**Supplementary Information**

**Wells et al. Distinct spread of DNA and RNA viruses among mammals amid prominent role of domestic species**

**SI Figures**

Figure S1 (eigenvector centrality measures for DNA and RNA virus sharing) …..…….…. 2

Figure S2 (effects of sampling bias on eigenvector centrality measures) …..……….…..…. 3

Figure S3 (host species with highest centrality measures) ……………………...……….…. 4

Figure S4 (network plot of mammal-virus associations) …..……………………...…….…. 5

**SI Tables**

Table S1 (test statistics for multiple pairwise comparison) ………………..………………. 6

Table S2 (table of mammal-virus interactions) ……………………….……………………. 7

**SI Methods**

Statistical analysis details …………………………………………….……………………. 9

Box S1 (model code for modelling virus sharing among hosts) ………..…………………. 14

Box S2 (model code for modelling proportion of zoonotic viruses) ………………………. 16

**SI References** …………………………………………………………………......………. 17



**Figure S1**. Eigenvector centrality scores (box plots and species data points) of host species from different mammalian orders, depicting their relative importance in virus sharing and spread across the entire network of mammal-virus associations. Larger values refer to host species sharing more viruses with others, especially with host species that are also well connected. Artiodactyla and Cetacea are presented as separate groups because of their distinct ecological niche, mammalian orders with few species are merged into the group ‘other’. Grey points represent measures for wild and red points measures for domestic mammalian host species and humans.



**Figure S2**. Effects of sampling bias on eigenvector centrality measures for mammalian host species. The left panel shows the correlation strength (Spearman’s R) in eigenvector centrality measures for all host species (n = 725 species) calculated from the full dataset with those from data subsets, in which different proportions of interactions have been randomly removed according to the proportional numbers of research articles (red points) and published sequences (blue points) for underlying host-virus interactions. The proportion of removed interactions ranged from 0.05 to 0.30 with a total of 200 generated data subsets (with 4 subsets for each removal rate). The right panel shows The Kruskal-Wallis D test statistic from comparing eigenvector centrality measures of domestic species with those from wildlife for the same data subsets. All Kruskal-Wallis tests were linked to p-values < 0.001.



**Figure S3**. List of mammalian host species (n= 33 species) which are among the top ten species with highest eigenvector centrality measures for any of the data subsets, in which different proportions of interactions have been randomly removed according to the proportional numbers of research articles (red points) and published sequences (blue points) for underlying host-virus interactions. The proportion of removed interactions ranged from 0.05 to 0.30 with a total of 200 generated data subsets (with 4 subsets for each removal rate). Each point indicates the inclusion of a species among the top ten species for the given subset. The ten species with largest eigenvector centrality measures from the full data set are marked with an asterisk.



**Figure S4**. Bipartite network plot of the associations between host species from different mammalian orders with different families of virus species based on interaction records among the 725 host species and 1,785 virus species in the database. Upper nodes represent virus families and lower nodes mammalian orders. The widths of edges represent the relative number of interactions recorded at species levels (i.e. mammal and virus species assigned to the respective groups). Note that larger node sizes and edge widths can results from multiple interactions of the same species (e.g. a single virus associated with multiple host species from the same order) or more species within a group (e.g. multiple virus species connect to multiple host species).

**Table S1**. Test statistics for Dunn’s Kruskal-Wallis multiple pairwise comparison of network eigenvector centrality measures (all viruses in database) for mammalian host species from different orders. Species of Artiodactyla and Cetacea are presented as separate groups because of their distinct ecological niche, mammalian orders with few species are merged into the group ‘other’. Adjusted p-values (Padj) were calculated following the Benjamini-Yekuteili method (Benjamini & Yekutieli, 2001), values of Padj < 0.05 are depicted with a single asterisk (“\*”) and Padj < 0.01 with two (“\*\*”).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Primates | Rodentia | Lagomorpha | Eulipotyphla | Carnivora | Perissodactyla | Artiodactyla | Cetacea | Chiroptera | other |
| Primates | - | 5.3\*\* | - | - | - | - | - | - | - | - |
| Rodentia | - | - | - | - | - | - | - | - | - | - |
| Lagomorpha | - | - | - | - | - | - | - | - | - | - |
| Eulipotyphla | -5.5\*\* | -2 | - | - | - | - | - | - | - | -2.2 |
| Carnivora | 3.3\*\* | 8.8\*\* | - | 7.9\*\* | - | - | - | - | 2.9\* | 7.3\*\* |
| Perissodactyla | - | - | - | - | - | - | - | - | - | - |
| Artiodactyla | 2.3 | 7.1\*\* | - | 7\*\* | -0.7 | - | - | - | 1.8 | 6\*\* |

