Characterizing parasite generalism illuminates patterns of host-parasite associations

Park, A.W.

Farrell, M.J.

Schmidt, J.P.

Huang, S.

Dallas, T.A.

Pappalardo, P.

Drake, J.M.

Stephens, P.R.

Poulin, R.

Nunn, C.L.

Davies, T.J.

## Abstract

The distribution of parasites among mammalian hosts is complex and represents a differential ability or opportunity to infect different host species. Here, we take a macroecological approach to understanding why some parasites show a tendency to infect species widely distributed in the host phylogeny (phylogenetic generalism) while others are constrained to closely related hosts. Using a database on over 1400 parasite species that have been documented to infect between 1-69 terrestrial mammal host species, we characterize the phylogenetic generalism of parasites using standard effect sizes for three metrics: mean pairwise phylogenetic distance (PD), maximum PD, and phylogenetic evenness. We find that the majority of multi-host parasites are not phylogenetic specialists, with host relatedness being equivalent to that expected from a random sample of host species. Bacteria and arthropod parasites are typically the most generalist, viruses and helminths are intermediate, and protozoa are on average the most specialist. While viruses and helminths have similar mean pairwise PD on average, the viruses exhibit higher variation as a group. Close contact transmission is the transmission mode most associated with specialism. Most parasites with low phylogenetic evenness (associating with discrete groups of species in the host phylogeny) are helminths and viruses.

## Keywords

Parasite, multi-host, generalism, transmission mode, phylogenetic, macroecology

## Introduction

Parasites that can infect a wide range of host species, i.e., generalist parasites, are often highlighted as threats to biodiversity conservation and public health due to their widespread impact and likelihood of emergence in novel hosts [1–5]. Conventionally, the generalism (conversely, host specificity) of parasites is quantified by the taxonomic breadth of their host species, for instance number of host species [6–8]. Theoretical and empirical studies using such generalism metrics have provided important insights into the ecology and evolution of parasites, but have been less successful in identifying which host species are at greatest risk of future emergent infectious diseases, and which parasites are likely to cause those outbreaks. One of challenges for prediction is to take into account how host species differ in their susceptibility to the same parasite [8,9]. The rapid development in integrating information on evolutionary history into biodiversity research [10,11] has inspired recent research aimed at shining evolutionary perspective on the mechanisms underlying parasite generalism and for identifying unknown existing or potential host species [9,12–14]. For example, mammalian parasites capable of infecting multiple host species from disparate evolutionary lineages (i.e., distant relatives in a phylogeny) are more likely to cause zoonotic disease outbreaks than others [15]. Here, we take this new approach of quantifying the phylogenetic relatedness among host species infected by the same parasite, i.e., a parasite’s phylogenetic generalism (or phylogenetic host specificity) to identify the factors that determine whether parasites can spread across host lineages. We also present a novel index to characterize the spread of a parasite on the host phylogeny. For example, parasites may exhibit patterns consistent with creeping through the phylogeny via related hosts, taking large leaps across host clades, or both.

The distribution of parasites across host species has previously been shown to depend on the phylogenetic relatedness among host species [16–19]. Closely related host species not only share common evolutionary history, and thus common parasites, at least until divergence in the more recent past [20,21], they also show higher similarity in their biology than distantly related hosts [22–24], including characteristics associated with immune responses, and behaviors that can modify potential for host switching [25–27]. However, two recent studies, using relatively restricted datasets on parasites in different mammalian clades, showed large variation in the degree of phylogenetic generalism. Most parasites infecting multiple primate hosts also appeared to be phylogenetic generalists [15], but for carnivores, many multi-host parasites were constrained by host phylogeny, especially helminths and viruses [17]. Earlier work has also shown that many helminth species tend to exhibit taxonomically restricted host ranges [7,9,28] but suggested that viruses, due to their rapid mutation rates, should more readily adapt to new hosts that are not necessarily closely related to existing hosts [29–31]. These mixed results presented in different systems invite a broader-scale investigation on the question: why do some parasites overcome the physiological and ecological barriers between distantly related host species, while others are localized within a host phylogenetic tree?

