
49. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>.
50. Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
51. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. <https://doi.org/10.1038/nmeth.3176>.
52. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
53. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>.
54. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
55. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
56. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>.
57. Sayers, E.W., Beck, J., Bolton, E.E., Bourexis, D., Brister, J.R., Canese, K., Comeau, D.C., Funk, K., Kim, S., Klimke, W., et al. (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 49, D10–D17. <https://doi.org/10.1093/nar/gkaa892>.
58. Zhang, Y.Z., Shi, M., and Holmes, E.C. (2018). Using metagenomics to characterize an expanding virosphere. *Cell* 172, 1168–1172. <https://doi.org/10.1016/j.cell.2018.02.043>.
59. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. <https://doi.org/10.1093/nar/gks1219>.
60. Shi, W., Shi, M., Que, T.C., Cui, X.M., Ye, R.Z., Xia, L.Y., Hou, X., Zheng, J.J., Jia, N., Xie, X., et al. (2022). Trafficked Malayan pangolins contain viral pathogens of humans. *Nat. Microbiol.* 7, 1259–1269. <https://doi.org/10.1038/s41564-022-01181-1>.
61. Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., Granger, B., Green, L., Howd, T., Mason, T., et al. (2018). Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* 19, 332. <https://doi.org/10.1186/s12864-018-4703-0>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Samples are described in Table S1	This paper	N/A
Critical commercial assays		
RNeasy® Plus Universal Mini Kit	QIAGEN	Cat. #73404
KAPA RNA HyperPrep Kit with RiboErase (HMR)	KAPA Biosystems	Cat. #KK8561
KAPA Unique Dual-Indexed Adapters Kit	KAPA Biosystems	Cat. #KK8727
Deposited data		
Sequencing data of all RNA libraries	This paper	NCBI-SRA BioProject: PRJNA953205
Data of all virus genomes	This paper	NCBI-Genbank: MT210604-MT210623, MZ328236-MZ328303, OM030289-OM030338, ON321888-ON321889, OQ715366-OQ716293, OQ802697-OQ802786
Software and algorithms		
Trimmomatic (v0.39)	Bolger et al. ⁴⁸	http://www.usadellab.org/cms/index.php?page=trimmomatic
Trinity (v2.9.0)	Grabherr et al. ⁴⁹	https://github.com/trinityrnaseq/trinityrnaseq/wiki
Megahit (v1.1.3)	Li et al. ⁵⁰	https://github.com/voutcn/megahit
Diamond Blastx (v0.9.24.125)	Buchfink et al. ⁵¹	N/A
blastn (v2.6.0)	–	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Bowtie2 (v2.3.5.1)	Langmead and Salzberg ⁵²	https://sourceforge.net/projects/bowtie-bio/
SAMtools (v0.1.19)	Li ⁵³	http://www.htslib.org/
Geneious Prime (v2021.0.3)	–	https://www.geneious.com/prime/
R package pheatmap (v1.0.12)	–	https://cran.r-project.org/web/packages/pheatmap/index.html
MAFFT (v7.450)	Katoh and Standley ⁵⁴	https://mafft.cbrc.jp/alignment/software/
TrimAl (v1.4.rev22)	Capella-Gutiérrez et al. ⁵⁵	http://www.drive5.com/muscle/manual/index.html
IQ-TREE (v1.6.12)	–	http://www.iqtree.org/
FigTree (v1.4.4)	–	http://tree.bio.ed.ac.uk/software/figtree/
R package olsrr (v0.5.3)	–	https://cran.r-project.org/web/packages/olsrr
R package vegan (v2.5-7)	–	https://cran.r-project.org/web/packages/vegan
R package stats (v4.1.0)	–	https://cran.r-project.org/web/packages/stats
Cytoscape (v3.9.1)	Shannon et al. ⁵⁶	https://cytoscape.org/
Other		
Sequencing systems	illumina	Novaseq6000
SILVA database	Quast et al. ⁵⁹	https://www.arb-silva.de
Genbank database	Sayers et al. ⁵⁷	https://www.ncbi.nlm.nih.gov/genbank/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and information should be directed to and will be fulfilled by the lead contact, Dr. Yong-Zhen Zhang (zhangyongzhen@fudan.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The sequence reads generated in this study are available at the NCBI Sequence Read Archive (SRA) database under BioProject accession PRJNA953205. All viral sequences generated in this study have been deposited in GenBank under the accession numbers MT210604-MT210623, MZ328236-MZ328303, OM030289-OM030338, ON321888-ON321889, OQ715366-OQ716293, OQ802697-OQ802786. All other data and code used in this study are available in the [supplemental information](#).

METHOD DETAILS

Study design and sample collection

This study was designed to better understand the diversity, ecology and evolution of viruses in bats, rodents and shrews. To ensure that sample collection was representative, four locations were selected for the field survey of wild mammals in Hubei (Jingmen and Wufeng) and Zhejiang (Wenzhou and Longquan) provinces, China ([Figures 1](#) and [S1](#)). These locations represent different natural habitats in subtropical regions, including two agricultural areas with woodland and mountain components (Jingmen and Wenzhou) and two mountain forests (Longquan and Wufeng) located in coastal (Zhejiang province) and inland (Hubei province) areas, respectively. The study habitats in both Longquan and Wufeng are also at higher altitude (600–1200 meters) than those in Jingmen and Wenzhou (<500 meters), and are less affected by human activities.

Rodents and shrews were captured using cages via baits in agricultural areas with woodland regions in Jingmen and Wenzhou, and from forested areas in Longquan and Wufeng, while bats were captured in mountain caves. To help ensure the richness and abundance of the small mammals captured and to minimize sampling bias, two strategies were adopted. First, the small mammals were captured by team members from local Centers for Disease Control and Prevention (CDC) in Longquan, Wenzhou and Wuhan, who have more than ten years of experience in field surveys with strong local knowledge of the biodiversity and behaviour of animals sampled. Second, all mammals were captured during winter and early spring of 2016–2017, during which time bats are in hibernation, with multiple species co-habiting in their roosts, and rodents and shrews face food shortages. Hence, as best we could, we tried to sample animals in manner that reflected their relative population sizes in nature.

