# Distinct spread of DNA and RNA viruses among mammals amid prominent role of domestic species

**Running head: Virus sharing in mammalian networks**


## Abstract

**Aim:** Emerging infectious diseases arising from pathogen spillover from mammals to humans comprise a substantial health threat. Tracing virus origin and predicting the most likely host species for future spillover events are major objectives in One Health disciplines. We assessed patterns of virus sharing among a large diversity of mammals, including humans and domestic species.

**Location:** Global.

**Time period:** Current.

**Major taxa studied:** Mammals and associated viruses.

**Methods:** We used network centrality analysis and trait-based Bayesian hierarchical models to explore patterns of virus sharing among mammals. We analysed a global database that compiled the associations between 1,785 virus species and 725 mammalian host species as sourced from automatic screening of meta-data accompanying published nucleotide sequences between 1950 – 2019.

**Results:** We show that based on current evidence, domesticated mammals hold the most central positions in networks of known mammal-virus associations. Among entire host-virus networks, Carnivora and Chiroptera hold central positions for mainly sharing RNA viruses, while Ungulates hold central positions for sharing both RNA and DNA viruses with other

host species. We revealed strong evidence that DNA viruses were phylogenetically more host specific than RNA viruses. RNA viruses exhibited low functional host specificity despite an overall tendency to infect phylogenetically related species, signifying high potential to shift across hosts with different ecological niches. The frequencies of sharing viruses among hosts and the proportion of zoonotic viruses in hosts were larger for RNA than DNA viruses.

**Main conclusions:** Acknowledging the role of domestic species in addition to host and virus traits in patterns of virus sharing is necessary to improve our understanding of virus spread and spillover in times of global change. Understanding multi-host virus sharing pathways adds focus to curtail disease spread.

**Keywords**

Global virus spread, disease emergence, disease risk assessment, host-parasite interaction, pathogen spillover, zoonotic disease risk, network analysis

# 1. INTRODUCTION

Pathogen spillover and cross-species transmission between animals and humans is a major source of infectious diseases and a considerable global public health burden (Jones et al., 2008; Karesh et al., 2012). Understanding the factors that enable or facilitate these processes is a crucial step for such events to be predicted. Host shifting, that is the colonization of a new host species by a pathogen, requires a certain level of overlap in species traits ('ecological fitting') in order to overcome barriers of cross-species transmission and for

survival and reproduction within novel host species (Woolhouse et al., 2005; Parrish et al., 2008; Agosta et al., 2010). In the search for mechanisms and enabling conditions that may help to predict the future emergence of infectious diseases from animal populations, the necessity of considering entire host species communities amongst underpinning biogeographic structure and connectivity have been recently emphasized (Poulin, 2010; Fenton et al., 2015; Clark et al., 2018; Wells et al., 2018).

Network analyses that describe the connections of different host species in terms of parasite sharing have proven useful in analysing host specificity and parasite spread (Gómez et al., 2013; Luis et al., 2015), particularly since they offer the opportunity to explore community-wide pathogen spread (the distribution of a pathogen among host species, a pattern emerging from past and contemporary host shifting events that connect host species as nodes in a network). Other recent 'big data' studies of mammal-virus associations have explored whether host traits and geographic distribution can predict those species that most likely harbour undiscovered viruses that may cause future pandemics using trait-based regression analysis (Han et al., 2015; Luis et al., 2015; Olival et al., 2017). Such approaches may lead to increased predictability of future pandemics.

Yet despite important advances in virus discovery and analytical approaches, our understanding of virus sharing and their spread through entire networks of mammalian host species remains limited. The challenge of assessing different animal species in their role for virus spread is understandable, as detailed information about virus sharing across entire communities became only recently available (Wardeh et al., 2015; Olival et al., 2017) amid the challenge that many virus species remain unknown (Carroll et al., 2018).

We address this knowledge gap by exploring the role of different mammalian species in the spread of viruses through entire host communities. In particular, we tested whether domestic species (livestock and companion animals) play a major role in virus spread and

76 spillover among humans and wildlife. To this end, there are strong reasons why domesticated
77 animals should cover central positions in networks of host-virus associations. Domesticated
78 animals share large numbers of viruses and other parasites with humans (Morand et al., 2014)
79 and were recently reported to play crucial roles in the sharing of helminth parasites between
80 humans and wildlife (Wells et al., 2018). Moreover, the large numbers of domestic animals
81 compared to those of wildlife (Bar-On et al., 2018), and close contact between them and
82 people, creates ground for frequent and multilateral exposure. For entire networks of viruses
83 and mammalian host associations, we also expect different patterns of virus sharing for the
84 two different genome types of DNA and RNA viruses. Greater rates of replication error and
85 higher genetic diversity in RNA virus populations have been proposed to increase their host
86 range through more frequent host shifting and adaptation to distantly related host species,
87 whereas DNA viruses and retroviruses are assumed to be more host-specific due to stronger
88 codivergence with their hosts over much longer evolutionary timescales (Cleaveland et al.,
89 2001; Jackson & Charleston, 2004; Geoghegan et al., 2017; Longdon et al., 2018). With the
90 mounting recognition that host use in parasites seems to be more constrained by ecological
91 opportunity than by evolutionary history, there is an urgent need to understand and quantify
92 pathogen spread and host shifting capacity in response to specific traits at global scale (Nylin
93 *et al.*, 2018; Wells & Clark, 2019). Yet, to date little comprehensive work has explored of
94 whether host sharing and virus spread at the network level differ among these types of viruses
95 and whether they interact with the various groups of mammals in different ways. We used
96 network centrality analysis and Bayesian hierarchical models to quantify the extent of virus
97 sharing among different mammalian host species and the proportion of zoonotic viruses
98 carried in different hosts. If domestic species are key drivers of virus spread, we expect them
99 to occupy central positions in networks of pathogen sharing at the human-domestic animal-

wildlife interface, whereby variation in the host specificity of viruses may curtail their spread among the diversity of mammalian hosts at global scale.

# 2. METHODS

## 2.1 Virus-host data

We extracted mammal-virus species-level interactions from the Enhanced Infectious Diseases Database (EID2) (Wardeh et al., 2015) in the version from March 2019. In brief, EID2 utilises automated mining procedures to extract information on pathogens, their hosts and locations from two sources: 1) the meta-data accompanying nucleotide sequences published in the National Center for Biotechnology Information (NCBI) Nucleotide database (www.ncbi.nlm.nih.gov/nuccore); and 2) titles and abstracts of publications indexed in the PubMed database (www.ncbi.nlm.nih.gov/pubmed). To date, EID2 has extracted information from > 7 million sequences (and processed 100M+ sequences), and >8 million titles and abstracts. EID2 imports the names of organisms and their taxonomic hierarchy from the NCBI Taxonomy database (http://www.ncbi.nlm.nih.gov/Taxonomy/), and aligns it with an exhaustive collection of alternative names. In general, EID2 follows the NCBI definitions of 'species' and 'subspecies', with unclassified and uncultured species being denoted as 'no rank'.

The data of interest for this study were associations of mammalian species (including humans) with different virus species, independent of location records. We considered a mammalian species to be host to a virus if at least one NCBI meta-data set accompanying a published sequence detailed an association between the virus (or any of its subspecies or strains) and the host (or any of its subspecies), including detailed information about the sampling location (e.g. country/county where the association was recorded). We used this conservative approach rather than the full range of information collated from sequence

125 records and text mining in order to reduce any possible bias from experimental infection
126 studies. However, while we assume that sampling locations are most likely recorded as
127 metadata for natural infection, we are aware that our dataset may include non-natural
128 infections.

129 Virus species were assigned to genome type (DNA, RNA or other/unspecified)
130 following NCBI taxonomy as utilised by EID2. Mammal species synonyms and taxonomic
131 orders were standardized using the taxonomy of Wilson and Reeder (2005), the online
132 version of IUCN Red List and Integrated Taxonomic Information System, ITIS (accessed
133 May 2018). This revision enabled us to match the most recent host names to trait data.