We address this by analyzing a large amount of data describing mammalian host-parasite associations. Approximately half of the parasite species studied (781/1490) infect at least two host species from four mammalian orders: Artiodactyla, Carnivora, Perissodactyla and Primates. We compare the phylogenetic generalism of these parasite species in relation to two main factors: the higher-taxon parasite group (i.e., arthropods, bacteria, helminths, protozoa and viruses) and the transmission modes of each parasite species (i.e., close-contact transmission, complex life-cycle transmission, environmental transmission, and vector-borne transmission). Our aim is to identify patterns of parasite sharing by different hosts in relation to host phylogeny, and the principles that govern these patterns. Our study is unparalleled in terms of the diversity of host and parasite species considered. This breadth is made possible by the newly published Global Mammal Parasite Database (GMPD) 2.0 [32], which includes an extensive record of parasite occurrence in free-ranging populations of over 400 mammal species from the focal orders, coupled with further information on the taxonomy and characteristics of the parasite species. In addition to the host range of multi-host parasite species, we also examine a commonly neglected component of ecological parasitology -- the number of single-host parasites in relation to parasite taxonomy and transmission mode, as well as the position of their host in the phylogeny.

## Materials and Methods

Records of parasite associations with terrestrial mammals were obtained from the GMPD [32,33]. These records include Latin binomials and taxonomic classification for host and parasite species, and transmission mode for the majority of parasite species (80%). Transmission modes were assigned in the GMPD based on an extensive literature review [32,33], and comprise: close contact, complex life-cycle (i.e., with intermediate hosts transitioning parasites to definitive hosts, including via predator-prey interactions), environmental, and vector-borne transmission. The GMPD data were analyzed in R [34] to establish the number and identity of parasite species per host species and the number and identity of host species per parasite species. The analysis includes 404 terrestrial mammal host species. Because Perissodactyla is under-represented, with only 10 species compared to 115, 178 and 101 for Carnivora, Primates and Artiodactyla, respectively, we combine this order with Artiodactyla to form an ungulate group consisting of 111 species.

A published phylogeny of mammals [35,36] was used to obtain the phylogenetic distance (PD) between all pairs of hosts. Although more recent mammal phylogenies have been published, we use this version because it matches to the mammal taxonomy in the GMPD. Previous work has demonstrated that pairwise distance metrics, which we use here, are not generally biased by small difference in phylogenetic topology [37], and thus we believe our analyses are robust. For the set of hosts of each parasite species, the mean pairwise PD between hosts was calculated, as was its standard effect size [38]. The latter was obtained using the R package *picante* [39], under a null model in which the community data matrix (host-parasite associations) was randomized with the independent swap algorithm [40], maintaining occurrence frequencies of host and parasite species, and weighting the PD of selected host pairs based on their abundance in the GMPD. This measure captures the average relatedness of the host species of a given parasite species on a standardized scale, allowing for easier comparison among parasites. Two further standard effect size metrics were calculated on each parasite species' host set assuming the same null: maximum PD of any two hosts in a parasite's set, and the ratio of the mean minimum (nearest host) PD to the maximum PD [41]. The former metric provides a standardized measure of each parasite's breadth across the host phylogenetic tree (hereafter termed "span"). The latter metric provides a measure of the tendency for the set of host species to exhibit an aggregated or clumped distribution within the span (hereafter termed "aggregation") and is achieved by calculating the PD of each host species to its nearest host, averaging this across all hosts and scaling by the span. The standard effect size calculations return a z-score and a p-value, where z-scores below -1.96 (phylogenetically specialist parasites) are typically significantly different from the null expectation, assuming the null model generates a normally-distributed set of scores [42]. Scores in the interval -1.96<z<+1.96 are equivalent to a random host set under the null model, and are therefore consistent with relatively generalist parasites. While z-scores greater than +1.96 are theoretically possible, they do not occur under most null models. Consequently, we use the term generalist to describe parasites that infect host species whose relatedness is equivalent to a random host set under the null model. Host evolutionary distinctiveness, measured as millions of years of evolutionary separation, was estimated directly from the mammal phylogeny [35,36] using the *evol.distinct* function in the R package *picante* [39], with the *equal splits* option in which shared branches are apportioned equally among descendant lineages [43]. Each hosts's terminal branch length was also recorded as an additional measure of evolutionary isolation.

