

Luis, et al. (2015) Network Analysis of Host-Virus Communities in Bats and Rodents Reveals Determinants of Cross-Species Transmission

Supplementary Methods

Network Formulation

We removed rabies from the bat network because it was an extreme outlier, with 80 host species (and evidence of species-specific variants); thus, every species with rabies would on average have a degree of 79 more than species without rabies. This then gives a bi-modal degree distribution and gives too much weight to one virus species in determining the network structure, warranting its removal. We also removed taxonomic order-virus associations that were only identified by antibodies, because of potential cross-reactivity among closely related viruses.

In the database in (Luis *et al.*, 2013), we used what the International Committee on Taxonomy of Viruses (ICTV) defined as a virus species (King *et al.*, 2012). For example if a host species had two different strains of the same virus species, we only considered this a single virus. However, the hantaviruses in rodents are quite diverse and the ICTV's criteria of at least 7% amino acid sequence difference and/or fourfold difference in neutralization assays may not reflect true host-virus pairs and viral sharing. For example, ICTV recognizes 20 different strains of Andes hantavirus. Although they are considered one species, there are clearly multiple non-overlapping stable host-virus pairs, such as Pergamino virus that infects *Akodon azarae*, and Bermejo virus that infects *Oligoryzomys chacoensis*. Neither of these two strains infects the host of the other strain. Therefore, to better reflect true viral sharing, we considered them separate viruses. However, to test the sensitivity of this classification we also calculated network

metrics using the ICTV classification. We also performed an additional secondary analysis; to address the possibility that recent increased attention on bats may have led to increased sampling in bats and artificially increased the apparent connectivity of the bat network, we also created networks using only viral accounts before the year 2000.

Classification within the *Miniopterus* genus is very dynamic and several species appear to be complexes, with cryptic species. *Miniopterus schreibersii* has recently been reclassified into three species following Tian *et al.* (2004) and Appleton *et al.* (2004), thus we split *M. schreibersii* sensu lato into *M. schreibersii* (Europe/N. Africa/Near East), *M. fuliginosus* (Asia), and *M. oceanensis* (Australasia) (Tian *et al.*, 2004; Appleton *et al.*, 2004). Similarly, *Miniopterus schreibersii* is now not considered to occur in Eastern and Southern Africa, and because of frequent misidentification between *Miniopterus natalensis* and *Miniopterus schreibersii* (IUCN, 2010), we reclassified references to *M. schreibersii* in Eastern and Southern Africa as *M. natalensis*.

To account for sampling bias, we additionally calculated the quantitative linkage densities and connectance for sampling effort-corrected networks. To do this, we regressed the weight of each edge against the logged number of citations of the least sampled host species for each edge (e.g., Gómez *et al.*, 2013). We additively rescaled the residuals so that the lowest value was 1, and used them for the edge weights in the new networks. These new weights would then reflect the amount of viral sharing between species relative to sampling effort, under the assumption that the measure of sampling effort should be from the lesser studied species. These sampling effort-corrected networks were used only for calculating an adjusted mean weighted degree and adjusted weighted connectance.

Species traits were compiled from online databases and the literature as described in Luis *et al.* (2013). We examined body mass, number of litters per year, litter size, maximum longevity, torpor use, International Union for Conservation of Nature (IUCN) conservation status, geographic distribution

area, latitude of the midpoint (centroid) of the species distribution, number of other species in the same taxonomic order that are sympatric, and for bats only, migratory classification, diet, gregariousness, and propensity to roost in caves. Torpor expression was treated as a categorical variable with 3 categories: (1) no evidence of torpor use, (2) some torpor use, but not true hibernation (minimum body temperature $\geq 11^{\circ}\text{C}$), and (3) true hibernation (body temperature $< 11^{\circ}\text{C}$). We categorized migratory status of bats as (1) species that can be broadly categorized as sedentary or only local (< about 100km) migrants, (2) species that can be generalized as regional migrants (about 100-500 km), and (3) species that can be categorized as long-distance migrants (>500 km). This information was compiled from the literature, with species reported to be present in an area year-round considered evidence for category 1. See supplemental Figures S4-S7 in Luis *et al.* (2013) for plots of the raw data. Bat species traits that were new to these analyses were gregariousness, propensity to roost in caves and diet. Gregariousness was treated as a continuous variable after placing species in bins according to order of magnitude of typical colony sizes, for example (1) <10 , (2) 10-100, up to (6) $>100,000$. For propensity to roost in caves, bats were classified into 2 broad categories, (1) mostly using caves or (2) mostly using other roosts. All of these trait data were available for 53 species of bat and 74 species of rodent, and thus, the analyses using trait data were on these subsets of species. Diet was classified as mostly 1) frugivorous, 2) nectarivorous, 3) insectivorous, or 4) sanguivorous.

Network Statistics

To calculate p-values for the network statistics in Table 1, we performed 10,000 permutations of the networks, where we shuffled the edge weights. To calculate p-values for number of links, mean degree, mean weighted degree, connectance, and quantitative connectance, we took the edge lists for both the bat and rodent networks, which consist of each possible pair-wise combination of bat species, and each pair-wise combination of rodent species and the observed number of viruses shared between the two species (edge weights). We then randomly shuffled the edge weights (including zero-weight edges) across both bats and

rodents and reformed the bat and rodent networks from these shuffled weights. We calculated the metrics for both the bat and rodent networks for each permutation. The p-value listed in Table 1 was the number of times we observed a result that was at least as extreme as the observed differences between the bat and rodent networks, divided by 10,000 (the number of permutations). Since the permutations randomly shuffle the connections, the permuted networks would lose their assortativity and transitivity; therefore, the p-values for those can only tell us if the observed networks have significantly more assortativity or transitivity than a random network (not whether the bat and rodent networks are significantly different). Therefore for these metrics we permuted the edge weights within bats and within rodents, and there are two p-values listed, one each for the bat and rodent networks, respectively.

Community Detection

Host species were grouped into ‘communities’, partitions of highly interconnected nodes with fewer connections to nodes in other communities, by using the community detection algorithm described in Blondel *et al.* (2008), which maximizes the modularity between groups of the weighted networks. Communities are formed by maximizing the modularity between groups, defined as $Q = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j)$, where A_{ij} is the weight of the edge between i and j , $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to i , c_i is the community to which i is assigned, the delta function is 1 if $u=v$ and 0 otherwise, and $m = \frac{1}{2} \sum_{ij} A_{ij}$. Self edges (the number of viruses each species had) were ignored, and several iterations with different initiations were performed to maximize the modularity score.

As an alternative to using modularity to assign nodes to communities, we also used clique percolation theory (Palla *et al.*, 2005). In this method, nodes can belong to multiple communities. Here, communities are unions of smaller fully connected subgraphs that share nodes, called k-cliques, where k refers to the number of nodes in the subgraphs. A k-clique-community is the union of all k-cliques that can be reached through a series of adjacent k-cliques (those that share k-1 nodes) (Palla *et al.*, 2005). We chose a k

of 5 because it resulted in a community structure as highly structured as possible (see Palla *et al.*, 2005, supplementary information). The downside of this method for these networks is that each virus that has at least 5 hosts will make a fully connected subgraph of its hosts. This led to many communities with a great amount of overlap between them. For example, bat communities 7 and 8 shared 4 of 6 species— all 4 of the shared species were in the *Rhinolophus* genus and belonged to the same community using the modularity method. The two additional members in each community connected them to a separate community by the modularity method. Therefore we concentrate on the communities identified by maximizing the modularity, but outline (in black in figures 3 and 4) the species which belonged to multiple communities in the k-clique communities to illustrate the amount of overlap that occurs between the communities.