134 Of the 724 non-human mammalian host species in our data set, we considered 21
135 species as 'domestic' (including the major commensal rodent species) and all other as
136 'wildlife'. Domestic species were banteng (*Bos javanicus*), yak (*B. mutus*), cow (*B. taurus*),
137 water buffalo (*Bubalus bubalis*), bactrian camel (*Camelus bactrianus* and *C. ferus*),
138 dromedary (*C. dromedarius*), dog (*Canis familiaris* and *Canis lupus*), goat (*Capra aegagrus*),
139 guinea pig (*Cavia porcellus*), wild ass (*Equus africanus*), donkey (*E. asinus*), horse (*E.*
140 *caballus*), cat (*Felis catus*), guanaco (*Lama guanicoe*), house mouse (*Mus musculus*), rabbit
141 (*Oryctolagus cuniculus*), sheep (*Ovis aries*), brown rat (*Rattus norvegicus*), black rat (*R.*
142 *rattus*), pig (*Sus scrofa*) and vicugna (*Vicugna vicugna*). We constrained our domestic species
143 selection to these major domestic species only to showcase possible differences in pathogen
144 sharing, while we are aware that there are some additional species that may be considered to
145 be domestic animals.

146 We generated four different measures of sampling effort for each mammalian host species,
147 namely 1) number of PubMed-indexed publications (summed over all associated virus
148 species), 2) number of virus sequences recorded (summed over all associated virus species),
149 3) Shannon diversity of publication records, accounting for the proportional number of

publications for each associated virus species and 4) Shannon diversity of sequence records, accounting for the proportional numbers of sequence records for each associated virus species. For Shannon indices larger values are linked to overall larger number of records and a more even distribution of records among different virus species, i.e. higher overall sampling coverage (Magurran, 2004). We generated these multiple indices as proxies of sampling intensity, as the true sampling effort is not known. This is because records of species interactions in the literature are arguably 'presence-only' records and rarely report the lack of interactions or the number of host individuals examined that would reduce the number of pseudo-absences in biotic interaction data (Little, 2004; Wells et al., 2013).

### 2.2 Mammalian host phylogeny and ecological trait data

A goal of this study was to assess whether variation in the phylogenetic and ecological similarities of mammalian species predict patterns of virus sharing (i.e., pairs-wise phylogenetic and ecological distances that are calculated among all possible combinations of viable host species) and the proportion of zoonotic viruses (i.e., viruses infecting humans and at least one other animal species) associated with different host species. We gathered ecological trait data from the PanTHERIA (Jones et al., 2009) and EltonTraits 1.0 (Wilman et al., 2014) databases to characterise all of the sampled mammals using a range of traits likely to impact on their suitability as hosts for viruses.

Selected traits were: body mass, which is a key feature of mammals in terms of their metabolism and adaptation to environments; average longevity, litter size and the average number of litters per year as demographic parameters that could be relevant for within-host dynamics of viruses; diet breadth (calculated as a Shannon diversity index based on the proportional use of 10 diet categories as presented in EltonTraits); range area, which we expect to affect the exposure to other mammalian host species; average temperature and

175 average precipitation within a host's distribution as an indicator of climatic niche; latitudinal
176 centroid of distribution as an indicator of the general habitat and climate within which hosts
177 are occurring across a gradient from tropical to polar environments; and habitat as multiple
178 binary indicators of whether a species uses 1) forest, 2) open vegetation, and/or 3)
179 artificial/anthropogenic habitats. Information on specific habitat utilisation was compiled
180 from the International Union for the Conservation of Nature (IUCN) database
181 (http://www.iucnredlist.org). Missing trait data were randomly imputed (as part of the
182 Bayesian sampling approaches, see model codes in Supporting Information Appendix S1).
183 We did not include a larger set of ecological traits in our analysis to avoid collinearity issues.
184 Phylogenetic relationships between sampled mammal species were estimated from a
185 recent mammalian supertree (Fritz et al., 2009). We used this tree to compute pairwise
186 phylogenetic distances based on a correlation matrix of phylogenetic branch lengths (Paradis
187 et al., 2004) and also a vector of phylogenetic distance to humans for all other mammalian
188 host species. We also quantified pairwise ecological distance between sampled mammal
189 species based on a generalised form of Gower's distance matrices (Gower, 1971) using
190 weighted variables based on all of the ecological trait variables described above, following
191 methods in Pavoine *et al.* (2009). Phylogenetic and ecological distance matrices as well as
192 vectors of trait variables were scaled (dividing by the maximum for each distance matrix), so
193 all distance measures ranged from zero to one. Data formatting and analyses were conducted
194 in R version 3.4.3 (R Development Core Team, 2017) and used the packages *ape* (Paradis et
195 al., 2004) for phylogenetic distance calculations and *ade4* (Dray & Dufour, 2007) for
196 ecological distance calculations.
197
198 **2.3 Statistical analysis**

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, which we describe in detail in the SI Appendix. In brief, we used the following approaches:

*Centrality of host species in networks of virus sharing*

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, betweenness centrality, and closeness centrality (all Spearman $r \geq 0.76$). 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.

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). To account for sampling variation that could bias centrality measures (larger sample sizes may increase the relative number of interactions reported for poorly sampled host species)(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 consistency of results from subsets and the full dataset.

*Hierarchical model of virus sharing among host species*

We generated a binary $N \times N$ adjacency matrix with $z(i,j) = 1$ if the pair of host species $i$ and $j$ were recorded to share any virus and $z(i,j) = 0$ otherwise (with $i$ and $j \in 1,\ldots,N$ and $j \neq i$). The probability $\varphi(i,j)$ that two host species share any virus can be linked to $z(i,j)$ with a Bernoulli distribution given as

$$z(i,j) \sim Bernoulli[\varphi(i,j)].$$

We used the logit-link function to model variation in $\varphi(i,j)$ as

$$\text{logit}[\varphi(i,j)] \sim \eta(i) + \beta phyl_{order}(i) * dist_{phyl}(i,j) + \beta ecol_{order}(i) * dist_{ecol}(i,j) + \beta domest(i) + \mathcal{B}_{bias}\, sqrt[X_{bias}(i)X_{bias}(j)].$$

Here, $\eta(i)$ is the species-specific intercept, which is further modelled with a hierarchical hyperprior $\eta(i)$ as $\sim N[H_\eta(order), \sigma_\eta(order)]$; the hyperprior $H_\eta$ accounts for the 'average' virus sharing probability of species from different orders, while the variance $\sigma_\eta$ accounts for the deviation of species-level virus sharing-probabilities from the respective order-level hyperprior. The coefficients $\beta_{phyl}$ and $\beta_{ecol}$ account for variation in virus sharing with increasing phylogenetic and ecological distance from $i$. The coefficient $\beta_{domest}$ 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). The coefficients $\mathcal{B}_{bias}$ account for variation in relation to the four different proxies of sampling efforts described above, i.e. they control for sampling variation in the probabilistic model framework. Covariates from proxies of sampling efforts were generated as the square-rooted product of pairwise proxy variables. We fitted the model in a

249 Bayesian framework with Markov Chain Monte Carlo (MCMC) sampling in the software
250 JAGS version 4.3.0, operated via the R package *rjags* (Plummer, 2016).
251
252 *Hierarchical model of the proportion of zoonotic viruses carried by different host species*
253 We modelled the probability $\psi(i)$ that a virus recorded for a host species $i$ is zoonotic
254 (corresponding to the likely proportion of zoonotic viruses carried by a host species) using a
255 binomial distribution based on the number of zoonotic viruses $y(i)$ out of the total number of
256 viruses $w(i)$ as
257 $$y(i) \sim \mathcal{B}in[w(i), \psi(i)].$$
258 We then used the logit-link function to model variation in $\psi(i)$ among different host species
259 as
260 $$\text{logit}[\psi(i,t)] \sim \mu_{order}(i) + X(i)B.$$
261 Here, $\mu_{order}$ denote the order-specific average according to the taxonomic order of species $i$,
262 which was modelled with a Gaussian error structure and a common 'average' hyperprior
263 mean, i.e. $\mu_{order} \sim \mathcal{N}(H, \sigma^2)$. $X$ is a matrix of the 17 species-level covariates (including
264 phylogenetic distance to humans and the four proxies of sampling bias) described above and
265 $B$ is a vector of corresponding coefficient estimates. This model accounts for sampling
266 variation similar to the model of virus sharing (through variation partitioning among multiple
267 covariates that are assumed to either represent the relevant biological processes or proxies of
268 sampling bias). We fitted the model in a Bayesian framework in JAGS (Plummer, 2016).
269