The procedures for sampling and sample processing were reviewed and approved by the ethics committee of National Institute for Communicable Disease Control and Prevention of the China CDC. No statistical methods were used to predetermine sample size as this requires information on virus prevalence which was necessarily unknown. All mammals were caught alive and treated strictly according to the guidelines for the Laboratory Animal Use and Care from China CDC.

Animal dissections were performed under ether anesthesia, and every effort was made to minimize suffering. Five types of internal organs including gut (with feces), liver, spleen, kidney and lung were harvested for caught mammals. Tissue samples were kept in portable refrigerating equipment with dry ice before being transferred to -80°C for storage.

Host species identification was initially performed by experienced field biologists based on morphological characteristics, and was later confirmed by sequencing and analyzing the partial cytochrome c oxidase (COI) gene from each sample (600–700 nucleotides near 5' of the gene).⁵

RNA extraction, library construction, and sequencing

To describe the virome composition in these small mammals in the least biased way possible, the animals captured were selected for meta-transcriptomic analysis based on two criteria: (i) for rare species with fewer than 36 individuals from one sampling location, all were included in the analysis; (ii) for the remaining species, 36, 72, or 108 individual animals were randomly selected according to the total sample size, generating 1–3 biological replicates (i.e., each contained 36 individuals). As a result, >25% individuals from each of sampling locations were included for each dominant species ([Table S1](#)). There were five exceptions to these criteria because of the low quality of some samples: *Apodemus agrarius* from Wenzhou, *Rhinolophus affinis* and *Niviventer niviventer* from Longquan, and *Anourosorex squamipes* and *Crociodura attenuate* from Wufeng ([Table S1](#)).

RNA was extracted from various animal tissues based on tissue type, mammalian species and sampling location. Briefly, the same tissues (i.e., liver, spleen, lung, kidney and gut) of each of the 36 individual animals (or all individuals for rare species) from the same species from the same location were mixed in equal quantity and homogenized. However, as the individual numbers were low for some of rare species and/or organs, such that the sample amount obtained from did not meet the requirements for reliable RNA extraction and library construction, different organ tissues were also mixed for 41 libraries ([Table S2](#)). For example, library WFSPSG-ZhongHua contained spleen, kidney and liver samples from 13 individuals of *Apodemus draco* from Wufeng. To ensure the uniformity of sample homogenates, mixing of the 36 individuals was performed in two steps: (i) the samples from six individual animals were first mixed and homogenized using 1 ml QIAzol (Qiagen), and (ii) 100 μl of each homogenate generated in step (i) were pooled and then further homogenized using another 1 ml QIAzol. Total RNA was then extracted from the final homogenates using the RNeasy® Plus Universal Mini Kit (Qiagen) following the manufacturer's introductions. RNA quality was determined using an Agilent 4200 Bioanalyzer (Agilent Technologies) and quantified using NanoDropOne (Thermo Scientific, USA) before library construction and sequencing. All RNA solutions were stored at -80°C until library construction.

The KAPA RNA HyperPrep Kit with RiboErase (HMR) (KAPA Biosystems) was used for all library preparations. To provide an unbiased assessment of sample composition, no enrichment of viral content was conducted during sample processing.⁵⁸ The quantity and quality of RNA libraries were assessed using a Qubit4.0 fluorometer (Invitrogen) and an Agilent 4200 bioanalyzer (Agilent

Technologies). The average fragmentation size for these libraries was 300~350 bp. Accordingly, paired-end (150bp reads) sequencing for each RNA library was performed on the Illumina Novaseq 6000 platform by Novogene. Detailed information of all RNA libraries is provided in [Table S2](#).

Data processing

Sequencing reads were first adaptor- and quality-trimmed using the Trimmomatic program (v0.39)⁴⁸ with the following parameters: LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:36. The remaining reads were assembled *de novo* using the Trinity (v2.9.0)⁴⁹ or the Megahit (v1.1.3)⁵⁰ programs with default parameter settings. No filtering of host or bacterial reads was performed before assembly. To identify viral contigs, all assembled contigs were compared against the non-redundant (nr) protein database using Diamond Blastx (v0.9.24.125)⁵¹ with an e-value threshold of 1×10^{-5} . Sequences taxonomically annotated as from the kingdom of “Viruses” (using the “-taxonmap” option) based on the top blast hit were initially identified as potential virus sequences. For robustness, these potential virus contigs were subsequently compared against the non-redundant nucleotide (nt) database using blastn (v2.6.0) to remove host genome sequences, endogenous viral elements and artificial vector sequences. The resulting contigs were then further validated and quantified by read mapping (See next section).

Virus identification, quantification, and intra-host distribution

To quantify virus abundance, we first removed reads associated with ribosomal RNA by mapping quality-trimmed sequencing reads to rRNA contigs downloaded from the SILVA database.⁵⁹ The remaining reads were mapped to the above potential viral contigs using Bowtie2 (v2.3.5.1)⁵² with an end-to-end alignment. SAMtools (v0.1.19)⁵³ was used to sort and index these alignments, from which the read counts for each contig were obtained. Virus abundance was then calculated and normalized as the number of viral reads per million from the total non-rRNA reads in each library (RPM). To reduce false-positives, the data were filtered to viral contigs with an RPM ≥ 1 . The returned data was used to estimate the abundance of each viral family within a library.