MCMCglmm

In addition to the phylogenetic generalized least squares (PGLS) analyses, we used a Markov chain Monte Carlo approach to fit Bayesian generalized linear mixed models using the 'MCMCglmm' package in R (Hadfield *et al.*, 2010) to determine host traits correlated to a species' viral richness and its degree and betweenness in the network. This framework allows phylogenetic effects to be included in the model as a random variable, with the correlation in phylogenetic effects between two host species being inversely proportional to the time since those two host species shared a common ancestor (following a Brownian model of evolution). Models were run for 13,000 iterations with a burn-in of 3,000 iterations, a thinning interval of 10, and flat priors.

Supplementary Results

As an alternative to phylogenetic least squares (PGLS), we also ran models using MCMC generalized linear mixed effects models (Hadfield *et al.*, 2010). Using this method, we can include the phylogeny as

a random effect. The results were largely similar to the PGLS results (See Tables S5, S8, S13, S16, & S19). The most significant difference was the best predictors for the number of viruses in bats (Table S5). The top ranked model included only citations and phylogeny. The next 3 models did almost as well, and they included diet and gregariousness, in accordance with the PGLS model rankings. For degree in bats, the best model using MCMCglmm included gregariousness and phylogeny (Table S8), but not diet and sympatry as in the PGLS analyses; however our variable importance plots show these variables are not as important as gregariousness (Figure 2b). We were unable to get the models for betweenness in bats to run. The function had a hard time estimating some of the parameters, and model runs were very dependent on initial conditions. We lacked prior information to use stronger priors. For the number of viruses, degree, and betweenness in rodents, the best models by PGLS and MCMCglmm contained the same covariates (Table S13, 16, S19). For degree and betweenness, the glmm models with phylogeny as a random effect had approximately the same DIC as models without phylogeny, which matches the estimated lambda of 0 for the PGLS models.

References

- Appleton, B., McKenzie, J. & Christidis, L. (2004). Molecular systematics and biogeography of the bent-wing bat complex *Miniopterus schreibersii* (Kuhl, 1917) (Chiroptera: Vespertilionidae). *Molecular Phylogenetics and Evolution*, 31, 431–439.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, P10008.
- Gómez, J.M., Nunn, C.L. & Verdú, M. (2013). Centrality in primate–parasite networks reveals the potential

for the transmission of emerging infectious diseases to humans. *Proceedings of the National Academy of Sciences*, 110, 7738–7741.

Hadfield, J.D. *et al.* (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33, 1–22.

IUCN (2010). IUCN red list of threatened species. version 2010.4. <http://www.iucnredlist.org>.

King, A., Adams, M., Carstens, E. & Lefkowitz, E. (eds.) (2012). *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. San Diego: Elsevier.

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., Fooks, A.R., Rupprecht, C.E., Wood, J.L.N. & Webb, C.T. (2013). A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proceedings of the Royal Society B: Biological Sciences*, 280, 20122753.

Palla, G., Derényi, I., Farkas, I. & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435, 814–818.

Tian, L., Liang, B., Maeda, K., Metzner, W. & Zhang, S. (2004). Molecular studies on the classification of *Miniopterus schreibersii* (Chiroptera: Vespertilionidae) inferred from mitochondrial cytochrome b sequences. *Folia Zoologica*, 53, 303–311.

Supplementary Tables

Table S1. Loading values for principal components summarizing life history traits.

	PC1	PC2	PC3	PC4
Bats:				
litter size	0.47	-0.51	0.60	-0.41
litters per year	-0.70	-0.12	-0.03	-0.70
maximum longevity	0.51	0.51	-0.37	-0.58
body mass	-0.16	0.68	0.71	0.01
proportion of variance	0.37	0.32	0.18	0.13
Rodents:				
litter size	-0.26	0.87	-0.35	-0.23
litters per year	-0.47	-0.47	-0.74	-0.13
maximum longevity	0.61	-0.10	-0.19	-0.76
body mass	0.58	0.11	-0.55	0.59
proportion of variance	0.53	0.26	0.14	0.07

Table S2. Multiple regression on matrices of a subset of data, where the response variable is the number of viruses shared between two species.

Model	R ²	p
Bats		
~litter size+phylogeny+sympatry+citations	0.265	0.0001
~diet+phylogeny+sympatry+citations	0.260	0.0001
~longevity+phylogeny+sympatry+citations	0.259	0.0001
~migration+phylogeny+sympatry+citations	0.258	0.0001
~latitude+phylogeny+sympatry+citations	0.258	0.0001
~litters per year+phylogeny+sympatry+citations	0.255	0.0001
~cave roosting+phylogeny+sympatry+citations	0.255	0.0001
~mass+phylogeny+sympatry+citations	0.254	0.0001
~torpor+phylogeny+sympatry+citations	0.253	0.0001
~gregariousness+phylogeny+sympatry+citations	0.253	0.0001
~phylogeny+sympatry+citations	0.253	0.0001
~phylogeny+sympatry	0.234	0.0001
~sympatry	0.220	0.0001
~phylogeny	0.069	0.0001
~citations	0.045	0.0002
Rodents		
~longevity+phylogeny+sympatry+citations	0.174	0.0001
~torpor+phylogeny+sympatry+citations	0.172	0.0001
~litters per year+phylogeny+sympatry+citations	0.171	0.0001
~mass+phylogeny+sympatry+citations	0.171	0.0001
~latitude+phylogeny+sympatry+citations	0.171	0.0001
~litter size+phylogeny+sympatry+citations	0.170	0.0001
~phylogeny+sympatry+citations	0.170	0.0001
~citations	0.097	0.0001
~sympatry	0.091	0.0001
~phylogeny	0.017	0.0002
Bats and Rodents combined		
~litter size+phylogeny+sympatry+citations	0.153	0.0001
~litters per year+phylogeny+sympatry+citations	0.153	0.0001
~torpor+phylogeny+sympatry+citations	0.152	0.0001
~mass+phylogeny+sympatry+citations	0.152	0.0001
~latitude+phylogeny+sympatry+citations	0.151	0.0001
~longevity+phylogeny+sympatry+citations	0.151	0.0001
~phylogeny+sympatry+citations	0.151	0.0001
~sympatry+citations	0.115	0.0001
~phylogeny+citations	0.098	0.0001
~sympatry	0.083	0.0001
~phylogeny	0.054	0.0001
~citations	0.051	0.0001

Table S3. Generalized least squares model rankings for the number of viruses identified in bats. (See Fig. 2a for variable rankings of the best model.)