270 # 3. RESULTS

271 Of 1,785 virus species associated with 725 different mammalian host species (including
272 humans) in our dataset, 405 species (23%) have been recorded to infect humans. Out of these,
273 138 species (34% virus species infecting humans) are recorded as zoonotic. Of these zoonotic

274 species, 56 (41%) were recorded in wildlife but not in any domestic species, while 21 species
275 (15%) were recorded in humans and domestic animals but not in any wildlife species; the
276 remaining 61 zoonotic viruses were recorded in both wildlife and domestic species. In turn,
277 87 (5%) of all recorded virus species were shared by at least one domestic and one wildlife
278 species without being associated with humans.
279 The virus species included 730 DNA virus species and 912 RNA virus species (73 classified
280 as 'others'), of which 24 (3% of DNA virus species) and 91 (10% of RNA virus species)
281 were recorded as zoonotic. The overall network topography for DNA versus RNA viruses
282 reveal distinct spread of these viruses among host species, mostly depicted by considerably
283 lower virus sharing across orders of host species for DNA viruses (**Figure 1**).
284
285 **3.1 Centrality of host species in networks of virus sharing and spread**
286 Eigenvector centrality measures were higher for domestic than wildlife host species (Kruskal-
287 Wallis $\chi 2 \geq 35$, df= 1, $p < 0.01$), indicating that domestic species were the most central
288 species (after humans) in the entire mammal-virus association network based on current
289 evidence. The ten most central position in the network of all virus species were occupied by
290 *Homo sapiens*, *Bos taurus*, *Sus scrofa*, *Ovis aries*, *Canis lupus*, *Capra hircus*, *Equus*
291 *caballus*, *Felis catus*, *Bubalus bubalis*, and *Mus musculus* (following order of descending
292 centrality).
293 Centrality measures also varied among the different taxonomic orders of host species (all
294 Kruskal-Wallis $\chi 2 \geq 162.4$, df = 9, $p < 0.01$) (**Figure 2**). Specifically, eigenvector centrality
295 measures for all virus species were largest for wildlife species of the taxa Carnivora,
296 Chiroptera, Artiodactyla and Primates compared to other taxa (Rodentia, Eulipotyphla,
297 others) according to post-hoc multiple comparisons (Supporting Information, **Table S1**).
298 RNA viruses but *not* DNA viruses accounted for relatively larger centrality scores for

299 Carnivora and Chiroptera (both Mann–Whitney U test of group-level comparisons $p < 0.01$),
300 whereas centrality scores calculated for RNA and DNA viruses appeared to be of
301 indistinguishable ranks for Artiodactyla (Mann–Whitney U test $p = 0.52$) (Supporting
302 Information, **Figure S1**).
303 Centrality measures calculated from subsets of the underpinning adjacency matrix for all
304 viruses, with 5 – 30% of interactions removed according to number of published sequences
305 and publications, revealed a 4-fold stronger decline in correlations for the number of
306 published sequences than publications, but for all subsets, correlations with centrality
307 measures from the full data set remained reasonably high (i.e., all Spearman's $R > 0.6$ for
308 centrality measures with up to 30% of interactions removed; Supporting Information, **Figure**
309 **S2**). For these data subsets, there were a total of 28 host species that emerged as the top ten
310 host species according to centrality measures calculated from data subsets (Supporting
311 Information, **Figure S3**). However, despite this uncertainty in which host species occupied
312 the most central positions, the findings of significant larger centrality measures for domestic
313 than wildlife species hold true for all subsets (all Kruskal-Wallis tests with $\chi 2 \geq 18.3$, df= 1, p
314 < 0.01) (Supporting Information, **Figure S2**). Likewise, centrality measures varied among the
315 different taxonomic orders for all subsets (all Kruskal-Wallis tests with $\chi 2 \geq 22.3$, df= 1, p <
316 0.01) with the same order showing the largest centrality measures than for the full data set.
317
318 **3.2 Virus sharing among host species**
319 Analysing virus sharing patterns in a probabilistic hierarchical modelling framework
320 confirmed the prominent role of domestic animals in virus sharing across the entire network.
321 Wild mammalian host species were ca. 5.7 times (95% credible intervals [CIs] of odds ratio 5
322 – 9.3) more likely to share virus species with humans and ca. 4.2 times (odds ratio 4.9 – 5.5)
323 more likely to share virus species with domestic animals than with any other wild species.

Any pair of domestic species was ca. 70 times (odds ratio: 49.4 – 102.5) more likely to share viruses than any pair of two wildlife species. Humans shared DNA viruses ca. 33 times (odds ratio: 7 – 147) more often with any domestic species than DNA viruses were shared among any pair of two wildlife species, but we found no evidence that RNA viruses were shared more frequently by humans and any domestic species than among any pair of wildlife species (odds ratio: 1 – 126).

We found the highest frequencies of sharing RNA virus with any other mammalian species for species of the orders Chiroptera and Carnivora (averaging frequencies of 0.5 – 2% according to CIs of sharing RNA viruses with other species), whereas DNA virus sharing frequencies were mostly below 0.2% (according to upper bounds of CIs except for the orders Perissodactyla and Cetacea, for which large CIs indicated imprecise estimates)(**Figure 3**). For most host orders (except Cetacea) and both virus genome types, we found virus sharing to be more likely with closely related species (negative values for coefficients $\beta_{phyl}$ that depict increasing virus sharing for smaller phylogenetic distances among pairs of host species). Phylogenetic clustering of host species (which translates into higher phylogenetic host specificity for the viruses) was stronger for DNA viruses compared to RNA viruses shared by Primates, Carnivora, Artiodactyla and Chiroptera (**Figure 3**), signifying a general tendency of higher host specificity in terms of phylogenetic similarity for DNA viruses compared to RNA viruses. This tendency, however, is not true for viruses shared by Rodentia, as phylogenetic host specificity appeared to be relatively stronger for RNA than DNA viruses associated with species from this order (**Figure 3**).

Notably, phylogenetic host specificity for RNA viruses shared by Primates was relatively low, suggesting more frequent host sharing with more phylogenetically distant host species than in other orders (**Figure 3**). We found species of the orders Primates, Carnivora, Artiodactyla and Chiroptera to share RNA viruses with any other hosts of larger functional

distances than expected by chance, indicating low functional specificity of theses viruses (positive values for coefficients $\beta_{ecol}$)(**Figure 3**); however, functional distances among host species were generally less meaningful in describing patterns of virus sharing among pairs of host species than phylogenetic distances, as depicted by smaller effect sizes (**Figure 3**). Virus sharing among host species increased with the four proxies of sampling bias for both DNA and RNA viruses (all CIs of odds ratios 1.03 – 3.03 except for the relationships of 'Shannon diversity of publication records' ~ RNA virus sharing and 'number of publications' ~ DNA virus sharing), indicating that sampling efforts impact the topography of currently known mammal-virus networks.

### 3.3 Proportion of zoonotic viruses in different host species

We found Primates to harbour the overall largest proportions of zoonotic viruses with a group-level average of 51% (CI of 40 – 63% for respective $\mu_{order}$)(**Figure 4**), followed by slightly lower proportion of zoonotic viruses in Rodentia, Carnivora, Artiodactyla and Chiroptera (all respective $\mu_{order}$ CIs ranging between 12 – 46%) (**Figure 4**). The proportion of zoonotic viruses carried by domestic species was 1.8 times higher than in wildlife (odds ratio of 2.8 and CI of 1.8 – 4.3). RNA virus species accounted for the highest proportions of zoonotic viruses in all mammalian groups, averaging to 38% (CI of 15 – 64% according to hyperprior $H_{RNA}$) compared to only 9% (CI of 2 – 24% according to hyperprior $H_{DNA}$) of the DNA viruses in mammalian hosts being zoonotic.