Potential host associations for the obtained viral contigs were preliminarily identified based on the taxonomic information of the Blastx results and were further confirmed according to their phylogenetic relationships to viruses with known host associations. Briefly, viral contigs that fell into known vertebrate and/or invertebrate-associated virus groups were retained, while those that clustered with bacterial, fungal or plant virus groups were excluded. Consequently, 13,913 of the 34,427 identified viral contigs were considered to be bacterial, fungal, and plant-associated and were thus excluded from subsequent analyses. Viral contigs assumed to have vertebrate and/or invertebrate hosts were then subjected to more precise species assignments. Contigs with unassembled overlaps were merged (by host species and sample location) to obtain longer viral contigs using Geneious Prime (v2021.0.3). Next, all viral contigs were annotated using Geneious Prime, and those covering the RdRp for RNA viruses or the replicase for DNA viruses were retained. Species assignment of the resulting viral contigs was performed using the species demarcation criteria of each virus genus laid down by the ICTV.³¹ For genera that lacked clear species demarcation criteria, a relatively strict threshold of 80% amino acid identity to known virus species for the RdRp or replicase was used. Details of the species demarcation criteria used in this study are provided in [Table S4](#). These criteria were also used to identify novel virus species and identical virus species across sequencing libraries and mammal species. Specifically, if a virus species (i.e., that passing the species demarcation described in [Table S4](#)) was discovered in more than one mammalian species, it was thought to be able to jump between these host species. To reduce the impact of possible index-hopping, viruses were assumed to be the result of contamination from another library if the read count of a virus in one library was less than 0.1% of the highest read count for that virus among the other libraries within the same sequencing lane.^{6,60,61} Overall, this process resulted in the identification of 1095 viral sequences assigned to 669 vertebrate- and invertebrate-associated virus species ([Table S3](#)).

To determine the intra-host distribution of these viruses, Bowtie2 (v2.3.5.1) was used to map the non-rRNA reads of each tissue library to each of the respective virus contigs identified in the same mammal species from the same sampling location. Virus abundance in each tissue was calculated as RPM ([Table S3](#)). To aid visualization, only viruses identified in independent tissue libraries (i.e., unmixed tissues) were plotted in the heatmap ([Figures 3 and S3; Table S5](#)). The organ that had the highest abundance of a virus compared to other organs was assumed to represent the primary target organ of the virus in question. Similarly, the tissue with the second highest abundance of the same virus was defined as the secondary organ. After filtering with a threshold of RPM ≥ 1 , the organ distribution of each virus was visualized in R (v4.1.0) using the *pheatmap* package (v1.0.12).

To further explore the distribution of viruses within mammal host, we performed correlation analysis on the highest and the second highest abundance of some notable viruses in tissue libraries, employing linear, nonlinear and exponential models. Briefly, four models (lm, gam, loess, and nls) were used to fit the trend line by calling the function *stat_smooth* in the R package ggplot2. As these models showed a similar trend – i.e., that viruses were detected in other internal organs when their abundance in the primary target organs reached a particular level and the thresholds predicted by these models were similar (i.e., the x value when y=0) – the simple linear regression model was adopted to display the results and predict the threshold ([Figure 3C](#)). A principal components analysis (PCA) was performed to assess the difference of intra-host distribution virus among viral families and mammalian hosts. The scattered distribution of viruses from a family or a mammal host indicates that viruses from the corresponding family or host are more likely to be distributed in multiple organs. The *ggrepel* package (v0.9.1) was used for visualization of the PCA result in R.

Virus confirmation and genome extension

To confirm the presence of the viruses detected, RT-PCR assays were performed using specific primers designed based on the assembled viral contigs. The target PCR products were validated by Sanger sequencing. For the seven newly identified chuviruses, we examined their prevalence in each organ of each mammal individual from the same sequencing library (Table S7).

We also attempted to obtain complete or nearly complete genomes of some important and newly identified viruses using RT-PCR assays, RNA circularization and the 5'/3' RACE kits (Takara Bio). Among these, the non-RdRp gene segments of segmented RNA viruses (e.g., arenavirus, hantaviruses, reoviruses) were identified based on sequence similarity to the amino acid sequences of related reference viruses, and further validated using criteria described previously, namely: (i) a similar sequencing depth of each segment, and (ii) conserved sequences in non-coding regions of each segment.^{4,6}

Phylogenetic analysis

Vertebrate and invertebrate-associated virus sequences were first categorized into major viral clades based on the Diamond Blastx analysis. To validate the taxonomic assignment and examine the phylogenetic relationships among the newly identified viruses, phylogenetic analysis was performed on each of these viral clades using either the RdRp (RNA viruses) or replicase (DNA virus) proteins. To this end, 1887 reference genomes comprising available sequences of all viral clades detected in this study were downloaded from GenBank. These reference genomes included most member species of each viral clade classified by ICTV, as well as the top Diamond Blastx hit of the newly-identified viruses. The virus sequences identified in this study were aligned with the reference sequences of the same viral clade using the L-INS-i algorithm implemented in the program MAFFT (v7.450).⁵⁴ Ambiguously aligned regions were removed by TrimAl (v1.4.rev22). Maximum likelihood phylogenetic trees were then estimated based on the amino acid alignments using IQ-TREE (v1.6.12) employing the best-fit substitution model selected by the setting “-m MFP”. Branch support was assessed using 1,000 SH-like approximate likelihood ratio test (SH-aLRT) replicates. All trees were visualized using FigTree (v1.4.4).

Ecological factors governing virome composition and viral cross-species transmission in small mammals

To determine the ecological factors that have impacted the number of virus species (i.e., virus species richness), total virus abundance, and the number of cross-species transmitted viruses, we fitted three independent sets of GLM models that contained all possible combinations of the available and potential variables. Briefly, for models of virus species richness and abundance, six variables comprising host order, host sample size, host geographic distribution (i.e., present in one, two, three or four locations), sampling location, host habitat altitude (mountain or agricultural areas), and host habitat ecology (inshore or inland) were included. For models of the number of cross-species transmitted viruses, virus species richness per host species and total virus abundance per host species were included with the same variables as the other two models.