model	AICc	npar	weight	p	R	λ
~ cit. + diet	241.05	6	0.280	5.65×10^{-6}	0.66	0.00
~ cit. + gregariousness + diet	241.13	7	0.268	5.25×10^{-6}	0.68	0.00
~ cit. + PC1 + gregariousness + diet	242.07	8	0.168	6.38×10^{-6}	0.70	0.00
~ cit. + PC1 + diet	242.43	7	0.140	9.47×10^{-6}	0.67	0.00
~ cit. + PC1	245.61	4	0.029	4.23×10^{-5}	0.54	0.20
~ cit.	246.20	3	0.021	3.35×10^{-5}	0.44	0.36
~ cit. + PC1 + gregariousness	246.28	5	0.020	6.79×10^{-5}	0.57	0.19
~ cit. + gregariousness	247.19	4	0.013	9.32×10^{-5}	0.46	0.41
~ cit. + sympatry	247.79	4	0.010	1.26×10^{-4}	0.48	0.30
~ cit. + sympatry + PC1	248.04	5	0.009	1.58×10^{-4}	0.54	0.19
~ cit. + area	248.31	4	0.007	1.64×10^{-4}	0.43	0.38
~ cit. + latitude	248.50	4	0.007	1.79×10^{-4}	0.46	0.32
~ cit. + sympatry + gregariousness	248.73	5	0.006	2.20×10^{-4}	0.51	0.35
~ cit. + sympatry + PC2	249.11	5	0.005	2.63×10^{-4}	0.56	0.00
~ cit. + PC1 + migration	249.25	6	0.005	2.53×10^{-4}	0.57	0.17
~ cit. + gregariousness + PC2	249.63	5	0.004	3.36×10^{-4}	0.46	0.41
~ cit. + migration	250.18	5	0.003	4.37×10^{-4}	0.46	0.36
~ cit. + sympatry + gregariousness + PC2	251.10	6	0.002	5.89×10^{-4}	0.57	0.00
~ cit. + gregariousness + migration	251.53	6	0.001	7.15×10^{-4}	0.47	0.41
~ cit. + sympatry + migration	251.71	6	0.001	7.73×10^{-4}	0.51	0.28
~ sympatry + gregariousness	258.27	4	0.000	2.38×10^{-2}	0.39	0.00
~ sympatry	258.66	3	0.000	2.94×10^{-2}	0.32	0.00
~ 1	261.16	2	0.000			0.16

cit. indicates the logged number of citations on Web of Science. PC's are principal components on life history traits (see Table S1 & Fig. S6). λ is the strength of phylogenetic dependence (see Methods). ~ 1 indicates the null (intercept) model.

Table S4. Coefficients for the best generalized least squares model for the number of viruses identified in bats (number of viruses $\sim \log(\text{citations}) + \text{diet}$).

	Value	Std.Error	t-value	p-value
(Intercept)	-3.072869	2.9507236	-1.041395	0.3029
citations	1.367932	0.3376619	4.051188	0.0002
diet-fruit	4.360703	2.5798984	1.690262	0.0975
diet-insects	2.855044	2.5916522	1.101631	0.2761
diet-nectar	4.806748	2.9921075	1.606476	0.1147

Table S5. MCMCglmm model rankings for number of viruses in bats.

model	DIC
fixed= ~citations, random=~phylogeny	239.03
fixed= ~citations+gregariousness+diet, random=~phylogeny	239.07
fixed= ~citations+diet	239.41
fixed= ~citations+diet, random=~phylogeny	239.52
fixed= ~citations+gregariousness, random=~phylogeny	239.77
fixed= ~citations+gregariousness+sympatry+diet, random=~phylogeny	240.03
fixed= ~citations+migration, random=~phylogeny	240.42
fixed= ~citations+PC1+gregariousness, random=~phylogeny	240.82
fixed= ~citations+latitude, random=~phylogeny	240.89
fixed= ~citations+sympatry+migration, random=~phylogeny	240.93
fixed= ~citations+gregariousness+sympatry, random=~phylogeny	242.89
fixed= ~citations+sympatry, random=~phylogeny	243.85
fixed= ~citations+sympatry+PC2, random=~phylogeny	248.01
fixed= ~citations	252.75
fixed= ~gregariousness+sympatry, random=~phylogeny	257.40
fixed= ~1, random=~phylogeny	261.38
fixed= ~1	261.98

Table S6. Generalized least squares model rankings for degree in the bat viral sharing network. (See Fig. 2b for variable rankings of the best model.)

model	AICc	npar	weight	p	R	λ
\sim cit. + gregariousness + sympatry + diet	416.38	8	0.140	0.0006	0.65	0.00
\sim cit. + gregariousness + sympatry	416.54	5	0.129	0.0015	0.53	0.28
\sim cit. + gregariousness	417.42	4	0.084	0.0024	0.36	0.44
\sim cit. + sympatry + PC2	417.44	5	0.083	0.0023	0.57	0.00
\sim cit. + PC1 + gregariousness	417.68	5	0.073	0.0026	0.47	0.41
\sim cit. + sympatry	417.89	4	0.066	0.0030	0.45	0.27
\sim cit.	418.20	3	0.056	0.0028	0.28	0.40
\sim sympatry + gregariousness	418.87	4	0.040	0.0049	0.52	0.00
\sim cit. + PC1	418.94	4	0.039	0.0051	0.39	0.38
\sim cit. + sympatry + migration	418.95	6	0.039	0.0037	0.54	0.25
\sim cit. + gregariousness + cave roosting	419.38	5	0.030	0.0057	0.33	0.49
\sim cit. + latitude	419.48	4	0.029	0.0067	0.51	0.00
\sim cit. + sympatry + diet	419.67	7	0.026	0.0037	0.60	0.00
\sim cit. + gregariousness + PC2	419.75	5	0.025	0.0068	0.33	0.47
\sim cit. + PC3 + gregariousness	419.79	5	0.025	0.0069	0.37	0.44
\sim cit. + diet	419.91	6	0.023	0.0056	0.57	0.00
\sim cit. + sympatry + PC1	420.06	5	0.021	0.0078	0.46	0.27
\sim cit. + migration	420.18	5	0.020	0.0082	0.36	0.43
\sim cit. + area	420.28	4	0.020	0.0100	0.28	0.41
\sim cit. + PC1 + migration	420.43	6	0.018	0.0071	0.47	0.43
\sim cit. + gregariousness + migration	420.46	6	0.018	0.0072	0.41	0.46
\sim sympatry	422.12	3	0.008	0.0247	0.43	0.06
\sim cit. + PC1 + diet	422.16	7	0.008	0.0104	0.57	0.00
\sim 1	424.91	2	0.002			0.36

cit. indicates the logged number of citations on Web of Science. PC's are principal components on life history traits (see Table S1 & Fig. S6). λ is the strength of phylogenetic dependence (see Methods).

Table S7. Coefficients for the best generalized least squares model for degree in the bat network.

	Value	Std.Error	t-value	p-value
(Intercept)	-3.213794	5.558061	-0.5782221	0.5658
citations	2.854609	1.662416	1.7171448	0.0923
gregariousness	2.226032	1.238327	1.7976122	0.0784
sympatry	0.055026	0.017916	3.0712856	0.0035

Table S8. MCMCglmm model rankings for degree in the bat viral sharing network.

model	DIC
fixed= ~ citations+gregariousness, random=~ animal	405.03
fixed= ~ citations+PC1+gregariousness, random=~ animal	405.69
fixed= ~ citations+migration, random=~ animal	406.41
fixed= ~ citations, random=~ animal	408.24
fixed= ~ citations+gregariousness+sympatry, random=~ animal	408.74
fixed= ~ citations+sympatry+migration, random=~ animal	412.55
fixed= ~ 1	413.02
fixed= ~ citations+gregariousness+sympatry+diet, random=~ animal	413.45
fixed= ~ citations+Latitude.abs, random=~ animal	414.14
fixed= ~ citations+sympatry+PC2, random=~ animal	414.84
fixed= ~ citations+gregariousness+diet, random=~ animal	415.77
fixed= ~ citations+sympatry, random=~ animal	416.11
fixed= ~ gregariousness+sympatry, random=~ animal	418.16
fixed= ~ citations+diet, random=~ animal	418.39
fixed= ~ citations+gregariousness	427.25
fixed= ~ citations	428.27
fixed= ~ 1	430.72