## Results

Host communities of a parasite exhibited a bias towards being more related than expected by chance, indicated by negative z-scores for mean pairwise PD between host species (Fig. 1, global median standard effect size=-0.93). However, while 96% of parasites have negative standard effect sizes, only 12.4% of parasites have significantly negative standard effect sizes. No parasites were found to have significantly positive standard effect sizes.

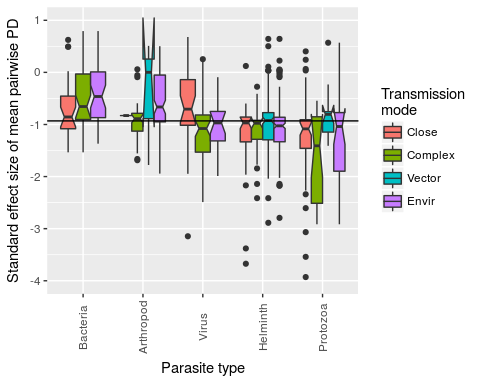


Figure 1. Boxplot of standard effect sizes of mean pairwise PD in each parasite group stratified by transmission mode. Negative values indicate that a parasite species tends to infect host species that are more closely related than expected by chance (under the null model). Conversely, positive values indicate the opposite. In terms of deviation from the null model, 12.4% of standard effect sizes are significantly negative (none of the positive values are statistically significant). Solid horizontal line indicates the global median standard effect size (-0.93).

The proportion of these significantly negative standard effect sizes varies across parasite type, with smallest to largest proportions corresponding to bacteria, arthropods, helminths, viruses and protozoa, respectively (test for equality of proportions, p<0.001, Fig. 2 top panel). The proportion of significantly negative standard effect sizes is not significantly different across parasite transmission modes (test for equality of proportions, p=0.65, Fig. 2 bottom panel). However, across the full range of standard effect sizes, each of the two covariates (parasite type and transmission mode) and their interaction explain significant differences between mean values (ANOVA: p<0.001 for parasite type, transmission mode and interaction, details in Table S1, supplementary material). Additionally, a less conservative null model that ignores the frequency of host records (unweighted) in the GMPD was compared with the weighted null model used here (Fig. S1, supplementary material). In general, standard effect sizes are highly correlated across the two null models, with more extreme negative effect sizes observed in the unweighted model, and viruses and bacteria being the parasite types most likely to be re-classified as phylogenetic specialists in an unweighted null model (Fig. S1, supplementary material). Further, excluding parasite species with only one known host species does not bias results as the frequency of parasite types and transmission modes is congruent across single- and multi-host parasites (Fig. S2, supplementary material).

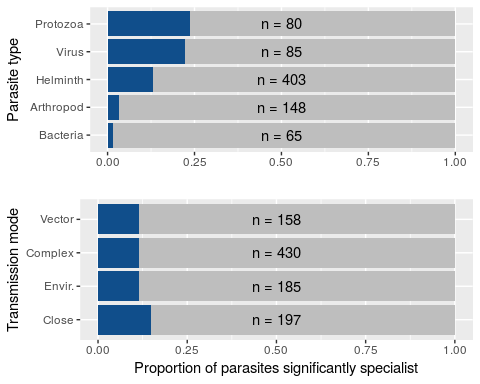


Figure 2. Proportion of parasites (blue bars) with sets of host species that are more related than expected by chance, grouped by parasite type (top panel) and transmission mode (bottom panel). Sample size () for each group is shown for each group.

Parasite type by transmission mode interactions are particularly driven by protozoa and viruses, and by close contact and vector-borne transmission (Fig. S3, supplementary material); close contact transmission is associated with atypical extreme specialism in these groups and vector-borne transmission is associated with atypical generalism in these same groups. Parasites that infect hosts only within one group (carnivores, primates or ungulates) exhibit patterns of host relatedness that are qualitativey similar to the overall pattern (Figs. S4 & S5, supplmentary material). Departures from the general pattern include the observation of several specialist bacteria infecting carnivores (Fig. S4, supplementary material) and the relative lack of importance of transmission mode on patterns of specialization for parasites infecting primates and ungulates (Fig. S5, supplementary material).