For the model predicting the potential of viral cross-species transmission, we used mixed GAM models with the cross-species transmission status of a virus set as a binary response variable. A virus family-level taxonomy that represents the relationships between viruses was included in the GAM models as a random effect. As the viruses newly identified in this study were too diverse to be represented in a single phylogeny (i.e., expansive sequence alignments are not robust), a variance-covariance matrix was generated to represent the virus taxonomic relationships.¹⁵ This matrix was generated at the virus family level because most of the newly identified viruses were only distantly related to known viruses and could not be annotated to lower taxonomic level (i.e., genus). This virus taxonomic random effect was combined with all possible combinations of additional variables including host order, host sample size, virus abundance, whether or not a virus was multi-organ distributed within the host or was invertebrate-associated. The GAM models were fit using the *gam* function in the R package *mgcv* (v1.8-41) with the “REML” method. In addition, we fitted two other independent sets of GAM models which added a random effect of virus family nested within host order or host family, respectively, to infer whether there is a host-specific virus family effect (Table 1; Models 5 and 6).

All fit models were subsequently ranked by AIC, and that with the lowest AIC value of each set was selected as the best-fit model. The deviance explained by each variable was calculated by comparing the full model including all variables and the submodels including all additional variables except the test variable. Accordingly, the deviance explained by a model was calculated as $(D_n - D_i)/D_n$, while the deviance explained by each variable in the full model was calculated as $(D_f - D_i)/D_n$, where D_n is the deviance of an intercept-only model, D_f is the deviance of the full model, and D_i is the deviance of submodel *i*. In addition, the partial effect of each variable in the best-fit model was calculated as the prediction of the estimate when keeping the other categorical variables at their most common value and the other numeric values at their median value.¹⁵

Mammal diversity

To determine the diversity of mammal species sampled, a phylogenetic tree was estimated using the COI amino acid sequences using the same method and parameters as described above. Principal coordinate analysis (PCoA) was conducted to evaluate the impact of habitat on mammalian composition, using the *vegan* (v2.5-7) and *stats* (v4.1.0) package in R. In addition, we performed a rarefaction analysis using R to assess whether the captured mammals reflected the true diversity of small mammals at each sampling location.

Validation of the method for sample mixing

To evaluate the impact of sample mixing on the diversity of the viruses identified, we mixed samples of different numbers and constructed RNA libraries. The gut and lung samples from 108 individuals of *Rattus norvegicus* from Wenzhou were selected for this purpose. Among these, the 36 individuals that were mixed and used to generate the libraries LCSC_hejia_36-1 and LCSF_hejia_36-1 (Table S2) were divided into three groups, each with 12 individuals. The gut and lung samples of these three groups were then mixed in equal quality and homogenized. At this point, six libraries for gut and lung samples were generated (e.g., LCSC_hejia_12-1 to LCSC_hejia_12-3 for gut). In addition, another 36 individuals that were mixed to generate the libraries LCSC_hejia_36-2 and LCSF_hejia_36-2 were added, and then divided into three groups of 24 individuals each. From this, six more libraries of gut and lung samples were generated, each containing 24 gut or lung samples. In total, we generated 18 libraries using mixed gut and lung samples of 12, 24 or 36 individuals, with three replicates for each mixed group (Table S2).

On the basis of the mixing process described above, viruses detected in libraries LCSC_hejia_12-1 and LCSC_hejia_12-2 were expected to be detected in library LCSC_hejia_24-1, while those found in LCSC_hejia_12-1, LCSC_hejia_12-2 and LCSC_hejia_12-3 should also be identified in LCSC_hejia_36-1, since the same individuals were used in these libraries. The total virome compositions generally met these expectations, with the exception of eight viral clades (Figure S2, in red). Notably, the missed detection of some viruses in libraries with greater sample mixing (i.e., 24 or 36 individuals) was remedied when biological replicates were employed, especially for viruses with high prevalence and abundance. For example, although caliciviruses were detected in two libraries of 12 gut samples (12-1C and 12-3C), they were not found in libraries 24-1C and 36-1C. However, these viruses could be identified when more samples were included (i.e., in libraries 24-2C, 24-3C, 36-2C). In sum, these data indicate that the sample mixing method employed had no obvious dilution effect and did not impact the detection of most viruses.

Host-virus correlation network

A host-virus correlation network was generated using the Cytoscape (v3.9.1) software⁵⁶ employing the “Prefuse Force Directed Layout” option. Node sizes were used to distinguish virus species (small nodes) and mammalian species (large nodes), node shapes were used to distinguish mammalian species sampled from different locations, and node colors were used to distinguish the three orders of small mammals. Topological analysis was performed using the “Analyze Network” function in Cytoscape.

Statistical analysis

To test whether the difference in mammalian composition between habitats and in virome composition between host orders and host habitats was significant, we performed a permutational multivariate analysis of variance (PERMANOVA, 9999 permutations) using the “*adonis2*” function in the R package *vegan* (v2.5-7), based on the Bray-Curtis distance matrixes generated using the “*vegdist*” function in the *vegan* package. Univariate statistical analysis was performed using the Wilcoxon rank sum test or the Kruskal-Wallis test in *ggpubr* package (v0.4.0) in R to compare continuous variables. Chi-squared tests or Fisher’s exact tests were used for the comparison of categorical variables, using the R *stats* package (v3.5.1). The validity of models was checked by testing overall the uniformity and dispersion of the simulated residuals using R *DHARMa* package (v0.4.6).