Table S9. Generalized least squares model rankings for betweenness in the bat viral sharing network. (See Fig. 2c for variable rankings of the best model.)

model	AICc	npar	weight	R	λ
~ cit. + migration	756.46	5	0.189	0.45	0.79
~ cit. + sympatry + migration	756.87	6	0.154	0.46	0.93
~ cit. + gregariousness + migration	757.46	6	0.114	0.48	0.78
~ cit. + gregariousness	757.66	4	0.104	0.36	0.82
~ cit. + PC1 + migration	757.79	6	0.097	0.49	0.79
~ cit.	758.64	3	0.064	0.23	0.93
~ cit. + sympatry + gregariousness	759.13	5	0.050	0.37	0.88
~ cit. + sympatry	759.54	4	0.041	0.24	0.98
~ cit. + PC1 + gregariousness	759.75	5	0.036	0.39	0.79
~ cit. + gregariousness + PC2	759.96	5	0.033	0.33	0.80
~ cit. + PC1	760.59	4	0.024	0.26	0.93
~ cit. + latitude	760.84	4	0.021	0.27	0.92
~ cit. + sympatry + PC2	761.57	5	0.015	0.31	1.00
~ cit. + sympatry + gregariousness + PC2	761.67	6	0.014	0.38	0.89
~ cit. + sympatry + PC1	761.96	5	0.012	0.24	0.98
~ sympatry + gregariousness	764.98	4	0.003	0.36	0.85
~ cit. + diet	765.49	6	0.002	0.39	0.92
~ cit. + sympatry + diet	766.55	7	0.001	0.36	0.98
~ 1	767.74	2	0.001		0.15
~ cit. + PC1 + diet	767.78	7	0.001	0.38	0.92
~ sympatry	768.21	3	0.001	0.11	0.98

cit. indicates the logged number of citations on Web of Science. PC's are principal components on life history traits (see Table S1 & Fig. S6). λ is the strength of phylogenetic dependence (see Methods).

Table S10. Coefficients for the best generalized least squares model for betweenness in bats.

	Value	Std.Error	t-value	p-value
(Intercept)	-137.55805	220.10945	-0.6249529	0.5349
citations	113.52532	39.70538	2.8591924	0.0062
migration-regional	207.90011	104.41604	1.9910744	0.0521
migration-long distance	-82.96197	122.93303	-0.6748550	0.5029

Table S11. Generalized least squares model rankings for number of viruses identified in rodents. (See Fig. 2d for variable rankings of the best model.)

model	AICc	npar	weight	p	R	λ
~ cit. + sympatry + PC3	369.09	5	0.350	1.335×10^{-10}	0.69	0.26
~ cit. + sympatry	371.09	4	0.129	2.017×10^{-10}	0.67	0.12
~ cit. + sympatry + IUCN + PC3	371.45	6	0.108	5.931×10^{-10}	0.69	0.26
~ cit. + PC3	371.81	4	0.090	2.898×10^{-10}	0.65	0.28
~ cit. + sympatry + area	372.21	5	0.074	6.153×10^{-10}	0.68	0.18
~ cit. + sympatry + torpor	372.35	6	0.069	9.135×10^{-10}	0.69	0.00
~ cit. + sympatry + latitude	373.06	5	0.048	9.331×10^{-10}	0.68	0.00
~ cit. + sympatry + IUCN	373.35	5	0.042	1.074×10^{-9}	0.67	0.13
~ cit. + sympatry + PC2	373.57	5	0.037	1.197×10^{-9}	0.68	0.00
~ cit.	375.08	3	0.017	5.711×10^{-10}	0.63	0.20
~ cit. + torpor	375.75	5	0.013	3.478×10^{-9}	0.66	0.00
~ cit. + area + latitude	376.09	5	0.011	4.103×10^{-9}	0.66	0.16
~ cit. + area	376.88	4	0.007	3.653×10^{-9}	0.63	0.24
~ cit. + IUCN	377.28	4	0.006	4.453×10^{-9}	0.63	0.20
~ sympatry	405.36	3	0.000	4.316×10^{-3}	0.33	0.02
~ 1	411.33	2	0.000			0.07

cit. indicates the logged number of citations on Web of Science. PC's are principal components on life history traits (see Table S1 & Fig. S6). λ is the strength of phylogenetic dependence (see Methods).

Table S12. Coefficients for the best generalized least squares model for the number of viruses in rodents.

	Value	Std.Error	t-value	p-value
(Intercept)	-5.570284	1.458	-3.821	0.0003
citations	1.549927	0.265	5.851	0.0000
sympatry	0.005471	0.002	2.222	0.0295
PC3	-1.044548	0.488	-2.141	0.0357

Table S13. MCMCglmm model rankings for number of viruses in rodents.

model	DIC
fixed= ~citations+sympatry+PC3, random=~phylogeny	364.28
fixed= ~citations+PC3, random=~phylogeny	367.03
fixed= ~citations+sympatry, random=~phylogeny	368.65
fixed= ~citations+sympatry+PC3	370.19
fixed= ~citations+sympatry+area, random=~phylogeny	370.20
fixed= ~citations+sympatry+torpor, random=~phylogeny	370.69
fixed= ~citations+sympatry+Latitude, random=~phylogeny	372.35
fixed= ~citations, random=~phylogeny	374.16
fixed= ~citations	376.40
fixed= ~1, random=~phylogeny	411.39
fixed= ~1	411.66

Table S14. Generalized least squares model rankings for degree in the rodent viral sharing network. (See Fig. 2e for variable rankings of the best model.)

model	AICc	npar	weight	p	R	λ
~ cit. + sympatry + latitude	571.94	5	0.411	7.00×10^{-11}	0.70	0.00
~ cit. + sympatry	573.57	4	0.182	8.68×10^{-11}	0.68	0.00
~ cit. + sympatry + IUCN	574.52	5	0.113	2.48×10^{-10}	0.69	0.00
~ cit. + sympatry + area	574.56	5	0.111	2.52×10^{-10}	0.69	0.00
~ cit. + sympatry + PC3	575.70	5	0.063	4.41×10^{-10}	0.68	0.00
~ cit. + sympatry + PC2	575.86	5	0.058	4.77×10^{-10}	0.68	0.00
~ cit. + sympatry + IUCN + PC3	576.85	6	0.035	1.07×10^{-9}	0.69	0.00
~ cit. + sympatry + torpor	577.36	6	0.027	1.37×10^{-9}	0.69	0.00
~ cit. + IUCN	590.05	4	0.000	3.28×10^{-7}	0.58	0.00
~ cit.	591.06	3	0.000	2.48×10^{-7}	0.55	0.00
~ cit. + PC3	593.22	4	0.000	1.60×10^{-6}	0.55	0.00
~ cit. + area	593.29	4	0.000	1.66×10^{-6}	0.55	0.00
~ cit. + torpor	593.64	5	0.000	2.76×10^{-6}	0.57	0.00
~ cit. + area + latitude	595.03	5	0.000	5.40×10^{-6}	0.55	0.00
~ sympatry	598.34	3	0.000	1.10×10^{-5}	0.48	0.00
~ 1	615.50	2	0.000			0.00

cit. indicates the logged number of citations on Web of Science. PC's are principal components on life history traits (see Table S1 & Fig. S6). λ is the strength of phylogenetic dependence (see Methods).