We found the proportion of zoonotic RNA viruses in different host species to increase with larger range area (odds ratio of 1.06 – 1.6). In contrast, there was no evidence that the proportion of zoonotic DNA viruses in different host species was linked to any species traits (all odds ratio estimates intersecting with 1). The proportion of zoonotic RNA viruses was smaller for host species with higher Shannon diversity scores of sequence records (odds ratio

374 of 0.6 – 0.8), suggesting that more intensive sequencing efforts of a large range of these
375 viruses increases the discovery of viruses confined to non-human hosts.
376 The associations between host species from different mammalian orders and viruses from
377 different families is illustrated in Supporting Information, **Figure S4**, data are presented in
378 Supporting Information, **Table S2**.
379

380 # 4. DISCUSSION

381 Pathogen spillover and the emergence of infectious diseases ultimately depend on how
382 pathogens conquer eco-evolutionary barriers to infect novel hosts (Lloyd-Smith et al., 2009),
383 but spatiotemporal variation in species interaction and pathogen transmission opportunities
384 are proximately driven by host occurrences and community assembly (Canard et al., 2014;
385 Stephens et al., 2016). It comes therefore as little surprise that globally pervasive mammal
386 groups, such as bats and rodents, are often considered to share as many viruses with humans
387 as do primates, our closest relatives (Calisher et al., 2006; Luis et al., 2013; Olival et al.,
388 2017). Our study adds novel insights into virus spread across mammalian communities.
389 Specifically, we provide the strongest evidence to date that domestic animals are the most
390 central species in mammalian host-virus interaction networks. We also found rather
391 distinctive patterns of how DNA and RNA viruses are shared and spread among different
392 mammalian groups, with bats and carnivores being most influential in spreading RNA viruses
393 and being only of minor role in spreading DNA viruses through the network. We emphasize
394 the dominant role of domestic species in virus sharing, since domestication status strongly
395 increases the chance of virus sharing among multiple mammalian hosts. Likewise, we found
396 domestic species also to carry larger proportions of zoonotic viruses than wildlife species
397 after accounting for phylogeny and other traits.

Our study concerns the contemporary pattern of virus sharing of mammal species rather than any specific co-evolutionary histories of host switching and origin of viruses. In many, perhaps most instances, this sharing indicates the possibility of cross-species transmission, either directly via contact, or indirectly via air, soil, water, fomites or vectors. The exceptionally high virus sharing of humans and domestic animals with other mammalian species suggest that these species play a crucial role in spreading viruses, as frequent virus acquisition and dissemination is the most plausible explanation for such intensive virus sharing. This may reflect the wide geographic distribution and contact opportunities to wildlife across biogeographic borders, given that domestic species are not particularly distinguished from wildlife in terms of ecological traits. In fact, contact opportunity and community assembly have been shown in a number of studies to impact pathogen sharing and host shifting (Cooper *et al.*, 2012; Clark *et al.*, 2018; Wells & Clark, 2019). Many pathogens, including viruses, can overcome species and environmental barriers to infect distantly related hosts and disperse across large geographic areas (Longdon et al., 2014; Wells et al., 2015), although strong constraints in host shifting may also cause biogeographic structure in pathogen diversity and zoonotic disease risk (Poulin, 2010; Murray et al., 2015). Beside the large geographic ranges and diverse habitats encroached by domestic species, their large populations sizes and high densities, that often exceeds those of wildlife populations (Bar-On et al., 2018), could further contribute to host shifting and pathogen spread. This could be especially the case if large population sizes facilitate contact opportunity, virus amplification and diversification caused by more intensive within-population transmission or other factors, warranting future research.

Our findings of larger proportions of zoonotic RNA viruses compared to DNA viruses carried in different mammals is consistent with previous research (Cleaveland et al., 2001; Kreuder Johnson et al., 2015; Olival et al., 2017) and is in line with our finding that mammal species

423 generally share RNA viruses more frequently with other hosts than DNA viruses. Here, we
424 reveal for the first time that these two major groups of viruses are differently spread across
425 entire networks of mammalian hosts, an important finding that remains largely unnoticed
426 when solely looking at the species richness and propensity of zoonotic viruses carried in
427 different wildlife species. Remarkably, Chiroptera and Carnivora hold central positions in
428 terms of virus sharing with other species for RNA viruses only, whereas Ungulates hold
429 central positions for sharing both RNA and DNA viruses with other host species. In practice,
430 these findings translate into a minor role of bats and carnivores for the spread of DNA viruses
431 (and relatively low risk that DNA viruses will spillover from these species to humans). We
432 also found that cattle (*Bos taurus*), pig (*Sus scrofa*), horse (*Equus caballus*) and sheep (*Ovis*
433 *aries*), which are globally the most abundant and economically important mammalian
434 livestock species (Thornton, 2010), are among those species with the relatively highest
435 centrality measures in terms of DNA virus sharing. Importantly though, it should be noted
436 that for all these species, the frequencies of sharing DNA viruses with other host species was
437 considerably lower than sharing RNA viruses regardless of centrality measures (as is also
438 true for group-level estimates for different mammalian orders as depicted in **Figure 3**). We
439 thus emphasize that aforementioned species have a *relative* crucial role in spreading DNA
440 viruses, whereas RNA viruses generally are much more frequently shared among mammalian
441 host species. In this context, our model framework for analysing patterns in host sharing
442 provides probabilistic estimates of the variation in the pairwise phylogenetic and functional
443 similarities of infected versus uninfected host species as a signal of host specificity. This tool
444 enables us to quantify host specificity of DNA versus RNA viruses in different groups of
445 hosts, resulting in refined and community-wide measures of previously notified higher host
446 specificity in DNA viruses compared to RNA viruses (Cleaveland et al., 2001; Jackson &
447 Charleston, 2004; Geoghegan et al., 2017). Notably, the low functional host specificity of

RNA viruses exhibited by viruses shared among hosts of Primates, Carnivora, Artiodactyla and Chiroptera (i.e., functional traits of pairs of host species infected by these viruses were larger than expected by chance) emphasises their capacity to cross ecological species barriers during host shifting events despite the overall tendency to infect phylogenetically related species.

The understanding of virological factors that ensure efficient virus replication and transmission within and among host species is in its infancy (Geoghegan et al., 2016). Consequently, disentangling host or virus traits as drivers of the differential spread of DNA and RNA viruses among different mammalian orders is currently not possible and requires additional research. Possible working hypotheses as to why primates and ungulates are of relatively high central importance in sharing DNA viruses could be linked to mechanisms that enable efficient within-host virus replication and population-level transmission. At the same time, exploring virus attributes of the major DNA virus families shared among these host species, namely Herpesviridae, Papillomaviridae and Adenoviridae (Supporting Information, **Figure S4**), may help to explain why these viruses are more likely to be shared by primates and ungulates but are less likely to cross host species barrier with regards to bats and carnivores. Moreover, the strong links of some RNA viruses such as the Bunyavirales to arthropod vectors (Marklewitz et al., 2015), requires further research into the role of host-vector associations and other transmission modes for the spread of viruses.

We recognize several shortfalls in analysing database records of host-pathogen associations. First, any record of a virus species in a host entirely relies on targeted molecular screening. Certain research foci such as the boost in coronavirus research linked to bats after the SARS pandemics (Drexler et al., 2014) may include a sampling bias difficult to capture when only accounting for publication or sequencing numbers as proxies for sampling bias, since the true presence/absence of viruses in non-target host species remains unknown.

Undoubtedly, major research efforts are linked to viruses of public health relevance, while there is a dearth of systematic pathogen surveillances in wildlife (Tompkins et al., 2015). If different sampling efforts for DNA and RNA are sufficiently captured by the proxies for sampling bias is unknown and warrants future research. Second, detecting a pathogen in any targeted host species depends on its prevalence in its host population and the number of sampled host individuals but such information is not always available from collated database records. With sparse data, any direct interpretation of absolute numbers of species richness and interactions could rather reflect the observation process than true biological patterns and processes (Wells et al., 2013), and we are therefore currently not able to explore such important properties in our study. Network topologies can be also biased by sampling and data aggregation (Farine & Whitehead, 2015). We control for research effort in our analysis by accounting for variation in relation to publications and sequencing numbers, as has been done previously (Gómez et al., 2013; Olival et al., 2017). However, as more complete data from systematic disease surveillance efforts becomes available, it will be desirable to improve such analysis to better distinguish true but undiscovered interactions from 'false zeros' among other sources of bias. Compiling host-pathogen interactions from the literature and published evidence may also lead to 'false positives' such as interactions recorded from laboratory infection studies only; we minimized this error in our study by considering only interactions backed by molecular sequence records with information about sampling location in the metadata. The ongoing sophistication and broad-scale application of molecular screening methods for detecting pathogen species and identifying lineage variation may also discover unexpected and cryptic interactions among previously disconnected groups (Doña *et al.*, 2019). Finally, we are aware that amalgamating species-specific host-pathogen interactions into $N \times N$ adjacency matrix as used for some network statistics comes at the cost of losing information about pathogen species identity and thus overall connectivity of host

498 species can no longer be traced back to particular pathogen species. Overall, network
499 connectivity and modularity are therefore community-level entities, while a focus on
500 particular virus species would require more detailed analysis of underlying species-level
501 interaction matrices.