**Table S2**. Table of interactions between mammalian host species from different orders with viruses from different families as recorded in our database. The number in parenthesis of column and row titles give the total number of host and virus species within the respective groups. Cell entries represent the respective number of host and virus species from each group recorded to interact.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Primates (97) | Rodentia (143) | Lagomorpha (7) | Eulipotyphla (36) | Carnivora (92) | Perissodactyla (5) | Artiodactyla (71) | Cetacea (1) | Chiroptera (206) | other (67) |
| Adenoviridae (89) | 9 / 19 | 3 / 3 | 0 / 0 | 0 / 0 | 7 / 6 | 1 / 2 | 5 / 11 | 1 / 0 | 39 / 48 | 4 / 6 |
| Anelloviridae (37) | 3 / 25 | 4 / 2 | 0 / 0 | 0 / 0 | 4 / 6 | 1 / 1 | 1 / 7 | 1 / 0 | 0 / 0 | 1 / 1 |
| Arenaviridae (40) | 1 / 8 | 39 / 35 | 0 / 0 | 1 / 1 | 0 / 0 | 0 / 0 | 0 / 0 | 1 / 0 | 0 / 0 | 2 / 3 |
| Arteriviridae (11) | 5 / 5 | 4 / 3 | 0 / 0 | 0 / 0 | 0 / 0 | 1 / 1 | 1 / 1 | 1 / 0 | 0 / 0 | 3 / 3 |
| Asfarviridae (1) | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 4 / 1 | 0 / 0 | 0 / 1 | 0 / 1 |
| Astroviridae (65) | 9 / 21 | 1 / 1 | 0 / 0 | 0 / 0 | 7 / 18 | 0 / 0 | 5 / 15 | 1 / 0 | 26 / 11 | 3 / 6 |
| Bornaviridae (3) | 1 / 2 | 0 / 0 | 1 / 1 | 2 / 2 | 2 / 1 | 1 / 2 | 1 / 1 | 1 / 0 | 2 / 1 | 3 / 1 |
| Caliciviridae (20) | 3 / 4 | 4 / 2 | 4 / 4 | 0 / 0 | 8 / 6 | 0 / 0 | 3 / 8 | 1 / 0 | 5 / 2 | 2 / 3 |
| Carmotetraviridae (1) | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 1 / 1 | 0 / 1 |
| Circoviridae (43) | 2 / 14 | 2 / 3 | 1 / 1 | 2 / 2 | 6 / 7 | 1 / 1 | 4 / 12 | 1 / 0 | 14 / 13 | 3 / 2 |
| Coronaviridae (95) | 3 / 9 | 9 / 7 | 1 / 1 | 2 / 3 | 10 / 5 | 1 / 1 | 11 / 11 | 1 / 0 | 69 / 71 | 7 / 11 |
| Dicistroviridae (1) | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 1 / 1 | 0 / 1 |
| Filoviridae (6) | 6 / 5 | 3 / 2 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 1 / 1 | 1 / 0 | 4 / 2 | 0 / 0 |
| Flaviviridae (65) | 19 / 22 | 26 / 14 | 1 / 1 | 1 / 1 | 2 / 2 | 1 / 10 | 17 / 22 | 1 / 0 | 33 / 17 | 13 / 12 |
| Genomoviridae (28) | 1 / 5 | 0 / 0 | 2 / 2 | 0 / 0 | 3 / 7 | 0 / 0 | 3 / 6 | 1 / 0 | 1 / 6 | 2 / 3 |
| Hantaviridae (120) | 1 / 12 | 63 / 62 | 0 / 0 | 32 / 45 | 1 / 1 | 0 / 0 | 0 / 0 | 1 / 0 | 11 / 8 | 8 / 14 |
| Hepadnaviridae (6) | 14 / 2 | 1 / 1 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 2 / 1 | 1 / 0 | 13 / 4 | 2 / 2 |
| Hepeviridae (3) | 1 / 1 | 0 / 0 | 0 / 0 | 0 / 0 | 1 / 1 | 0 / 0 | 2 / 1 | 1 / 0 | 6 / 1 | 0 / 0 |
| Herpesviridae (252) | 50 / 85 | 16 / 30 | 1 / 2 | 1 / 1 | 23 / 30 | 3 / 11 | 37 / 35 | 1 / 1 | 31 / 36 | 17 / 40 |
| Iflaviridae (1) | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 0 / 1 | 1 / 1 |
| Iridoviridae (1) | 1 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 0 / 1 | 0 / 1 |