Table S15. Coefficients for the best generalized least squares model for degree in rodents.

	Value	Std.Error	t-value	p-value
(Intercept)	-20.555	5.348	-3.843	0.0003
citations	4.399	1.120	3.927	0.0002
sympatry	0.056	0.011	5.066	0.0000
latitude	0.214	0.110	1.956	0.0545

Table S16. MCMCglmm model rankings for degree in rodents.

model	DIC
fixed= ~citations+sympatry+Latitude	571.02
fixed= ~citations+sympatry+Latitude, random=~phylogeny	571.43
fixed= ~citations+sympatry, random=~phylogeny	573.02
fixed= ~citations+sympatry+area, random=~phylogeny	573.87
fixed= ~citations+sympatry+PC3, random=~phylogeny	574.96
fixed= ~citations+sympatry+torpor, random=~phylogeny	576.14
fixed= ~citations	590.68
fixed= ~citations, random=~phylogeny	590.87
fixed= ~citations+PC3, random=~phylogeny	592.80
fixed= ~1, random=~phylogeny	615.29
fixed= ~1	615.30

Table S17. Generalized least squares model rankings for betweenness in the rodent viral sharing network. (See Fig. 2f for variable rankings of the best model.)

model	AICc	npar	weight	p	R	λ
~ cit. + sympatry + area	1055.01	5	0.758	1.11×10^{-16}	0.81	0.00
~ cit. + sympatry	1059.20	4	0.093	2.22×10^{-16}	0.79	0.00
~ cit. + sympatry + latitude	1060.89	5	0.040	1.11×10^{-15}	0.79	0.00
~ cit. + sympatry + IUCN	1061.41	5	0.031	1.44×10^{-15}	0.79	0.00
~ cit. + sympatry + PC2	1061.48	5	0.030	1.44×10^{-15}	0.79	0.00
~ cit. + sympatry + PC3	1061.49	5	0.030	1.55×10^{-15}	0.79	0.00
~ cit. + sympatry + torpor	1063.75	6	0.010	7.66×10^{-15}	0.79	0.00
~ cit. + sympatry + IUCN + PC3	1063.78	6	0.009	7.77×10^{-15}	0.79	0.00
~ sympatry	1072.20	3	0.000	4.64×10^{-14}	0.73	0.00
~ cit. + area + latitude	1111.10	5	0.000	5.11×10^{-5}	0.51	0.00
~ cit.	1116.39	3	0.000	3.69×10^{-4}	0.40	0.06
~ cit. + PC3	1117.52	4	0.000	1.01×10^{-3}	0.41	0.09
~ cit. + IUCN	1117.59	4	0.000	1.05×10^{-3}	0.41	0.00
~ cit. + area	1118.62	4	0.000	1.76×10^{-3}	0.40	0.21
~ cit. + torpor	1119.08	5	0.000	2.27×10^{-3}	0.42	0.03
~ 1	1126.90	2	0.000			0.00

cit. indicates the logged number of citations on Web of Science. PC's are principal components on life history traits (see Table S1 & Fig. S6). λ is the strength of phylogenetic dependence (see Methods).

Table S18. Coefficients for the best generalized least squares model for degree in rodents.

	Value	Std.Error	t-value	p-value
(Intercept)	-770.680	130.421	-5.909	0.0000
citations	107.359	25.274	4.248	0.0001
sympatry	2.552	0.254	10.029	0.0000
area	0.000	0.000	-2.534	0.0135

Table S19. MCMCglmm model rankings for betweenness in rodents.

model	DIC
fixed= ~citations+sympatry+area, random=~phylogeny	1053.66
fixed= ~citations+sympatry+area	1054.21
fixed= ~citations+sympatry, random=~phylogeny	1058.78
fixed= ~citations+sympatry+Latitude, random=~phylogeny	1060.19
fixed= ~citations+sympatry+PC3, random=~phylogeny	1060.64
fixed= ~citations+sympatry+torpor, random=~phylogeny	1062.69
fixed= ~citations, random=~phylogeny	1115.70
fixed= ~citations	1116.05
fixed= ~citations+PC3, random=~phylogeny	1116.20
fixed= ~1	1126.67
fixed= ~1, random=~phylogeny	1126.72

Supplementary Figures

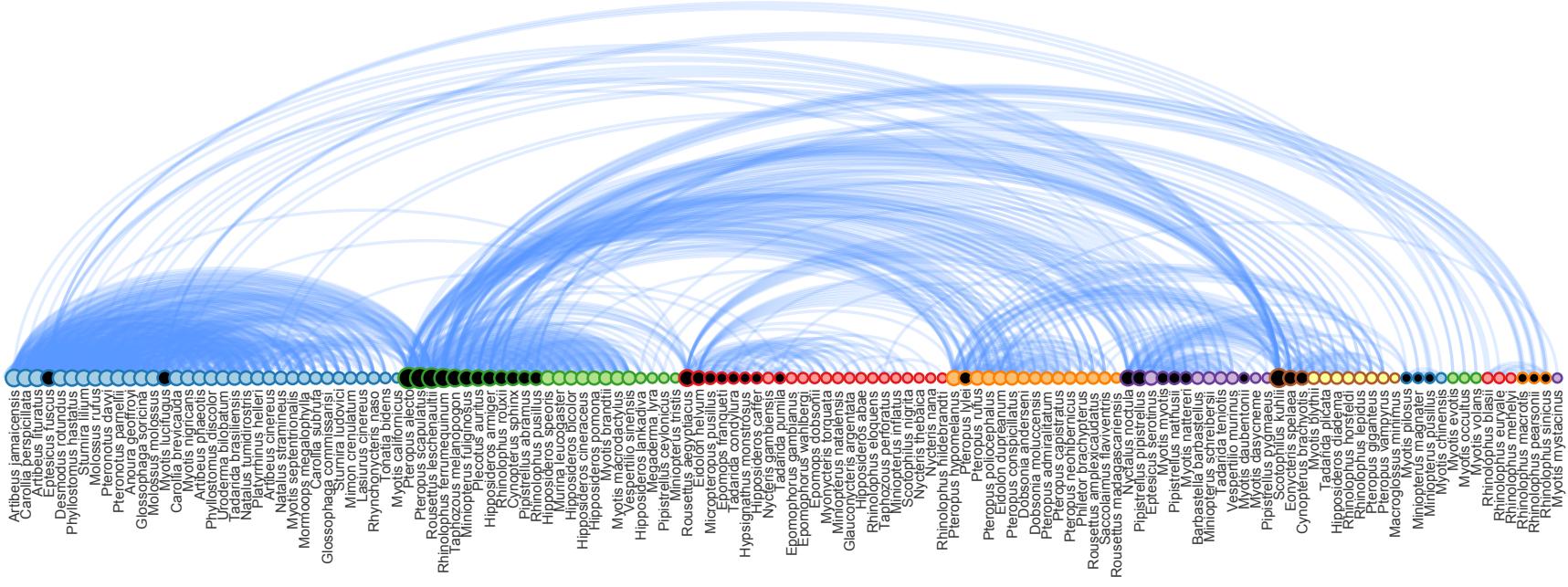


Fig. S1. Diagram of the viral sharing network in bats, organized by communities as determined by maximizing the modularity of the viral sharing network. Each circle represents a bat species, and links represent viruses shared between species, with the thickness of the line representing the number of viruses shared. The circle's size is proportional to the degree of the host species and is colored by community. Species symbols with a black center were classified as part of multiple communities in an alternative method for community identification (see text). The species in gray were not assigned to a community.