502 Our work reveals the importance of domestication status and phylogenetic clustering
503 on the importance of virus sharing among mammals, showcasing also the limited sharing of
504 DNA viruses by bats and carnivores in contrast to primates and ungulates species that readily
505 share both RNA and DNA viruses. The emergence of novel infectious diseases through
506 pathogen spillover is a hierarchical process. Ecological factors that determine the contact
507 opportunity between different host species pave the way for cross-species transmission, host
508 adaptation and subsequent within-host reproduction and transmission, which are then largely
509 controlled by ecophysiological and genetic factors. Future work that better accounts for virus
510 factors and host species community assembly may shed further light on why different types
511 of viruses spread differently among phylogenetic and functional groups of mammals and
512 foster better predictions of future disease emergence.

513

514

515 **REFERENCES**

516 Agosta, S. J., Janz, N. & Brooks, D. R. (2010) How specialists can be generalists: resolving
517 the "parasite paradox" and implications for emerging infectious disease. *Zoologia*, 27,
518 151-162. http://dx.doi.org/10.1590/S1984-46702010000200001

519 Bar-On, Y. M., Phillips, R. & Milo, R. (2018) The biomass distribution on Earth.
520 *Proceedings of the National Academy of Sciences*, 115, 6506-6511.
521 https://doi.org/10.1073/pnas.1711842115

522 Benjamini, Y. & Yekutieli, D. (2001) The control of the false discovery rate in multiple
523 testing under dependency. *The Annals of Statistics*, 29, 1165-1188.
524 https://doi.org/10.1214/aos/1013699998
525 Bonacich, P. & Lloyd, P. (2001) Eigenvector-like measures of centrality for asymmetric
526 relations. *Social Networks*, 23, 191-201. https://doi.org/10.1016/S0378-8733(01)00038-
527 7
528 Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. & Schountz, T. (2006) Bats:
529 important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews*, 19, 531-
530 545. https://doi.org/10.1128/CMR.00017-06
531 Canard, E. F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014)
532 Empirical evaluation of neutral interactions in host-parasite networks. *The American*
533 *Naturalist*, 183, 468-479. https://doi.org/10.1086/675363
534 Carroll, D., Daszak, P., Wolfe, N. D., Gao, G. F., Morel, C. M., Morzaria, S., Pablos-
535 Méndez, A., Tomori, O. & Mazet, J. A. K. (2018) The Global Virome Project. *Science*,
536 359, 872-874. https://doi.org/10.1126/science.aap7463
537 Clark, N. J., Clegg, S. M., Sam, K., Goulding, W., Koane, B. & Wells, K. (2017) Climate,
538 host phylogeny and the connectivity of host communities govern regional parasite
539 assembly. *Diversity and Distributions*, 24, 13-23. https://doi.org/10.1111/ddi.12661
540 Cleaveland, S., Laurenson, M. K. & Taylor, L. H. (2001) Diseases of humans and their
541 domestic mammals: pathogen characteristics, host range and the risk of emergence.
542 *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 356,
543 991-999. https://doi.org/10.1098/rstb.2001.0889
544 Cooper, N., Griffin, R., Franz, M., Omotayo, M. & Nunn, C. L. (2012) Phylogenetic host
545 specificity and understanding parasite sharing in primates. *Ecology Letters*, 15, 1370-
546 1377.

547 Costenbader, E. & Valente, T. W. (2003) The stability of centrality measures when networks
548 are sampled. *Social Networks*, 25, 283-307. https://doi.org/10.1111/j.1461-
549 0248.2012.01858.x
550 Doña, J., Serrano, D., Mironov, S., Montesinos-Navarro, A. & Jovani, R. (2019) Unexpected
551 bird–feather mite associations revealed by DNA metabarcoding uncovers a dynamic
552 ecoevolutionary scenario. *Molecular Ecology*, **28**, 379-390.
553 https://doi.org/10.1111/mec.14968
554 Dray, S. & Dufour, A.-B. (2007) The ade4 Package: implementing the duality diagram for
555 ecologists. *Journal of Statistical Software*, 22, 20. https://doi.org/10.18637/jss.v022.i04
556 Drexler, J. F., Corman, V. M. & Drosten, C. (2014) Ecology, evolution and classification of
557 bat coronaviruses in the aftermath of SARS. *Antiviral Research*, 101, 45-56.
558 https://doi.org/10.1016/j.antiviral.2013.10.013
559 Farine, D. R. & Whitehead, H. (2015) Constructing, conducting and interpreting animal
560 social network analysis. *Journal of Animal Ecology*, 84, 1144-1163.
561 https://doi.org/10.1111/1365-2656.12418
562 Fenton, A., Streicker, D. G., Petchey, O. L. & Pedersen, A. B. (2015) Are all hosts created
563 equal? Partitioning host species contributions to parasite persistence in multihost
564 communities. *The American Naturalist*, 186, 610-622. https://doi.org/10.1086/683173
565 Fritz, S. A., Bininda-Emonds, O. R. P. & Purvis, A. (2009) Geographical variation in
566 predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology*
567 *Letters*, 12, 538-549. https://doi.org/10.1111/j.1461-0248.2009.01307.x
568 Geoghegan, J. L., Duchêne, S. & Holmes, E. C. (2017) Comparative analysis estimates the
569 relative frequencies of co-divergence and cross-species transmission within viral
570 families. *PLoS Pathogens*, 13, e1006215. https://doi.org/10.1371/journal.ppat.1006215

571 Geoghegan, J. L., Senior, A. M., Di Giallonardo, F. & Holmes, E. C. (2016) Virological
572 factors that increase the transmissibility of emerging human viruses. *Proceedings of the*
573 *National Academy of Sciences*, 113, 4170-4175.
574 https://doi.org/10.1073/pnas.1521582113
575 Gómez, J. M., Nunn, C. L. & Verdú, M. (2013) Centrality in primate–parasite networks
576 reveals the potential for the transmission of emerging infectious diseases to humans.
577 *Proceedings of the National Academy of Sciences*, 110, 7738-7741.
578 https://doi.org/10.1073/pnas.1220716110
579 Gower, J. C. (1971) A general coefficient of similarity and some of its properties. *Biometrics*,
580 27, 857-871.
581 Han, B. A., Schmidt, J. P., Bowden, S. E. & Drake, J. M. (2015) Rodent reservoirs of future
582 zoonotic diseases. *Proceedings of the National Academy of Sciences*, 112, 7039-7044.
583 https://doi.org/10.1073/pnas.1501598112
584 Jackson, A. P. & Charleston, M. A. (2004) A cophylogenetic perspective of RNA-virus
585 evolution. *Molecular Biology and Evolution*, 21, 45-57.
586 https://doi.org/10.1093/molbev/msg232
587 Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L. & Daszak,
588 P. (2008) Global trends in emerging infectious diseases. *Nature*, 451, 990-994.
589 https://doi.org/10.1038/nature06536
590 Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K.,
591 Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M.J., Foster, J. K.,
592 Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., Teacher, A.,
593 Bininda-Emonds, O. R. P., Gittleman, J. L., Mace, G. M., Purvis, A. & Michener, W. K.
594 (2009) PanTHERIA: a species-level database of life history, ecology, and geography of