Helminths and viruses are the two parasite types whose average standard effect sizes for mean pairwise PD are intermediate among the five parasite types, and their mean values are very close to each other. However, it is notable that viruses exhibit much wider variation in degree of generalism (Fig. 3), containing many extreme specialists (large, negative standard effect sizes) but also many relative generalists. By contrast, helminths tend to be more consistent in their degree of generalism, and contain relatively few species that are at either extreme of the generalism-specialism continuum. This illustration underscores the importance of examining the variance in phylogenetic specialism of a parasite type in addition to its average level.

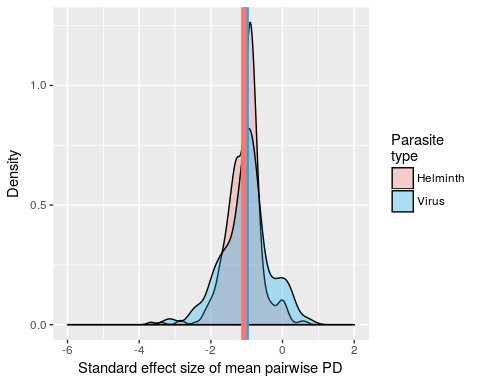


Figure 3. Probability density function of standard effect size of mean pairwise PD for helminths (red) and viruses (blue). The mean standard effect size is very close between the two types (group mean values denoted by color-coded vertical lines), but the virus group contains more extreme parasites in terms of both specialism and generalism. The complete density plot for all five parasite types is presented in the supplementary material (Fig. S6, supplementary material).

The standard effect size of maximum PD in a parasite's host set provides a comparative measure of the span that a parasite exhibits across the host phylogenetic tree. The standard effect size of the ratio of mean minimum to maximum PD measures the tendency for hosts of a parasite to aggregate in the mammal phylogeny. We found that parasite taxonomic groups exhibit variation in both of these metrics. Protozoa, viruses and helminths contain several parasite species whose span is significantly smaller than expected by chance (Fig. 4 points left of boxes in subplots). No parasite has a span that is bigger than expected by chance, which is not surprising since the spans associated with random host species selected in the null model are frequently large. Viruses and protozoa additionally contain some species whose hosts are more aggregated than expected by chance (Fig. 4, points below boxes in subplots). In contrast to parasite type, patterns of span and aggregation are similar across transmission modes (Fig. S7, supplementary material). An example illustration of two parasites, *Leptospira interrogans* and *Trypanosoma cruzi*, with similar span but very different aggregation patterns is provided in Fig. 5. The raw values for each metric of phylogenetic specificity, i.e., before applying standard effect size calculations, are comparable across all parasite types, with much overlap in values (Fig. S8, supplementary material). Helminths and protozoa exhibit somewhat smaller phylogenetic distances between host pairs than other parasite types, which may reflect either constraints to parasitism or differences in the phylogenetic structure of the clades of helminth and protozoan hosts versus hosts of other parasite types.

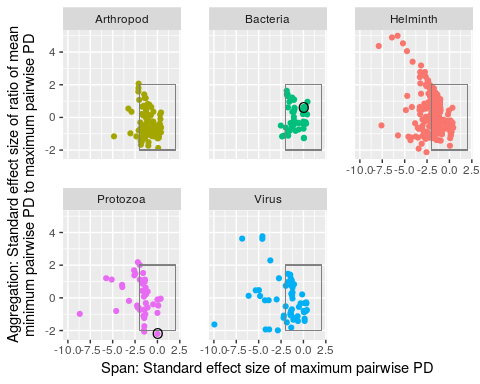


Figure 4. Scatter plot showing both span and aggregation of parasites by type. The x-axes provide a standard effect size of maximum pairwise PD of a parasite's host set (span) and the y-axes provide a standard effect size of the ratio of mean minimum pairwise PD to maximum pariwise PD (aggregation). Circled bacterial and protozoan parasites are *Leptospira interrogans* and *Trypanosoma cruzi*. These parasites have a similar span (), but very different host aggregation patterns (*L. interrogans* exhibiting a random distribution of host species, *T. cruzi* exhibiting an aggregated distribution of host species, illustrated further in Fig. 5). Boxes encompassing contain parasites whose standard effect sizes are typical of those expected from a random set of hosts under the null model. Negative standard effect sizes lower than -1.96 are typical of host sets exhibiting significantly small span (x-axis) and aggregated distribution (y-axis).