Supplemental figures

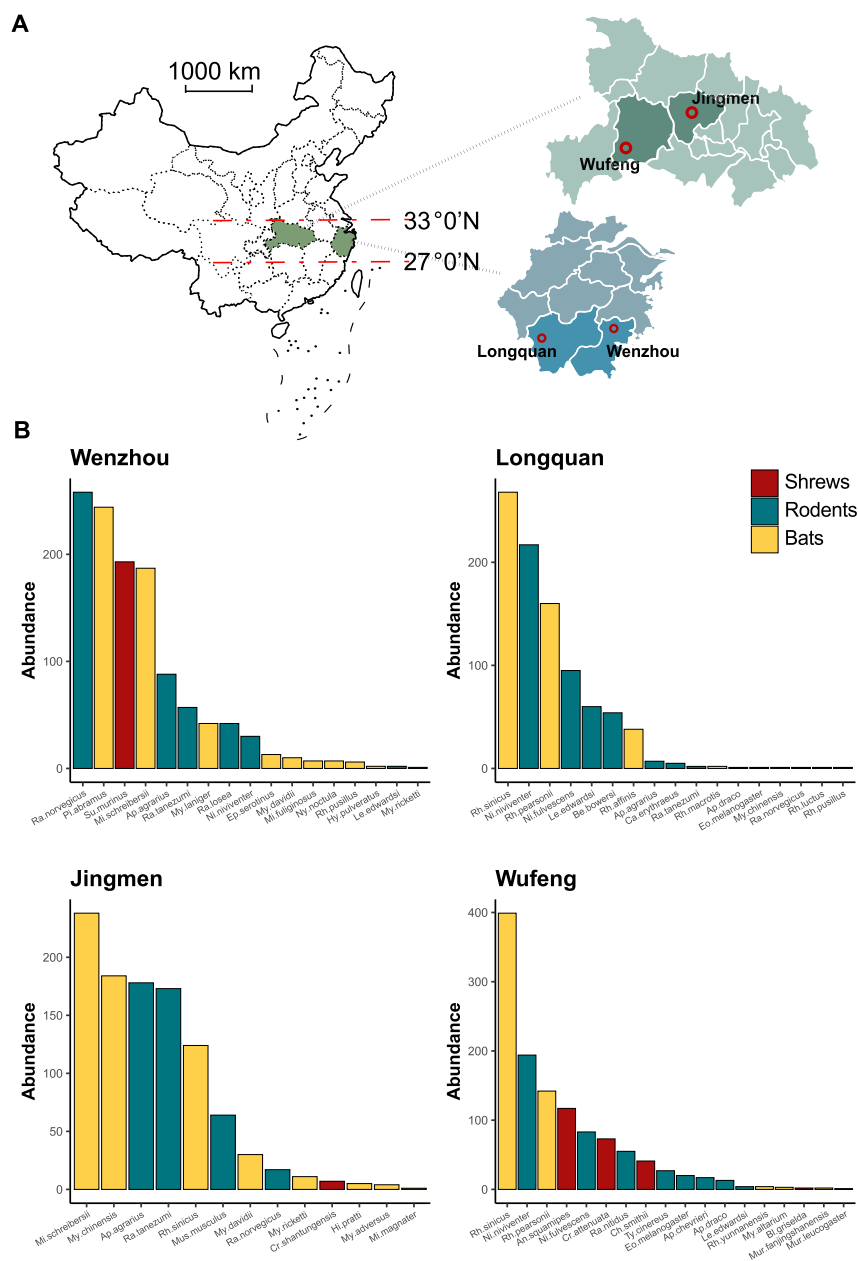
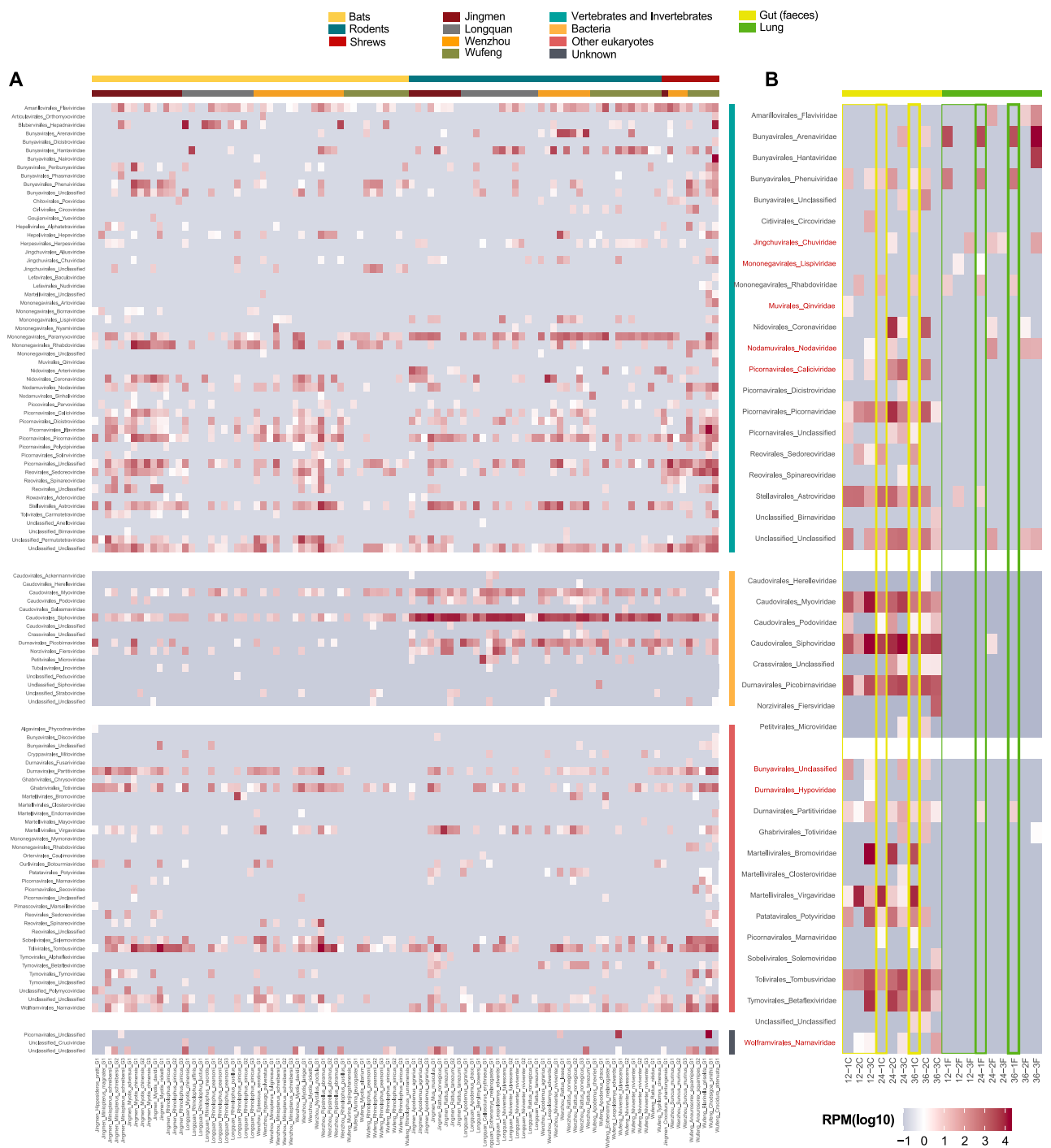


Figure S1. Wild small mammals sampled in this study, related to Figure 1

(A) Map of sampling locations.