595 extant and recently extinct mammals. *Ecology*, 90, 2648-2648.
596 https://doi.org/10.1890/08-1494.1
597 Karesh, W. B., Dobson, A., Lloyd-Smith, J. O., Lubroth, J., Dixon, M. A., Bennett, M.,
598 Aldrich, S., Harrington, T., Formenty, P., Loh, E. H., Machalaba, C. C., Thomas, M. J.
599 & Heymann, D. L. (2012) Ecology of zoonoses: natural and unnatural histories. *The*
600 *Lancet*, 380, 1936-1945.
601 Kreuder Johnson, C., Hitchens, P. L., Smiley Evans, T., Goldstein, T., Thomas, K., Clements,
602 A., Joly, D. O., Wolfe, N. D., Daszak, P., Karesh, W. B. & Mazet, J. K. (2015) Spillover
603 and pandemic properties of zoonotic viruses with high host plasticity. *Scientific Reports*,
604 5, 14830. https://doi.org/10.1016/S0140-6736(12)61678-X
605 Little, R. J. (2004) To model or not to model? Competing modes of inference for finite
606 population sampling. *Journal of the American Statistical Association*, 99, 546-556.
607 https://doi.org/10.2307/27590409
608 Lloyd-Smith, J. O., George, D., Pepin, K. M., Pitzer, V. E., Pulliam, J. R. C., Dobson, A. P.,
609 Hudson, P. J. & Grenfell, B. T. (2009) Epidemic dynamics at the human-animal
610 interface. *Science*, 326, 1362-1367. https://doi.org/10.1126/science.1177345
611 Longdon, B., Brockhurst, M. A., Russell, C. A., Welch, J. J. & Jiggins, F. M. (2014) The
612 evolution and genetics of virus host shifts. *PLoS Pathogens*, 10, e1004395.
613 https://doi.org/10.1371/journal.ppat.1004395
614 Longdon, B., Day, J.P., Alves, J. M., Smith, S. C. L., Houslay, T. M., McGonigle, J. E.,
615 Tagliaferri, L. & Jiggins, F. M. (2018) Host shifts result in parallel genetic changes
616 when viruses evolve in closely related species. *PLoS Pathogens*, 14, e1006951.
617 https://doi.org/10.1371/journal.ppat.1006951
618 Luis, A. D., O'Shea, T. J., Hayman, D. T. S., Wood, J. L. N., Cunningham, A. A., Gilbert, A.
619 T., Mills, J. N. & Webb, C. T. (2015) Network analysis of host–virus communities in

620 bats and rodents reveals determinants of cross-species transmission. *Ecology Letters*, 18,
621 1153–1162. https://doi.org/10.1111/ele.12491
622 Luis, A. D., Hayman, D. T. S., Shea, T. J., Cryan, P. M., Gilbert, A. T., Pulliam, J. R. C.,
623 Mills, J. N., Timonin, M. E., Willis, C. K. R., Cunningham, A. A., Fooks, A. R.,
624 Rupprecht, C. E., Wood, J. L. N. & Webb, C. T. (2013) A comparison of bats and
625 rodents as reservoirs of zoonotic viruses: are bats special? *Proceedings of the Royal*
626 *Society of London B: Biological Sciences*, 280, 20122753.
627 https://doi.org/10.1098/rspb.2012.2753
628 Magurran, A. E. (2004) Measuring biological diversity. Blackwell, Oxford.
629 Marklewitz, M., Zirkel, F., Kurth, A., Drosten, C. & Junglen, S. (2015) Evolutionary and
630 phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA
631 virus family. *Proceedings of the National Academy of Sciences*, 112, 7536-7541.
632 https://doi.org/10.1073/pnas.1502036112
633 Morand, S., McIntyre, K. M. & Baylis, M. (2014) Domesticated animals and human
634 infectious diseases of zoonotic origins: domestication time matters. *Infection, Genetics*
635 *and Evolution*, 24, 76-81. https://doi.org/10.1016/j.meegid.2014.02.013
636 Murray, K. A., Preston, N., Allen, T., Zambrana-Torrelio, C., Hosseini, P. R. & Daszak, P.
637 (2015) Global biogeography of human infectious diseases. *Proceedings of the National*
638 *Academy of Sciences*, 112, 12746-12751. https://doi.org/10.1073/pnas.1507442112
639 Nylin, S., Agosta, S., Bensch, S., Boeger, W.A., Braga, M.P., Brooks, D.R., Forister, M.L.,
640 Hambäck, P.A., Hoberg, E.P., Nyman, T., Schäpers, A., Stigall, A.L., Wheat, C.W.,
641 Österling, M. & Janz, N. (2018) Embracing colonizations: a new paradigm for species
642 association dynamics. *Trends in Ecology & Evolution*, 33, 4-14.
643 https://doi.org/10.1016/j.tree.2017.10.005

644 Olival, K. J., Hosseini, P. R., Zambrana-Torrelio, C., Ross, N., Bogich, T. L. & Daszak, P.
645 (2017) Host and viral traits predict zoonotic spillover from mammals. *Nature*, 546, 646.
646 https://doi.org/10.1038/nature22975
647 Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution
648 in R language. *Bioinformatics*, 20, 289-290.
649 https://doi.org/10.1093/bioinformatics/btg412
650 Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E. C., Burke, D. S., Calisher, C. H.,
651 Laughlin, C. A., Saif, L. J. & Daszak, P. (2008) Cross-species virus transmission and
652 the emergence of new epidemic diseases. *Microbiology and Molecular Biology*
653 *Reviews*, 72, 457-470. https://doi.org/10.1128/mmbr.00004-08
654 Pavoine, S., Vallet, J., Dufour, A.-B., Gachet, S. & Daniel, H. (2009) On the challenge of
655 treating various types of variables: application for improving the measurement of
656 functional diversity. *Oikos*, 118, 391-402. https://doi.org/10.1111/j.1600-
657 0706.2008.16668.x
658 Plummer, M. (2016) rjags: Bayesian graphical models using MCMC. R package version 4-6.
659 Poulin, R. (2010) Decay of similarity with host phylogenetic distance in parasite faunas.
660 *Parasitology*, 137, 733-741. https://doi.org/10.1017/s0031182009991491
661 R Development Core Team (2017) R: A language and environment for statistical computing.
662 R Foundation for Statistical Computing. https://cran.r-project.org/
663 Stephens, P. R., Altizer, S., Smith, K. F., Alonso Aguirre, A., Brown, J. H., Budischak, S. A.,
664 Byers, J. E., Dallas, T. A., Jonathan Davies, T., Drake, J. M., Ezenwa, V. O., Farrell, M.
665 J., Gittleman, J. L., Han, B. A., Huang, S., Hutchinson, R. A., Johnson, P., Nunn, C. L.,
666 Onstad, D., Park, A., Vazquez-Prokopec, G. M., Schmidt, J. P. & Poulin, R. (2016) The
667 macroecology of infectious diseases: a new perspective on global-scale drivers of

668 pathogen distributions and impacts. *Ecology Letters*, 19, 1159–1171.
669 https://doi.org/10.1111/ele.12644
670 Thornton, P. K. (2010) Livestock production: recent trends, future prospects. *Philosophical*
671 *Transactions of the Royal Society B: Biological Sciences*, 365, 2853-2867.
672 https://doi.org/10.1098/rstb.2010.0134
673 Tompkins, D. M., Carver, S., Jones, M. E., Krkošek, M. & Skerratt, L. F. (2015) Emerging
674 infectious diseases of wildlife: a critical perspective. Trends in Parasitology, 31, 149-
675 159. https://doi.org/10.1016/j.pt.2015.01.007
676 Wardeh, M., Risley, C., McIntyre, M. K., Setzkorn, C. & Baylis, M. (2015) Database of host-
677 pathogen and related species interactions, and their global distribution. *Scientific Data*,
678 2, 150049. https://doi.org/10.1038/sdata.2015.49
679 Wells, K., O'Hara, R. B., Morand, S., Lessard, J.-P. & Ribas, A. (2015) The importance of
680 parasite geography and spillover effects for global patterns of host–parasite associations
681 in two invasive species. *Diversity and Distributions*, 21, 477-486.
682 https://doi.org/10.1111/ddi.12297
683 Wells, K., O'Hara, R. B., Pfeiffer, M., Lakim, M. B., Petney, T. N. & Durden, L. A. (2013)
684 Inferring host specificity and network formation through agent-based models: tick–
685 mammal interactions in Borneo. *Oecologia*, 172, 307-316.
686 https://doi.org/10.1007/s00442-012-2511-9
687 Wells, K., Gibson, D. I., Clark, N. J., Ribas, A., Morand, S. & McCallum, H. I. (2018) Global
688 spread of helminth parasites at the human–domestic animal–wildlife interface. *Global*
689 *Change Biology*, 24, 3254-3265. https://doi.org/10.1111/GCB.14064
690 Wells, K. & Clark, N.J. (2019) Host specificity in variable environments. *Trends in*
691 *Parasitology*, 6, 452-465. https://doi.org/10.1016/j.pt.2019.04.001