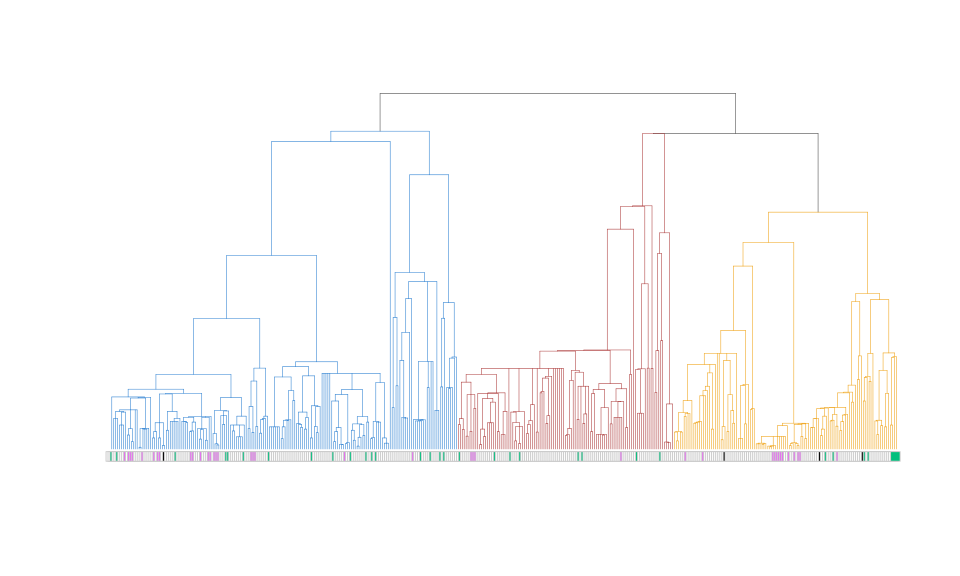


Figure 5. Examples of aggregated (purple) and random (green) distributions of parasites of similar span in the host phylogeny. The parasites are *L. interrogans* (green) and *T. cruzi* (purple), colored according to their groupings in Fig. 4. Black coloring indicates host species infected by both parasite species. The clade colors indicate primates (blue), ungulates (red), and carnivores (gold).

To investigate whether the evolutionary isolation of hosts influences host-parasite associations, we constructed a set of negative binomial generalized linear models, to accommodate over-dispersion in the number of parasite species per host species (variance to mean ratio ~25), meaning that a small number of host species are associated with a large number of parasite species. These models have the number of parasite species per host species as the response variable, and either host evolutionary distinctiveness or host terminal branch length as the main predictor variable (Fig. S9, supplementary material). Each model was fitted with and without the inclusion of a second predictor variable, the number of host records in the GMPD, which was used to control for sampling bias. Only one model showed that more evolutionarily distinct hosts (but not those subtending from longer terminal branches, and not in models that included the number of host records) had significantly fewer parasite species (Table S2, supplementary material). Even ignoring the indication that these patterns appear to be largely driven by the number of host records (a measure of host sampling bias that was significant in all models in which it was applied), host evolutionary distinctiveness was not strongly related to parasite species richness; one unit increase in evolutionary distinctiveness (millions of years of evolution) was associated with the reduction in parasite species richness by approximately one parasite species (supplementary material).