| Lipothrixviridae (1) | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 1 / 1 | 0 / 0 | 0 / 1 | 0 / 1 |
| Luteoviridae (1) | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 1 / 1 | 0 / 1 |
| Marseilleviridae (2) | 1 / 2 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 1 / 0 | 0 / 0 | 0 / 0 |
| Mimiviridae (1) | 1 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 0 / 1 | 0 / 1 |
| Nairoviridae (6) | 1 / 2 | 1 / 1 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 3 / 1 | 1 / 0 | 3 / 4 | 0 / 0 |
| Nodaviridae (1) | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 1 / 1 | 0 / 1 |
| Orthomyxoviridae (5) | 1 / 3 | 1 / 1 | 0 / 0 | 0 / 0 | 13 / 1 | 1 / 1 | 3 / 3 | 1 / 0 | 0 / 0 | 6 / 2 |
| Papillomaviridae (136) | 5 / 58 | 8 / 6 | 3 / 1 | 1 / 1 | 12 / 14 | 2 / 10 | 12 / 21 | 1 / 0 | 7 / 16 | 7 / 17 |
| Paramyxoviridae (152) | 5 / 12 | 16 / 90 | 1 / 1 | 2 / 10 | 38 / 6 | 1 / 1 | 13 / 9 | 1 / 0 | 50 / 31 | 12 / 6 |
| Parvoviridae (86) | 8 / 21 | 2 / 6 | 1 / 1 | 2 / 2 | 27 / 16 | 1 / 1 | 4 / 22 | 1 / 0 | 18 / 14 | 4 / 8 |
| Peribunyaviridae (29) | 4 / 17 | 4 / 5 | 1 / 1 | 0 / 0 | 3 / 3 | 2 / 2 | 6 / 7 | 1 / 0 | 2 / 2 | 3 / 3 |
| Phenuiviridae (20) | 1 / 12 | 7 / 5 | 0 / 0 | 2 / 1 | 1 / 2 | 0 / 0 | 6 / 3 | 1 / 0 | 5 / 4 | 3 / 4 |
| Picobirnaviridae (16) | 1 / 2 | 0 / 0 | 1 / 1 | 0 / 0 | 4 / 5 | 1 / 4 | 3 / 4 | 1 / 0 | 0 / 0 | 1 / 1 |
| Picornaviridae (155) | 9 / 95 | 20 / 19 | 1 / 1 | 2 / 2 | 8 / 9 | 1 / 2 | 15 / 27 | 1 / 0 | 21 / 7 | 10 / 9 |
| Pneumoviridae (8) | 2 / 3 | 1 / 1 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 2 / 2 | 1 / 0 | 0 / 0 | 1 / 2 |
| Polyomaviridae (53) | 7 / 20 | 6 / 6 | 0 / 0 | 0 / 0 | 5 / 5 | 0 / 0 | 4 / 6 | 1 / 0 | 14 / 15 | 1 / 1 |
| Poxviridae (37) | 7 / 12 | 13 / 6 | 2 / 2 | 0 / 0 | 15 / 7 | 1 / 3 | 15 / 15 | 1 / 0 | 3 / 3 | 7 / 6 |
| Reoviridae (34) | 6 / 11 | 9 / 6 | 1 / 1 | 0 / 0 | 8 / 8 | 1 / 5 | 18 / 11 | 1 / 0 | 27 / 7 | 9 / 6 |
| Retroviridae (48) | 64 / 24 | 2 / 3 | 0 / 0 | 0 / 0 | 10 / 3 | 1 / 1 | 9 / 15 | 1 / 0 | 6 / 4 | 9 / 7 |
| Rhabdoviridae (43) | 3 / 6 | 5 / 5 | 0 / 0 | 0 / 0 | 38 / 6 | 1 / 5 | 13 / 9 | 1 / 0 | 74 / 25 | 7 / 5 |
| Smacoviridae (17) | 5 / 7 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 3 / 10 | 1 / 0 | 0 / 0 | 0 / 0 |
| Solemoviridae (1) | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 1 / 1 | 0 / 1 |
| Togaviridae (15) | 3 / 9 | 4 / 6 | 0 / 0 | 0 / 0 | 1 / 1 | 1 / 7 | 2 / 3 | 1 / 0 | 0 / 0 | 5 / 4 |
| Tombusviridae (1) | 1 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 0 / 1 | 0 / 1 |
| Totiviridae (2) | 1 / 2 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 1 / 0 | 0 / 0 | 0 / 0 |
| Virgaviridae (1) | 1 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 1 | 0 / 0 | 0 / 1 | 0 / 1 |