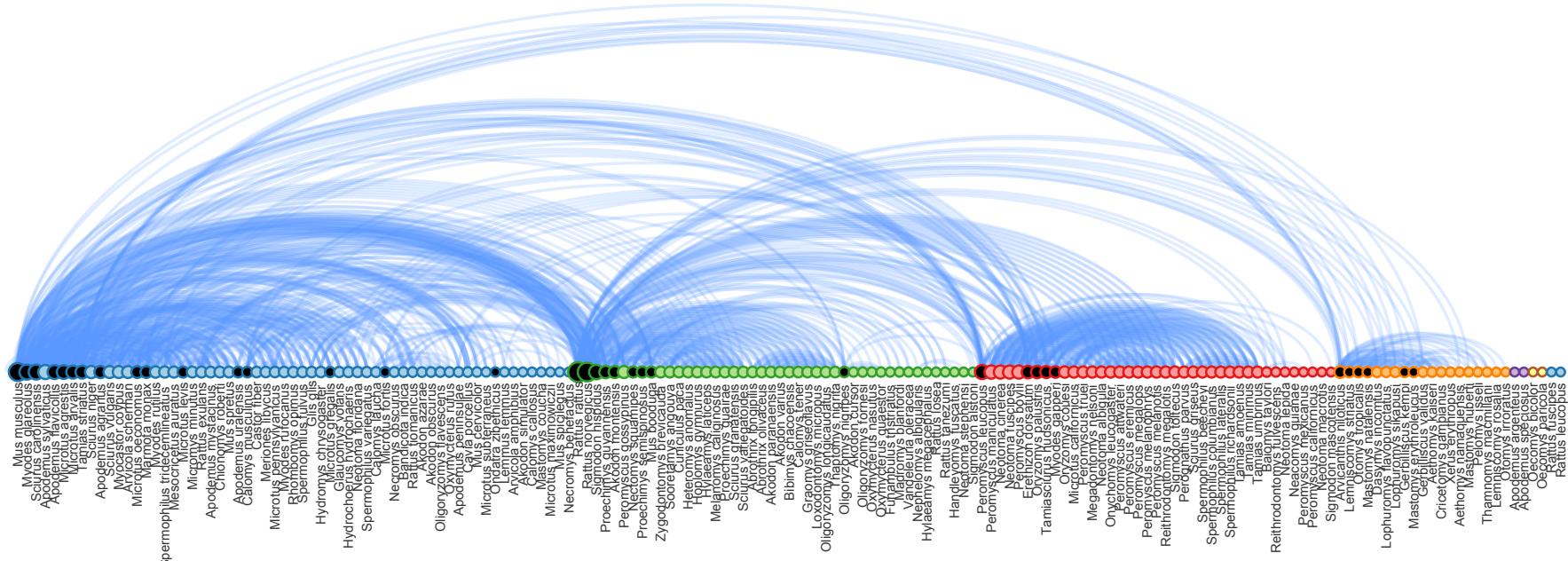
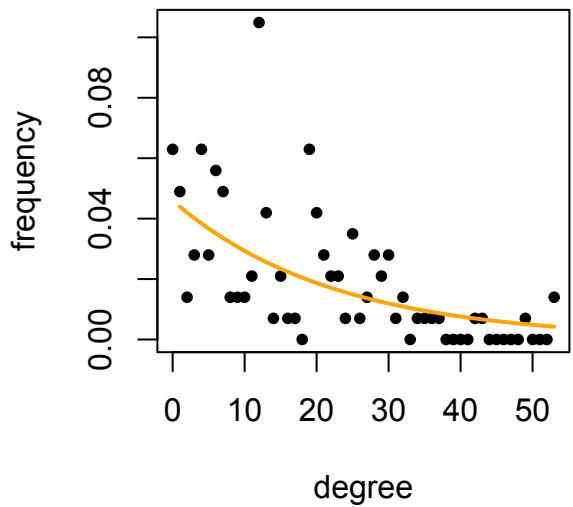


Fig. S2. Diagram of the viral sharing network in rodents, organized by communities as determined by maximizing the modularity of the viral sharing network. Each circle represents a rodent species, and links represent viruses shared between species, with the thickness of the line representing the number of viruses shared. The circle's size is proportional to the degree of the host species and is colored by community. Species symbols with a black center were classified as part of multiple communities in an alternative method for community identification (see text). The species in gray were not assigned to a community.

a)



b)

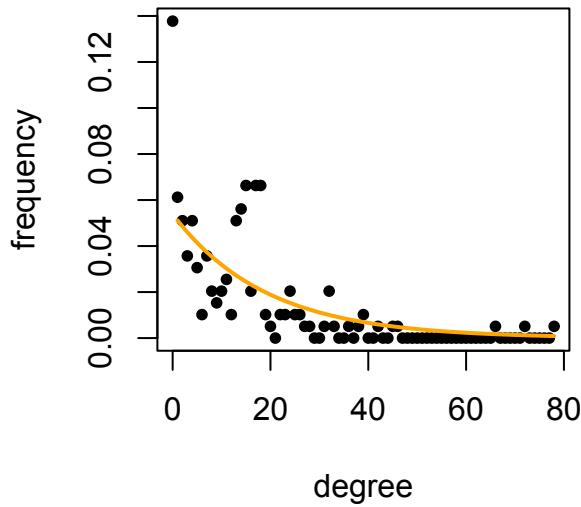


Fig. S3. The degree distributions for the a) bat and b) rodent viral sharing networks. The orange line is a maximum likelihood fit of an exponential distribution, with parameter 22.25 for bats and 19.05 for rodents.