(B) Sample size of each mammal species at each location.





(B) Abundance of invertebrate-associated viruses in internal organs.

(C) Relationship of virus abundance between the primary and the secondary target organs in each mammalian order.

(D) PCA plot showing the multi-organ distribution of viruses from three viral families in bats, rodents, and shrews.

Each point in (C) and (D) represents one virus in the family in question and one virus detected in one organ, respectively.

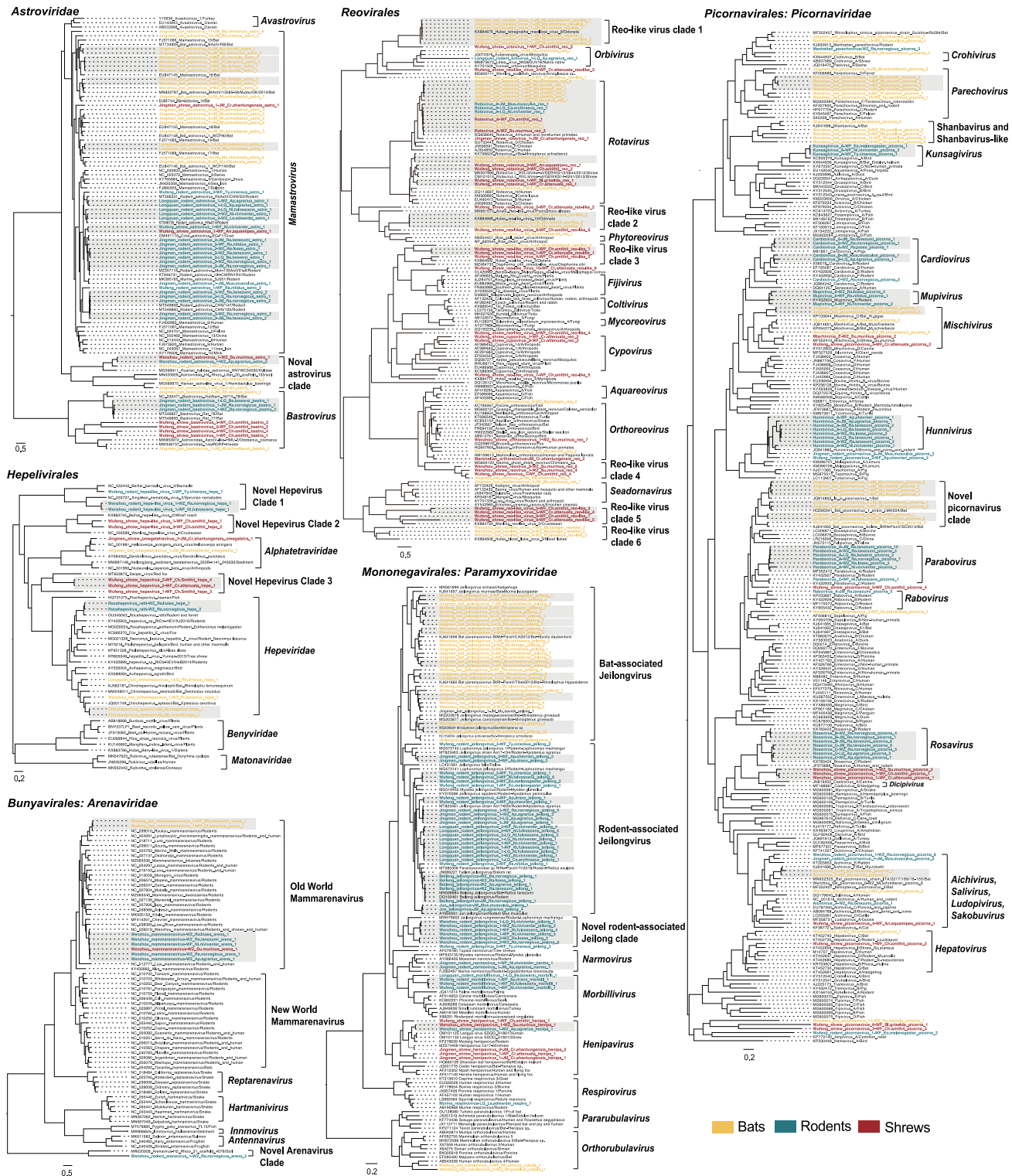


Figure S4. Phylogenetic trees of viruses with evolutionary significance or zoonotic potential, related to Figure 4

Phylogenetic trees were estimated based on the RdRp protein for RNA viruses and the replicase protein for DNA viruses. Viruses newly identified in this study are color marked by mammalian hosts, while previously published viruses are in black. The same viruses from different host species and sampling locations were labeled by gray squares. Branch lengths are scaled to the number of amino acid substitutions per site, as indicated by the scale bar. The tree is midpoint rooted for clarity only.