**Statistical analysis details**

The primary focus of this paper was to explore which mammalian host species might be the most important for spreading viruses due to their sharing of viruses with others, and we were interested in the phylogenetic and functional diversity of host species infected by different virus species. We addressed these aims using three different statistical approaches, namely 1) centrality measures of ecological network analysis to identify the most connected host species, 2) a hierarchical model to determine how the probability to share viruses among mammal species differ among host groups, wildlife versus domestic species and phylogenetic and functional distances between species, and, 3) a hierarchical model to explore variation in the proportion of zoonotic viruses among host species and groups of species. We conducted all analysis on interaction matrices including all virus species and only RNA versus DNA virus species, respectively.

**(a) Centrality of host species in networks of virus sharing and spread.**We calculated a number of different centrality measures to quantify the relative importance of host species in unimodal networks according to the number of viruses they share with others (Newman, 2003; Gómez *et al.*, 2013). For this, we generated an *N*×*N* adjacency matrix that depicts, for each possible pair of host species, the number of virus species they share. We calculated eigenvector centrality (a generalization of degree, which is the number of connections a host species has to others in terms of virus sharing; eigenvector centrality accounts both for the degree of a host species and those of connected species, i.e. it considers host species to be highly central if their connected species are connected to many other well-connected species (Bonacich & Lloyd, 2001)). Eigenvector centrality was strongly correlated with degree measures (unweighted degree, strength centrality, Opsahl degree centrality, all Spearman rank correlations with r ≥ 0.80), indicating that highly connected species (resulting in large degree values) were connected mainly to others with high degree. Eigenvector centrality was also correlated with betweenness centrality (accounts for the number of shortest paths crossing through a given node and represents relative importance in connecting otherwise disparate parts of a network (Brandes, 2001)) (Spearman r = 0.76) and also closeness centrality (an inverse measure of how many steps are required to link a host species to all others, that therefore describes its overall connectivity at network-level (Freeman, 1978) (Spearman r = 0.91). Thus, we present only results from eigenvector centrality and acknowledge that because of collinearity, it is not possible to distinguish further between the different components.

Centrality measures calculated from the interaction matrices of all and zoonotic viruses only, were strongly correlated (all Spearman rank correlations with r ≥ 0.82) and we present results from interaction matrices of all viruses and the two virus genome types only.