692 Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M. & Jetz, W. (2014)
693 EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals.
694 *Ecology*, 95, 2027-2027. https://doi.org/10.1890/13-1917.1
695 Wilson, D. E. & Reeder, D. M. (2005) Mammal species of the world. A taxonomic and
696 geographic reference (3rd ed.). Johns Hopkins University Press, Baltimore.
697 Woolhouse, M. E. J., Haydon, D. T. & Antia, R. (2005) Emerging pathogens: the
698 epidemiology and evolution of species jumps. *Trends in Ecology & Evolution*, 20, 238-
699 244. https://doi.org/10.1016/j.tree.2005.02.009
700
701

## 702 DATA ACCESSIBILITY STATEMENT

703 The data reported in this paper will be deposited at Dryad (*https://datadryad.org/*).
704

## 705 ORCID

706 *Konstans Wells* https://orcid.org/0000-0003-0377-2463
707 *Serge Morand* https://orcid.org/0000-0003-3986-7659
708 *Maya Wardeh* https://orcid.org/0000-0002-2316-5460
709 *Matthew Baylis* https://orcid.org/0000-0003-0335-187X
710

## 711 BIOSKETCH

712 **Konstans Wells** is a wildlife and disease ecologist researching on the consequences of
713 environmental change on biodiversity, wildlife, invasive species and host-parasite
714 interactions. The research team includes ecologists and epidemiologist who work at the
715 interface of biodiversity and health in times of global change.
716

# SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1**. Eigenvector centrality measures for DNA and RNA virus sharing.

**Figure S2**. Effects of sampling bias on eigenvector centrality measures.

**Figure S3**. Host species with highest centrality measures.

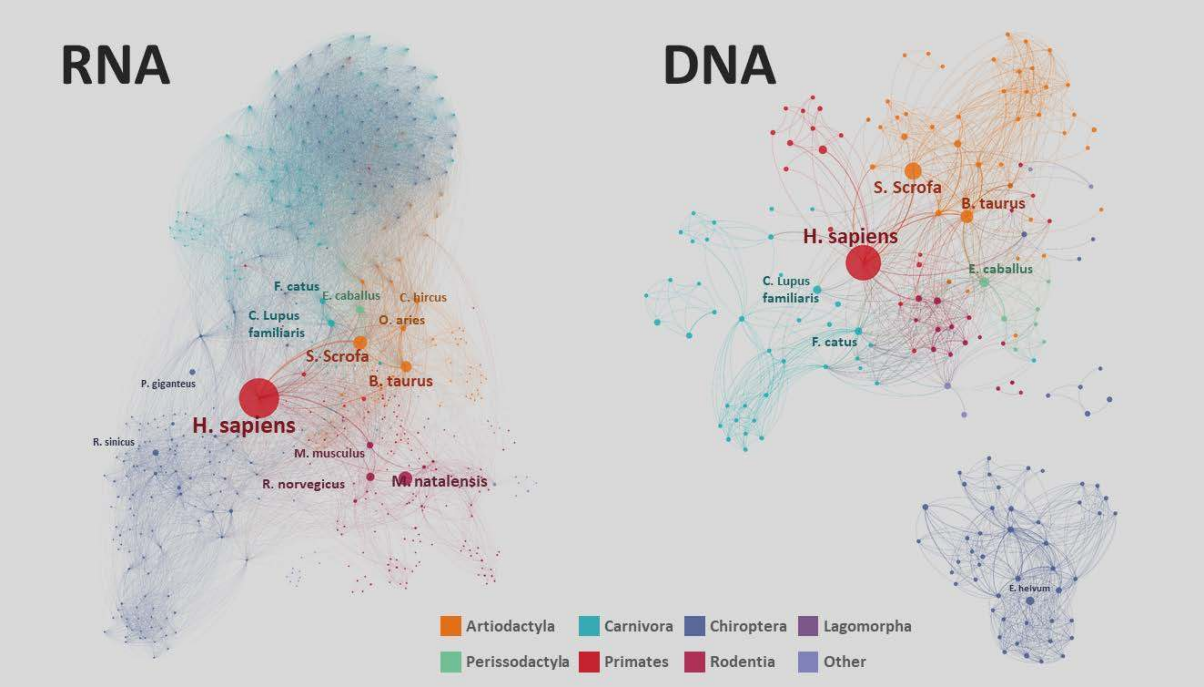
**Figure S4**. Network plot of mammal-virus associations.

**Table S1**. Test statistics for multiple pairwise comparison of centrality measures.

**Table S2**. Table of recorded mammal-virus interactions.

**Box S1**. Model code for modelling virus sharing among hosts.

**Box S2**. Model code for modelling proportion of zoonotic viruses.




735

736 **Figure 1**. Network plots of the sharing of RNA (left) and DNA viruses (right) among

737 mammalian host species. Each node represents a mammal species (total of n=725 species).

738 The size of the node depicts the number of virus species shared with other mammalian host

739 species, the width of edges is plotted proportional to the number of virus species shared

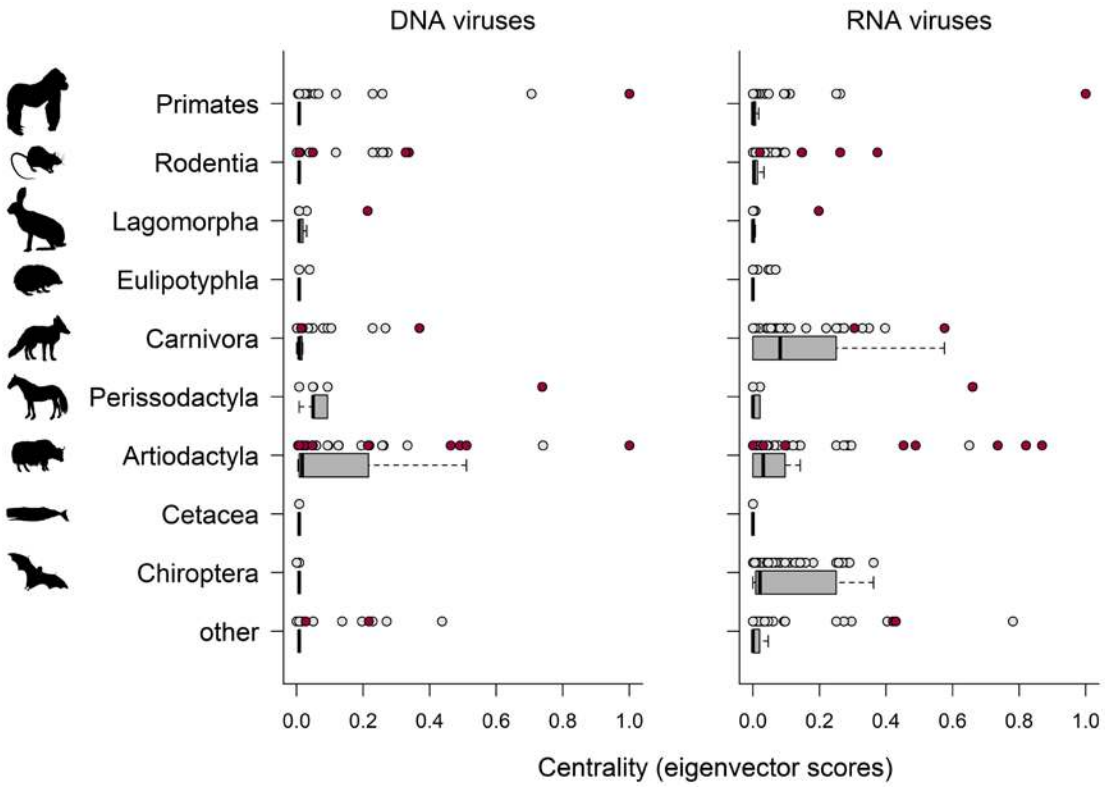
740 between pairs of hosts. Colour depict the different mammalian orders.