## Discussion

We have shown that most multi-host parasites are relatively unconstrained by the phylogenetic distance separating hosts, but among the 781 multi-host parasite species considered here, there is nonetheless a general tendency for parasites to infect mammalian host species that are comparatively closely related. Of the five parasite types, bacteria are the most generalist, while protozoa and viruses are more specalists. Transmission mode impacts the opportunity for parasites to encounter novel host species, for example via low intersepcific host contact rates, vector feeding preferences, disparate trophic links associated with complex life-cycle transmission and low parasite survivorship in the environment. However, transmission mode was less influential in determining parasite generalism compared with parasite taxonomy, where taxonomy likely captures numerous biological traits for which data is currently lacking, such as mutation rates and immunogenicity. Additionally, there were interactions between parasite type and transmission mode; both protozoan and viral parasites exhibit specialism when transmission is through close contact, in agreement with primate parasite research that used taxonomy as a way to classify host specificity [7]. Protozoan parasites that are environmentally transmitted are also often more specialist than expected by chance, whereas vector-borne protozoa are typically generalist. Some caution must be applied when jointly considering parasite type and transmission mode, as they are not truly independent predictors. For example, because of phylogenetic conservatism, transmission mode does not always vary across species in the same order, sometimes in the same phylum (e.g., all acanthocephalans and all cestodes have complex life cycles). This means that when comparing across transmission modes within a group like helminths, the taxonomic compositions of the different subsets of parasites are fundamentally different.

Previous research based on taxonomic definitions of generalism have tended to suggest that viruses and protozoa are relatively generalist and helminths relatively specialist [7]. Some of the differences between these preivous findings and the ones we present here emerge from our consideration of host phylogenetic breadth; taxonomic definitions may exagerate rare but large host species jumps by classifying a parasite as, for instance, associating with hosts of multiple orders, even if that parasite is most often associated with hosts within the same genus (where we capture large jumps with the complementary standard effect size for span). In addition, and in contrast with purely taxonomic definitions of generalism, examining the standard effect sizes of host phylogenetic diversity metrics allows for continuous, standardized measures that facilitate comparison across parasite species and with null expectations thereby providing more robust hypotheses testing.

By using phylogeny rather than taxonomic definitions we found, for example, that while viruses often infect closely related hosts, as has been observed previously [17], they are also a group that contains several parasites whose hosts are distantly related. Similarly, previous research on a subset of primate helminths demonstrated strong patterns of cospeciation, with occasional cross-clade host switching [44]. This pattern of host-parasite association is reflected in our observation that the level of host specificty of several helminths is equivalent to that of random host sets, and these parasites may both "jump" and "creep" through host phylogeny, evidenced by several examples of large-span and high-aggregation helminths.

While the underlying data in our analyses represent known host-parasite associations, they are not complete [45]. However, we established that our sample of multi-host parasites is not a taxonomically biased sample when compared to single-host parasites, and have established the likely biases associated with well-studied host species, as well as which parasite types are most impacted by these biases by comparing null modeling approaches. We acknowledge that the associations between hosts and parasites are not necessarily indicative of parasite fitness. Moreover, sharing may represent several different mechanisms. Parasites may jump between host species, establishing in each, as demonstrated by rabies virus [46]. Parasites can also maintain themselves in some reservoir host species and occasionally spill over to others, as evidenced, for example, in MERS coronovirus transmission from camels to humans [47]. In addition, some of what we call the same parasite in closely related hosts may indeed be different species, and reflect cospeciation events rather than any host shifts at all. Consequently, it is difficult to infer future potential for novel host acquisition from existing data based only on observations of presences, although promising statistical approaches have recently been put forward [48–50]. Plausibly, parasites with rapid evolution may be both good adaptors to, and explorers of, the space of host species, as is indicated by many of the viruses, which exhibit associations with dispersed clusters of host species in the mammal phylogeny, perhaps indicative of taking occasional leaps to novel host species followed by subsequent colonization of closely related host species. For successful host shifts, there must also be the opportunity for host jumps: reservoir and naïve hosts must come into contact, at least indirectly [51]. One obvious barrier to host shifts is geography. Host species geographical ranges have previously been shown to be one of the strongest predictors of viral parasite sharing among primates [16], but are only implicitly included here via the non-independence of range overlap and phylogenetic relatedness [52]. The explicit inclusion of geography is a promising line of macroecological inquiry that could provide further insights into parasite host range.