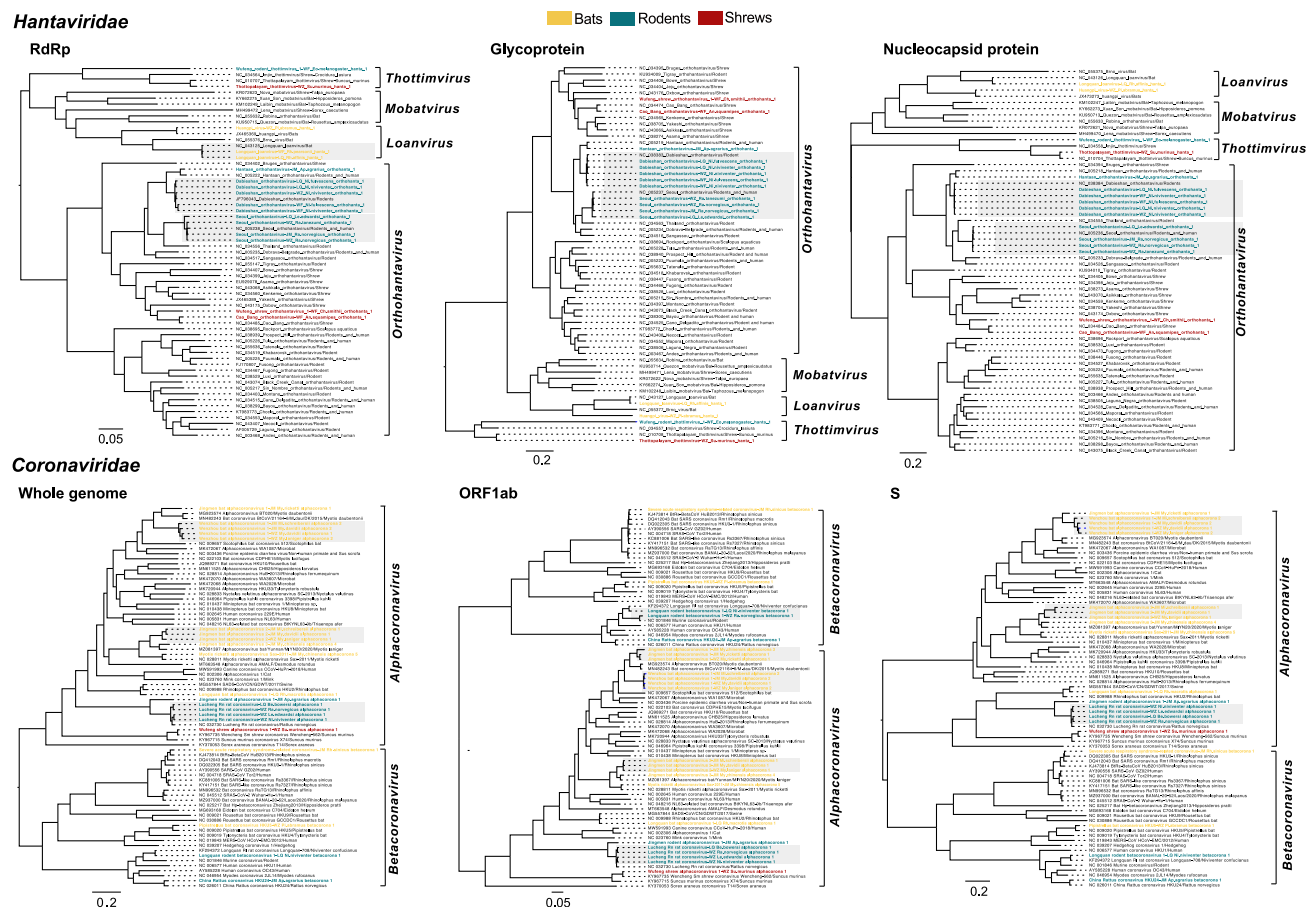


Figure S5. Phylogeny of the Hantaviridae and Coronaviridae, related to Figure 4
Figure legend follows that of Figure S4.