We used the non-parametric Kruskal-Wallis test to assess whether the eigenvector centrality measures differed between wildlife and domestic species and among host orders. We applied Dunn’s test for multiple comparisons (Benjamini & Yekutieli, 2001) among orders, whereby for simplicity, we merged orders with < 9 recorded host species into a single arbitrary group.

To account for sampling bias that could bias centrality measures (Costenbader & Valente, 2003), we randomly removed subsets of interaction records from the adjacency matrix used for calculating centrality measures. For this, we varied the proportion of removed interactions between 5 – 30% in each of 200 iterations following a uniform distribution. We used the relative proportion of publication and sequence numbers for each mammal-virus combination as two independent sets of probabilities of which interactions to remove. We then calculated centrality measured for each iteration and tested for correlations with centrality measures from the full data set (Spearman rank test) and possible differences between wildlife and domestic species (Kruskal-Wallis test). We also identified for each iteration the ten species with highest centrality measures.

**(b) Hierarchical model of virus sharing among host species.** We further analysed the *N*×*N* adjacency matrix of virus sharing among different pairs of host species in a probabilistic modelling framework to explore further host attributes that predict the frequency of virus sharing among host species. For simplicity, we generated a binary adjacency matrix with *z(i,j)* = 1 if the host species *i* and *j* were recorded to share any virus and *z(i,j)* = 0 otherwise (with *i* and *j* ∈ 1,…,*N* and *j* ≠ *i*). The probability *ϕ(i,j)* that two host species share any virus can be linked to *z(i,j)* with a Bernoulli distribution given as

*z(i,j) ~ Ɓernoulli[ϕ(i,j)].*

We used the logit-link function to model variation in *ϕ(i,j* ) as

logit*[ϕ(i,j)] ~ η(i)+ βphylorder(i) \* distphyl(i,j)* *+ βecolorder(i) \* distecol(i,j) + βdomest(i)* + *Ɓbias sqrt[Xbias(i)Xbias(j)].*

Here, *η(i)* is the species-specific intercept, which is further modelled with a hierarchical hyperprior *η(i)* as ~ N*[*H*η(order), ση(order)]*; the hyperprior H*η* accounts for the ‘average’ virus sharing probability of species from different orders, while the variance *ση* accounts for the deviation of species-level virus sharing-probabilities from the respective order-level hyperprior. The coefficients *βphyl* and *βecol*account for variation in virus sharing with increasing phylogenetic and ecological distance from *i*, whereby these coefficients were allowed to vary across host orders. Distances between infected host species that differ from random draws from potential host pools indicate specificity; lower values (i.e. 95% credible intervals < 0) indicate higher similarity between observed hosts than expected; values > 0 suggest that viruses infect more distantly related hosts than expected (Wells *et al.*, 2019).

The coefficient *βdomest* accounts for variation in virus sharing among all possible combinations between species classified as wildlife, domestic, or human compared to pairs of wildlife-wildlife species (a five-level categorical variable), while *Ɓbias* accounts for variation in relation to the four different proxies of sampling efforts described above. Covariates from proxies of sampling efforts were generated as the square-rooted product of pairwise proxy variables. We used pseudo priors for diagonal values of the adjacency matrix (i.e. *i* = *j*) in order to exclude irrelevant values of *z(i,j)* from the likelihood estimation.

We fitted the model in a Bayesian framework with Markov Chain Monte Carlo (MCMC) sampling in the software JAGS version 4.3.0, operated via the R package *rjags* (Plummer, 2016). We ran two chains of 50,000 iterations each for parameter adaptation, then sampled 1,000 posterior parameter estimates. See Supplementary information (SI Appendix, **Box S1**) for details.

**(c) Hierarchical model of the proportion of zoonotic viruses carried by different host species.** To examine whether mammalian host species differ in their propensity to carry zoonotic viruses (i.e., those recorded also in humans), we used a probabilistic hierarchical model to estimate the proportion of zoonotic viruses carried by each host species. We modelled the probability that a virus recorded for a host species *i* is zoonotic (corresponding to the likely proportion of zoonotic viruses carried by a host species) using a binomial distribution. This is because for any virus recorded, there is a probability *ψ(i)* it has been also recorded in humans given the number of zoonotic viruses *y(i)* out of the total number of viruses *w(i)* recorded for *i*. We thus assumed

*y(i) ~ Ɓin[w(i), ψ(i)]*.