By extending taxonomic definitions of generalism, using a large set of host and parasite species, and using standard effect sizes for complementary measures of average host relatedness, span, and aggregation in the host phylogeny, the nuanced patterns of host-parasite associations are clarified. Further biological interpretation of the patterns presented will require extensive accumulation and accessibility of additional variables that are known to impact host specificity. These additional variables include the size of parasites, their mode of reproduction, and mutation rates of these parasitic organisms [29,53]. Such data will provide a means to address a variety of questions. For example, is the low specificity of bacteria attributable to mutation rates, horizontal gene transfer or the need to maintain broadly-acting virulence genes to compete with other microbes in the environment? [54]. Does virus architecture, as summarized by the Baltimore classification influence host specificity directly and via the frequency of vector-borne transmission? Is helminth host specificity constrained by number of intermediate hosts? [55] Addressing these kinds of questions will greatly improve our understanding of how parasite diversity is ecologically regulated, and which potential host species, including humans, are at risk of acquiring novel parasites. Zoonotic transmission of parasites from animal to human populations is a longstanding concern [3] and has led to the assembly of data sources that identify common parasite types and transmission modes associated with emerging infectious diseases in humans [3,56]. Our study supports these efforts by characterizing the potential pool of mammalian parasites by their general ability to associate with different hosts across mammalian phylogeny.

## Acknowledgements

The Macroecology of Infectious Disease Research Coordination Network, funded by NSF (DEB 131223) provided useful discussions and support for this work. C. Nunn also acknowledges NSF (BCS 1355902). S. Huang is supported by the Alexander von Humboldt Postdoctoral Fellowship.

## References

## Supplementary material

### Parasite taxonomy, transmission mode and their interaction explain significant variation in mean pairwise PD

#simple test to motivate investigation of factors and their interaction  
basic.aov<-aov(lm(nri.flat$mpd.obs.z~nri.flat$para.type\*nri.flat$tmode))  
summary(basic.aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## nri.flat$para.type 4 48.97 12.243 37.746 < 2e-16 \*\*\*  
## nri.flat$tmode 3 7.77 2.592 7.990 2.91e-05 \*\*\*  
## nri.flat$para.type:nri.flat$tmode 10 7.63 0.763 2.351 0.00962 \*\*   
## Residuals 952 308.79 0.324   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Table S1: ANOVA testing differences in standard effect size of mean pairwise PD according to parasite type, transmission mode and their interaction.

### An alternative null model ignoring host frequencies is less conservative

The null model implemented in the main text randomizes the matrix of host-parasite associations conserving margin totals of the presence/absence matrix and weights phylogenetic distances by the relative frequencies (in the GMPD) of the chosen host pair. The conservation of features of the observed matrix structure and integration of host species biases in the number of records in the database renders this null model relatively conservative (and known as a fixed-fixed, weighted null - where 'fixed' refers to the conservation of row or column sums). This means it is relatively unlikely to support the inference of non-random patterns in data. Here, we contrast this approach with a null model that does not weight phylogenetic distances by frequency of host records, but simply uses the presence/absence matrix (a fixed-fixed, unweighted null). The unweighted variant is less conservative; almost all pairwise comparisons of standard effect sizes generated under the two models lie above the one-to-one line (Fig. S1 - black line). For example, a score of under the unweighted model, translates to in the weighted model (i.e., less likely to lie below the threshold typically associated with statistically significant specificity). Among the parasite types, viruses followed by bacteria, exhibit the shallowest slopes in this comparison (Fig. S1 - colored lines) meaning they will have the greatest discrepancy regarding significant specificity associated with the null models, i.e., significant specialist under an unweighted null reverting to generalist when accounting for host occurrence frequencies.

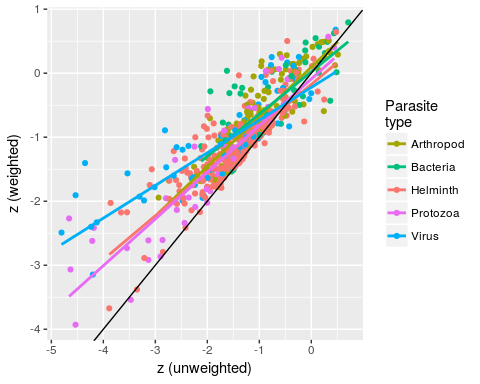


Figure S1: Correlation of the standard effect sizes (z scores) of mean pairwise PD under a model that ignores host frequencies in GMPD (unweighted), with a similar one that includes host frequencies, and was used in the main text (weighted). Colors represent different parasite types. The black line represents the one-to-one correspondence between the two measures and colored lines are linear regressions of the weighted scores on the unweighted scores by parasite type.