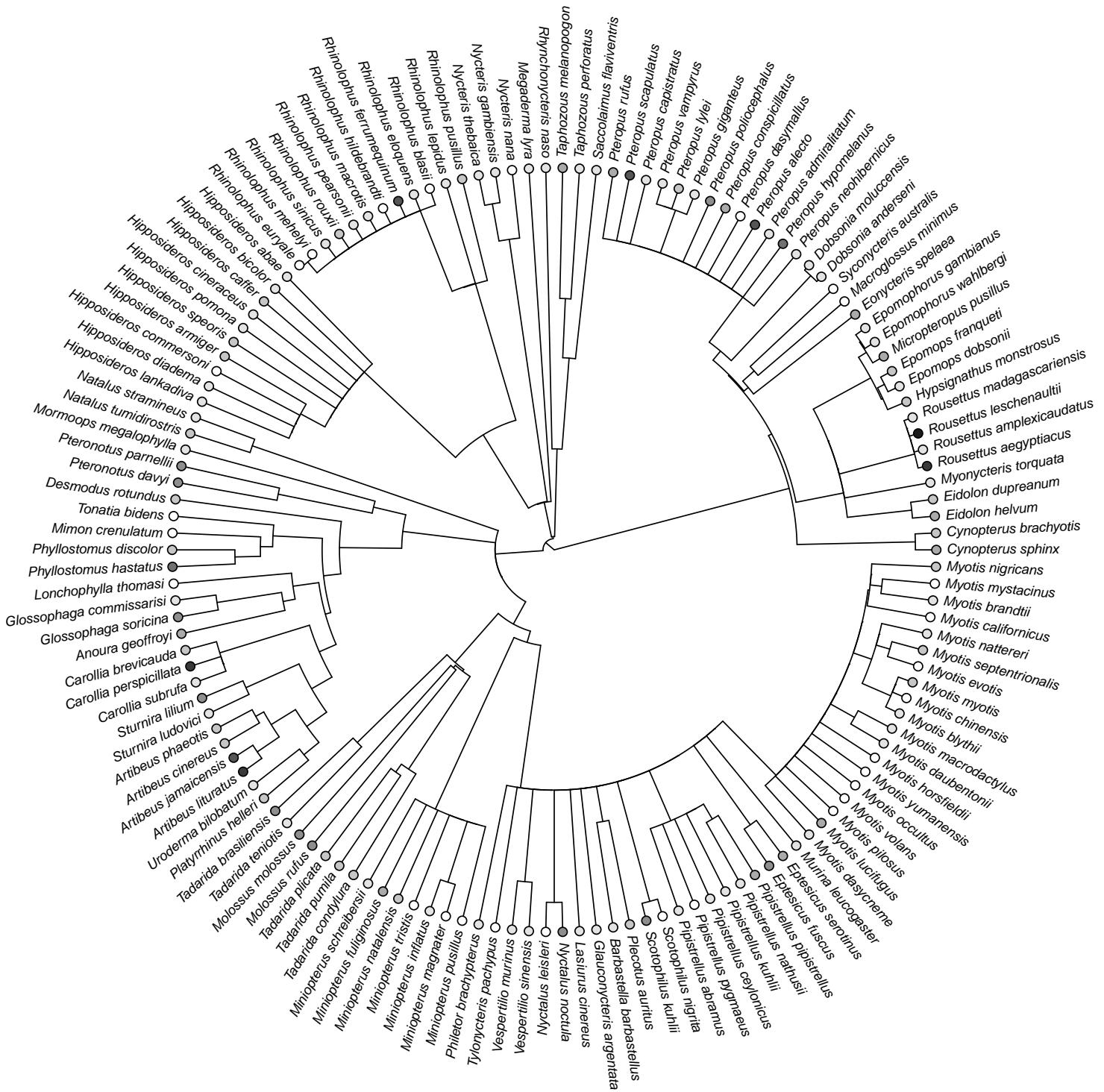


Fig. S4. Phylogenetic tree of the bat species in our network, with the darker shades of gray indicating more zoonotic viruses.

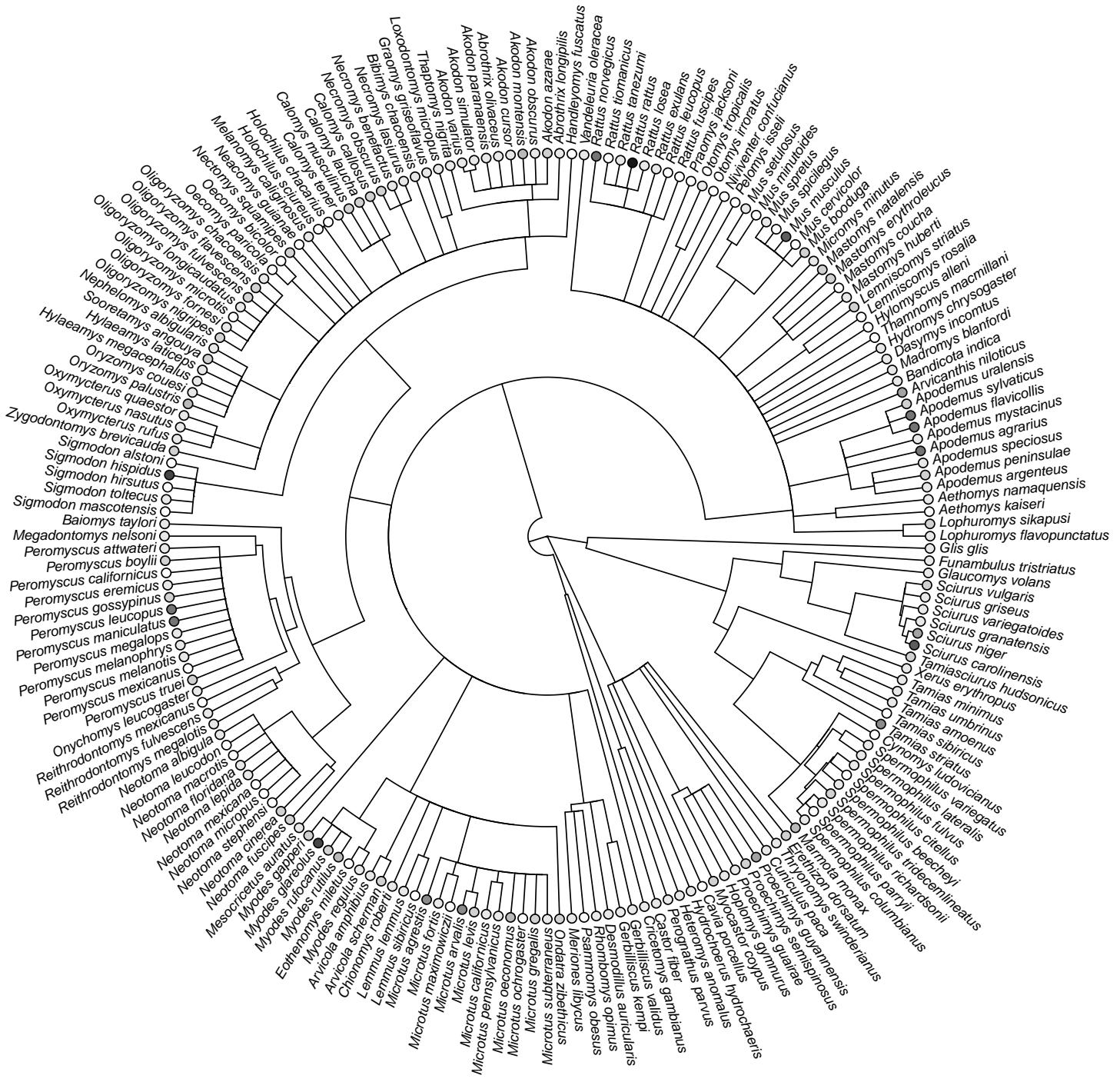


Fig. S5. Phylogenetic tree of the rodent species in our network, with the darker shades of gray indicating more zoonotic viruses.