We then used the logit-link function to model variation in *ψ(i)* among different host species as

logit*[ψ(i)] ~ µorder(i) + X(i)B.*

Here, *µorder* denotes the order-specific average according to the taxonomic order of species *i*, which were modelled with a Gaussian error structure and a common ‘average’ hyperprior mean, i.e. *µorder ~ Ɲ(H, σ2)*. *X* is a matrix of the 17 species-level covariates (including phylogenetic distance to humans and the four proxies of sampling bias) described above and *B* is a vector of corresponding coefficient estimates. We fitted the model in a Bayesian framework in JAGS (Plummer, 2016). See Supplementary information (SI Appendix, **Box S2**) for details.

**Box S1**. Model code for estimating the variation in the probability of virus sharing among mammalian host species. Data variables are as follows: *N* – number of host species in adjacency matrix; *adj.obs* - *N*×*N* adjacency matrix (binary) of virus sharing between different host species; *phyl.dist* – *N*×*N* matrix of scaled phylogenetic distances between pairs of host species; *ecol.dist* – *N*×*N* matrix of scaled functional distances between pairs of host species; *adj.type* – *N*×*N* matrix of type of relationships among pairs of host such as ‘wildlife-wildlife’ or ‘wildlife-domestic’; *evidence.1* – *N*×*N* matrix of research effort, given as the scaled square-root product of research efforts for the two respective host species, note that four indices of research efforts are used in the analysis (matrices *evidence.1,…, evidence.4*) as described in the methods.

|  |
| --- |
| model {  for(i in 1:N){  for(j in 1:N){  # Likelihood fit:  adj.obs[i,j] ~ dbern(phi.star[i,j, i.phi[i,j]])  # Indicator of diagonal/non-diagonal matrix entries  i.phi[i,j] <- 1 + equals(i,j)  phi.star[i,j, 2] ~ dbern(1) # Pseudoprior for diagonal matrix entries  phi.star[i,j, 1] <- phi[i,j]  # Logit-link model of virus sharing probability  logit(phi[i,j]) <- mu\_phi[i]  + b.phyl[order[i]] \* phyl.dist[i,j]  + b.ecol[order[i]] \* ecol.dist[i,j]  + c.type[adj.type[i,j]]  + b.evid[1] \* evidence.1[i,j]  + b.evid[2] \* evidence.2[i,j]  + b.evid[3] \* evidence.3[i,j]  + b.evid[4] \* evidence.4[i,j]  }  }  # Prior  for(i in 1:N){  mu\_phi[i] ~ dnorm(muTaxa\_phi[order[i]], tauTaxa\_phi[order[i]])  }  for(x in 1:norder){  muTaxa\_phi[x] ~ dnorm(muHyp\_phi, tauHyp\_phi)  tauTaxa\_phi[x] <- pow(sdTaxa\_phi[x], -2); sdTaxa\_phi[x] ~ dunif(0,10)  b.phyl[x] ~ dnorm(0, tau\_bphyl[x, i\_bphyl[x]])  b.ecol[x] ~ dnorm(0, tau\_becol[x, i\_becol[x]])  # Gibbs variable selection (GVS) prior  i\_bphyl[x] <- sel\_bphyl[x] + 1  sel\_bphyl[x] ~ dbern(0.05)  tau\_bphyl[x,1] <- 100  tau\_bphyl[x,2] <- pow(sd\_bphyl[x], -2); sd\_bphyl[x] ~ dexp(0.5)  i\_becol[x] <- sel\_becol[x] + 1  sel\_becol[x] ~ dbern(0.05)  tau\_becol[x,1] <- 100  tau\_becol[x,2] <- pow(sd\_becol[x], -2); sd\_becol[x] ~ dexp(0.5)  }  c.type[1] <- 0  for(c in 2:ntype){  c.type[c] ~ dnorm(0, tau\_type[c,i\_type[c]])  i\_type[c] <- sel\_type[c] + 1  sel\_type[c] ~ dbern(0.05)  tau\_type[c,1] <- 100  tau\_type[c,2] <- pow(sd\_type[c], -2); sd\_type[c] ~ dexp(0.5)  }  muHyp\_phi ~ dnorm(0, 0.01)  tauHyp\_phi <- pow(sdHyp\_phi, -2); sdHyp\_phi ~ dunif(0,10)  for(b in 1:4){  b.evid[b] ~ dnorm(0, tau\_evid[b,i\_evid[b]])  i\_evid[b] <- sel\_evid[b] + 1  sel\_evid[b] ~ dbern(0.05)  tau\_evid[b,1] <- 100  tau\_evid[b,2] <- pow(sd\_evid[b], -2); sd\_evid[b] ~ dexp(0.5)  }  ## Impute missing values  for(i in 1:N){  for(j in 1:N){  phyl.dist[i,j] ~ dnorm(0,1)  ecol.dist[i,j] ~ dnorm(0,1)  npaper.dist[i,j] ~ dnorm(0,1)  }  }  } |