741

742

743

744




745

746 **Figure 2**. Eigenvector centrality measures (box plots and species data points) of host species
747 from different mammalian orders, depicting their relative importance in virus sharing and
748 spread across networks for DNA viruses (left panel) and RNA viruses (right panel). Larger
749 values refer to host species sharing more viruses with others, especially with host species that
750 are also well connected. Artiodactyla and Cetacea are presented as separate groups because of
751 their distinct terrestrial/marine habitats, mammalian orders with few species are merged into
752 the group 'other'. Grey points represent measures for wild and red points measures for
753 domestic mammalian host species and humans.

754

755

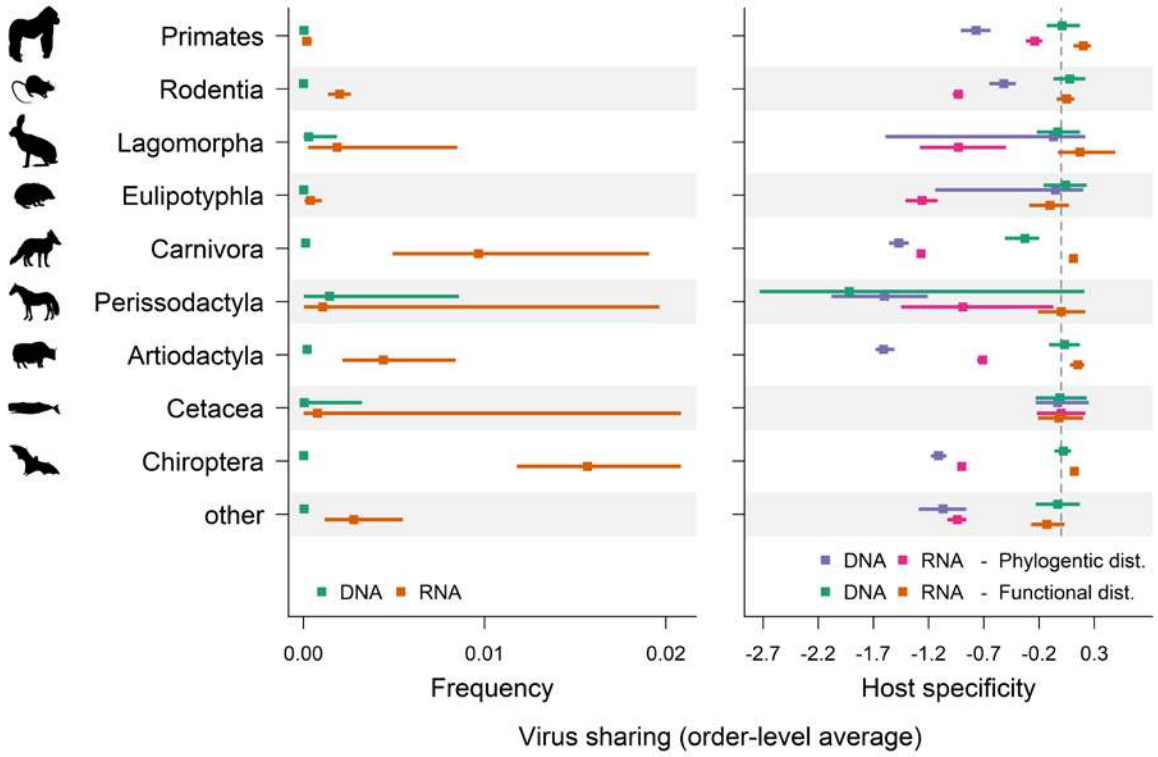



**Figure 3**. Order-level estimates of the average frequency mammalian species of the respective order share any of its associated viruses with another mammalian host species (left panel; parameter $H_{\eta}$*(order)* in model description). The right panel shows the relative extent of host specificity in virus sharing in terms of the relative difference between observed and expected phylogenetic and functional diversity of mammalian host species as estimated from regression coefficients. Values < 0 indicate pairs of infected hosts were more phylogenetically/functionally similar than expected based on random draws from regional mammalian species pools, indicating higher specificity in virus spread among mammalian species (corresponding to parameters *βphyl* and *βecol* in model description). All estimates are presented for the two subsets of DNA and RNA viruses. Boxes are posterior estimates and bars represent 95% credible intervals.

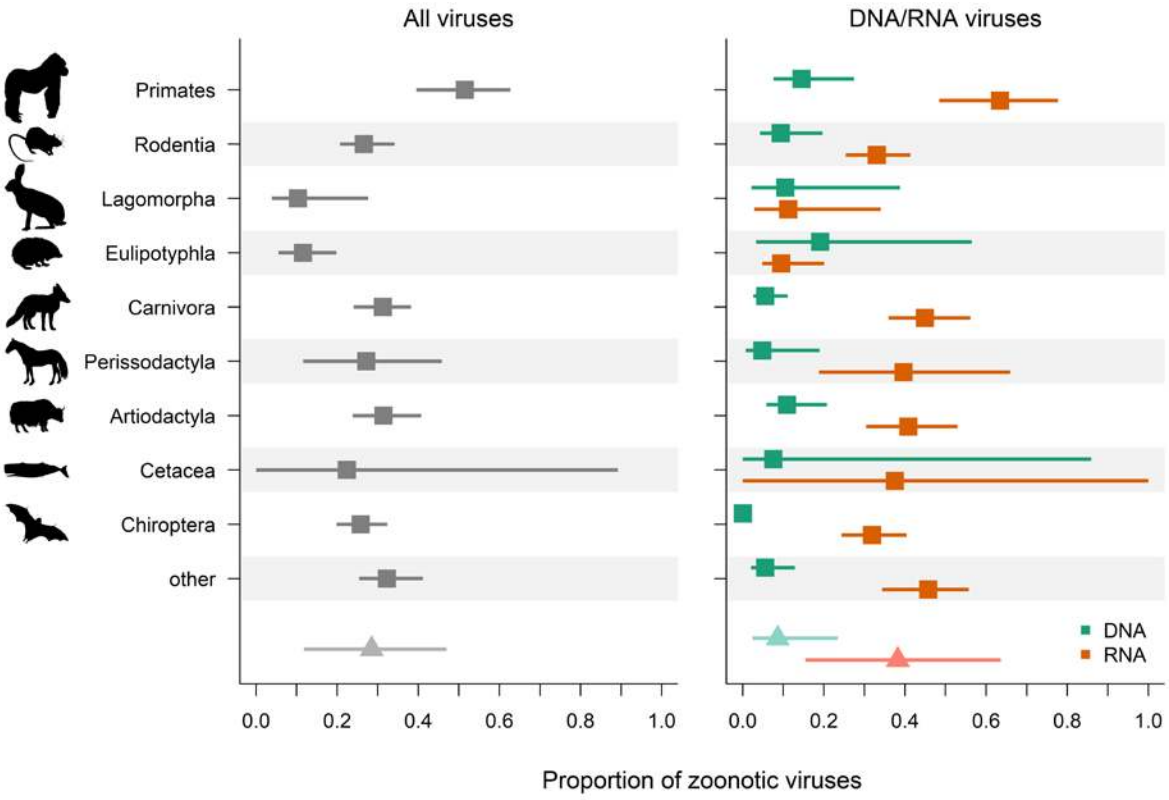



**Figure 4**. Estimated proportion of zoonotic viruses for mammalian host species from different orders (left panel: all n= 1,785 virus species in the database, right panel: estimates for the two main groups of n=730 DNA virus species and n=912 RNA virus species). Estimates represent the group-level averages ('hyperprior') from a Bayesian hierarchical model. The group "other" assembles all species from orders with < 9 species in the dataset. Boxes are posterior estimates and bars represent 95% credible intervals. The grey triangle and bar represent the overall average estimate according to a second-level hyperprior in the Bayesian model hierarchy.