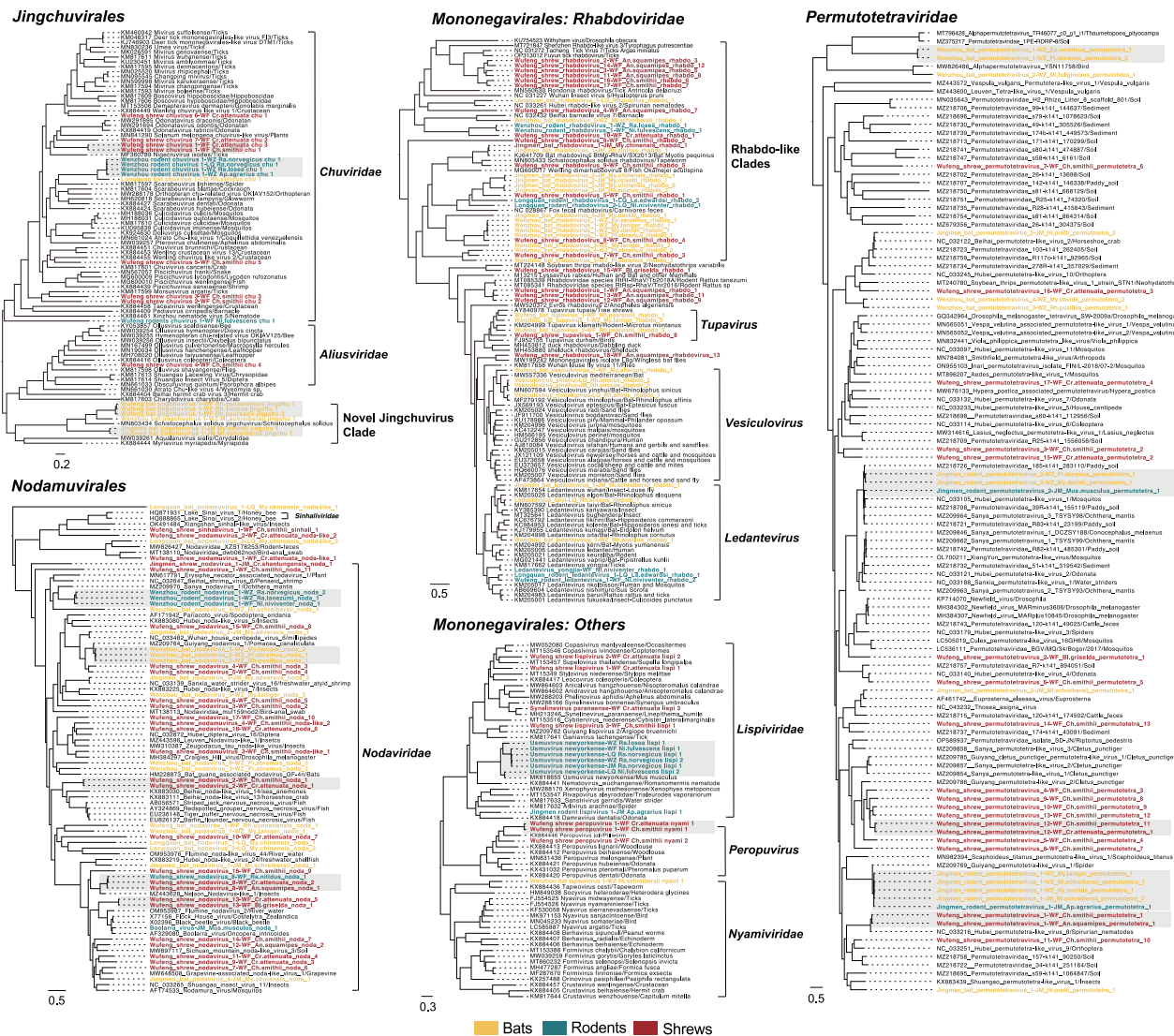


Figure S6. Phylogenetic trees of invertebrate-associated viral clades, related to Figure 4
Phylogeny of each viral clade based on the RdRp protein. Figure legend follows that of Figure S4.

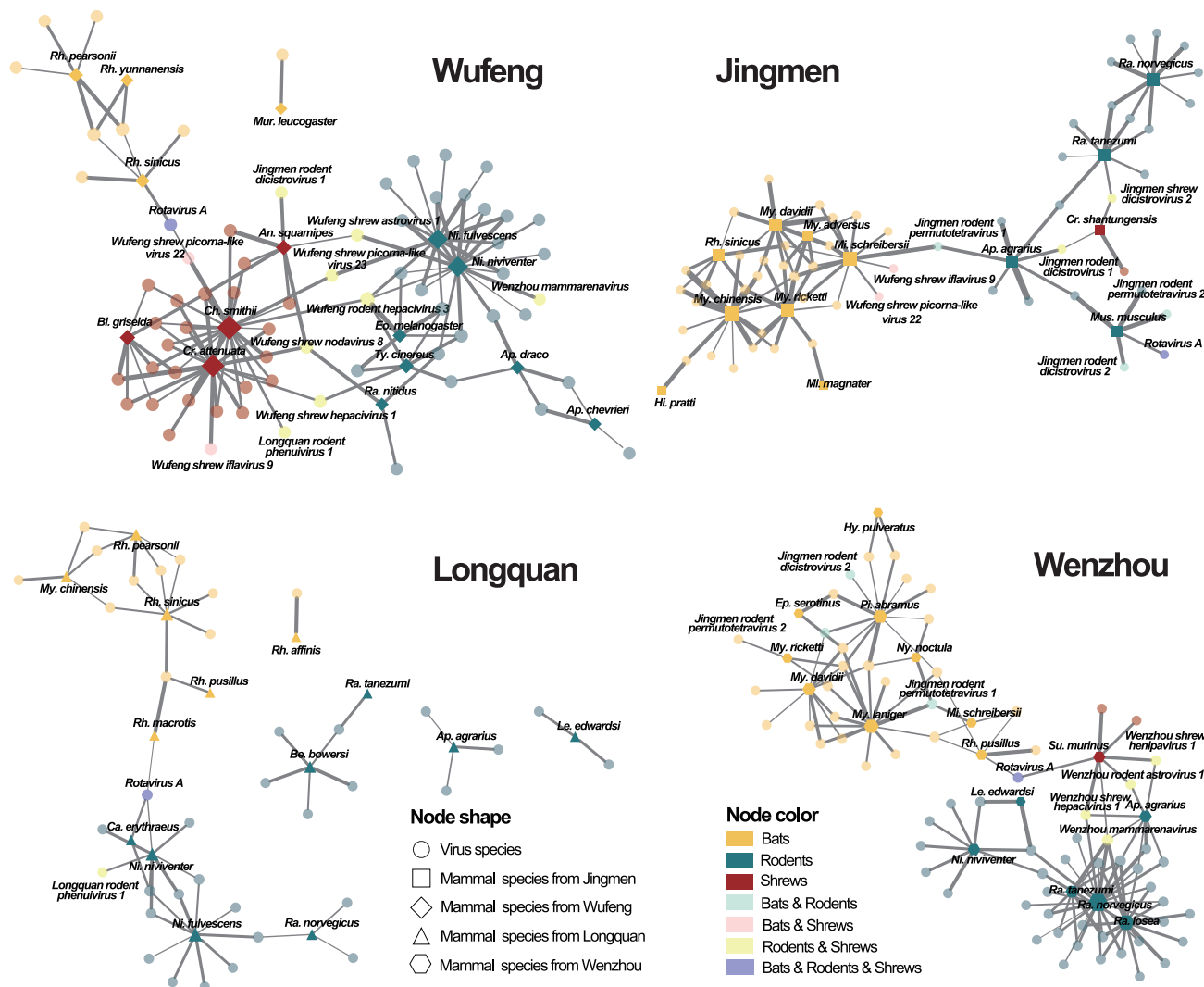


Figure S7. Host-virus correlation network for each sampling location, related to Figure 5

Node shapes stand for mammal species (non-transparent) and virus species (transparent). Node colors represent the host/mammal orders of viruses and mammals. Node and edge sizes are proportional to node degree and virus abundance, respectively.