**Box S2**. Model code for estimating the the proportion of zoonotic viruses carried by different host species. Data variables are as follows: *N* – number of host species;

*nvirus.zoon* – vector of length *N*, depicting for each host species the number of associated zoonotic virus species; *nvirus.all* – vector of length *N*, depicting for each host species the total number of associated virus species; *host.isdomest* – binary vector of length *N*, indicating for each host species of whether it is classified as ‘domestic’ or not; *Traits* – matrix of scaled host species traits used as covariates in the regression model.

|  |
| --- |
| model {  for (i in 1:N){  # Likelihood fitting the data to the model  nvirus.zoon[i] ~ dbin(p.zoonv[i], nvirus.all[i])  logit(p.zoonv[i]) <- mu.p[order[i]]  + beta.domest \* host.isdomest[i]  + inprod(Traits[i, ], B\_traits)  }  # Prior  for(o in 1:norder){  mu.p[o] ~ dnorm(muHyp\_p, tauHyp\_p[o])  tauHyp\_p[o] <- pow(sdHyp\_p[o], -2); sdHyp\_p[o] ~ dunif(0,10)  }  muHyp\_p ~ dnorm(0, 0.001)  beta.domest ~ dnorm(0, 0.01)  for(x in 1:ntrait){  B\_traits[x] ~ dnorm(0, tau\_trait[x, i\_tau\_trait[x]])  i\_tau\_trait[x] <- sel\_tau\_trait[x] + 1  sel\_tau\_trait[x] ~ dbern(0.05)  tau\_trait[x,1] <- 100  tau\_trait[x,2] <- pow(sd\_trait[x], -2); sd\_trait[x] ~ dexp(0.5)  }  #Impute missing trait data  for (i in 1:N){  for(x in 1:(ntrait-3)){  Traits[i,x] ~ dnorm(0, 1)  }  for(x in (ntrait-2):ntrait){  Traits[i,x] ~ dbern(0.5)  }  }  } |

**SI References**

Benjamini Y. & Yekutieli D. (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 29, 1165-1188. https://doi.org/10.1214/aos/1013699998

Bonacich, P. & Lloyd P. (2001) Eigenvector-like measures of centrality for asymmetric relations. *Social Networks* 23, 191-201. https://doi.org/10.1016/S0378-8733(01)00038-7

Brandes, U. (2001) A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology* 25, 163-177. https://doi.org/10.1080/0022250X.2001.9990249

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. doi:10.1073/pnas.1220716110

Newman, M.E.J. (2003) The structure and function of complex networks. *Siam Review* 45, 167-256. https://doi.org/10.1137/S003614450342480

Wells, K., Gibson, D.I. & Clark, N.J. (2019) Global patterns in helminth host specificity: phylogenetic and functional diversity of regional host species pools matter. *Ecography*, 42, 416-427. https://doi.org/10.1111/ecog.03886