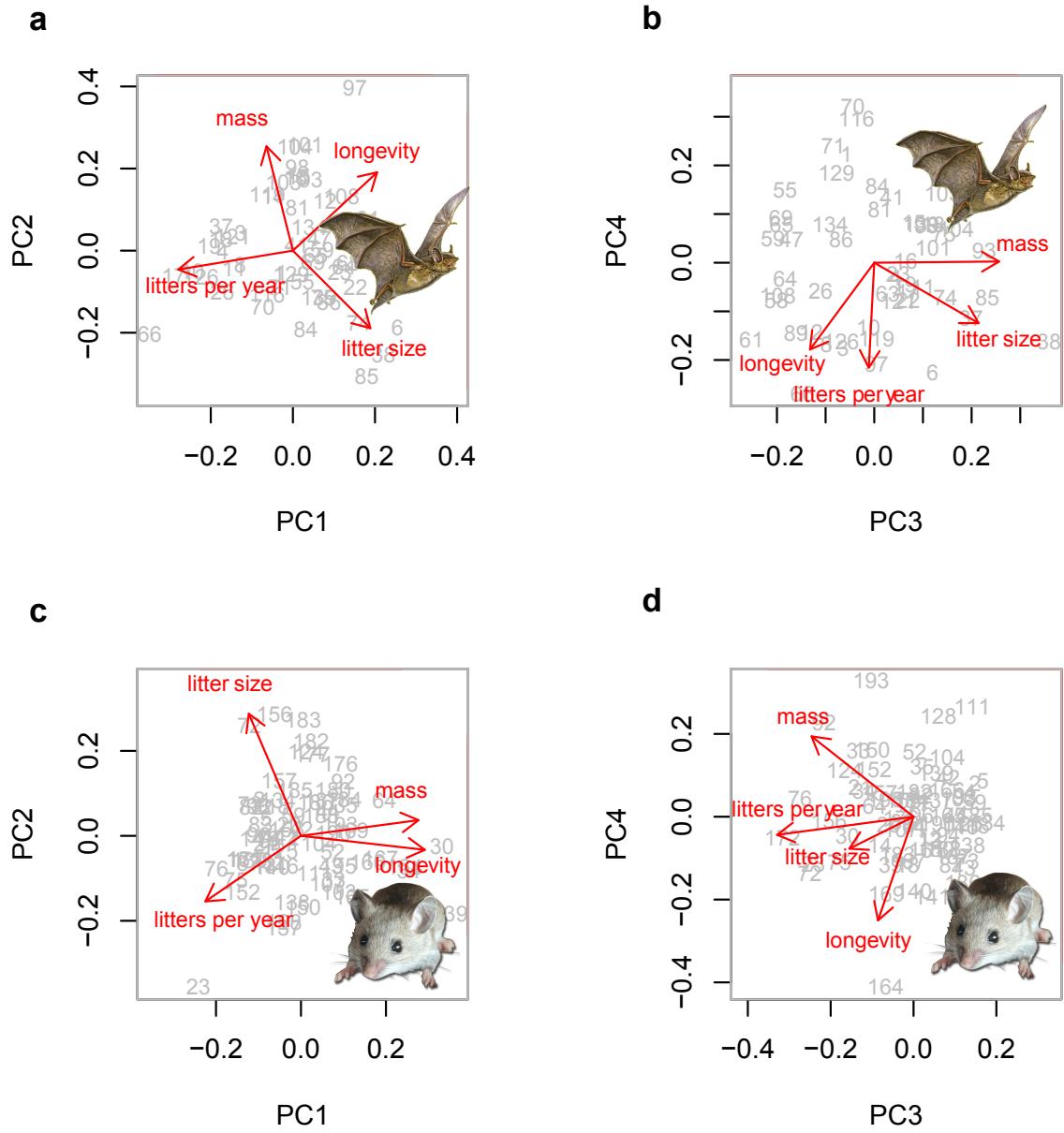


Fig. S6. Principal components analysis (PCA) for correlated life history traits in bats and rodents. a) Principal components 1 and 2 for bats; b) principal components 3 and 4 for bats; c) principal components 1 and 2 for rodents; and d) principal components 3 and 4 for rodents. See Table S2 for loading values and variance explained by each principal component.

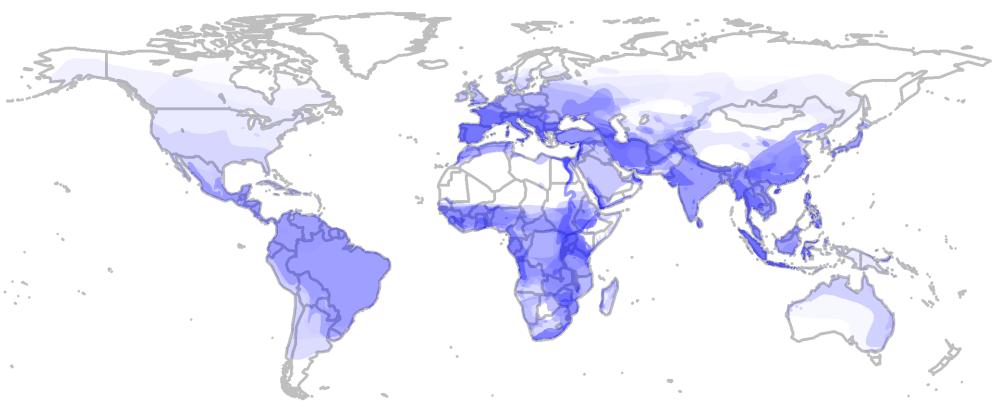


Fig. S7. Distributions of zoonotic bat viruses discounted by sampling effort. The shade of a virus' distribution is lighter than in Fig. 1 if its host(s) have been studied more (number of citations on Web of Science). We set the transparency of the color to $2/\text{sum}(\log_{10} \text{number of citations on Web of Science for all hosts})$, where 0 is fully transparent, and 1 is solid color. The Americas are a little lighter overall, reflecting how well these species have been studied. A few places in Africa become darker, indicating that these species have not been very well studied, and should perhaps be studied more thoroughly for their reservoir potential. Europe remains dark.

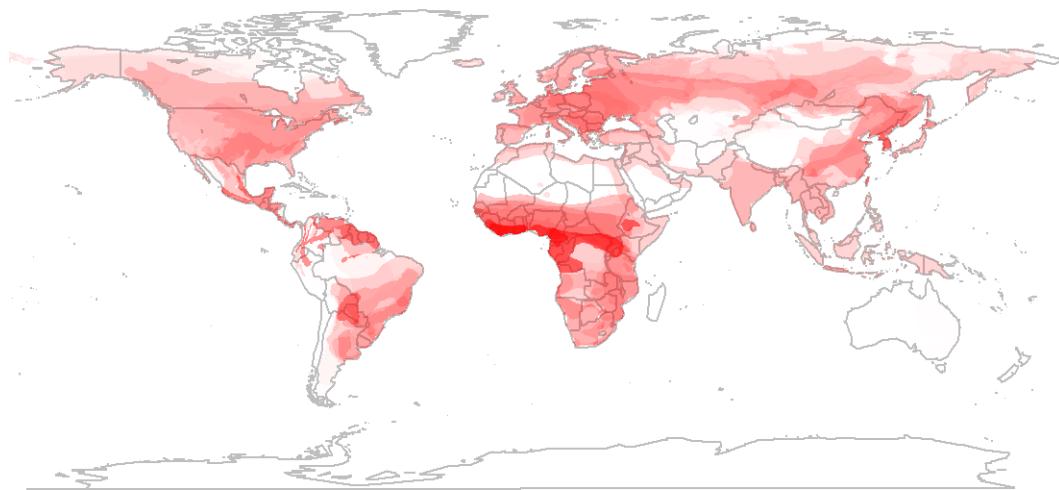


Fig. S8. Distributions of zoonotic rodent viruses discounted by sampling effort. The shade of a virus' distribution is lighter than in Fig. 1 if its host(s) have been studied more (number of citations on Web of Science). We set the transparency of the color to $2/\text{sum}(\log_{10} \text{number of citations on Web of Science for all hosts})$, where 0 is fully transparent, and 1 is solid color. Europe becomes lighter, indicating that the high viral richness identified in Europe may be due to sampling bias of these host species. Areas in central Africa become darker, perhaps suggesting that more virus discovery research should be directed to these areas.