### Single- vs. multi-host parasites

The main analysis is centered on parasite species that infect at least two host species. Including parasite species for which only one terrestrial mammal host is identified in the database provides an opportunity to compare single-host parasites and multi-host parasites. Quantitatively, single host parasites () are almost as common as multi-host parasites (), and broadly reflect the composition of parasite taxonomy and transmission modes observed in multi-host parasites (Fig. S2). A few notable exceptions include the virus group, which is under-represented among single-host parasites in general, and protozoan complex life-cycle parasites, which are more commonly single-host (with a mammal species as the final, definitive host) than multi-host. Additionally, in arthropods, close contact transmission has a higher frequency in single-host parasites, while environmental transmission has a higher frequency of multi-host parasites. However, generally this analysis indicates that our main results are not considerably biased by excluding single-host parasites.

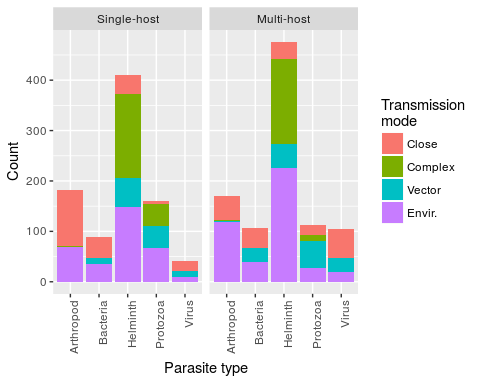


Figure S2: Number of single-host and multi-host (>1 host species) parasites as a function of parasite type and transmission mode. Parasites with more than one reported transmission mode (360/1476) are counted in each valid transmission mode category.

### Interactions affecting host specificity are largely driven by protozoa and viruses and by close contact and vector-borne transmission

Here, we consider each transmission mode in turn and score each parasite as either transmitting by that mode (1) or not (0). While several parasite types exhibit different patterns of host relatedness according to this 1/0 scoring, we note in particular that protozoan parasites are more likely to exhibit specialism under close contact transmission and generalism under vector-borne and complex life-cycle transmission modes. Similarly, viruses undergoing close contact transmission are much more likely to be specialist than vector-borne viruses.

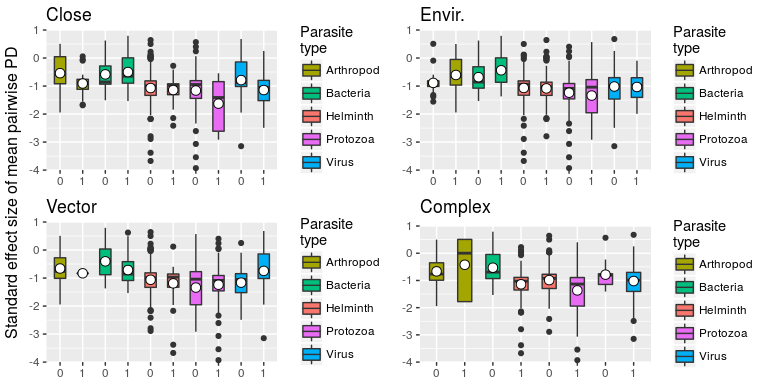


Figure S3: Standard effect size of mean pairwise PD between hosts stratified by parasite type (colored bars) and presence (1) or absence (0) of each transmission mode (named subplots), indicating interactions between parasite type (particularly protozoa and viruses) and transmission mode (particularly close contact and vector-borne transmission)

### Effect of host order on relatedness of hosts within a parasite's host set

Several parasites exclusively infect hosts within a host order. Grouping host species by order allows us to inspect how results on host specificity are driven by the non-monophyletic structure of a tree with three host orders. Here, the null models are constructed via random mammalian host species sampling according to the host taxonomic bias of the parasite. For example, a parasite known to infect 5 carnivores, 5 primates and 0 ungulates would have its random sets constructed from 5 randomly selected carnivores and 5 randomly selected primates. The plots separate parasites that infect hosts of multiple orders, carnivores only, primates only and ungulates only. We observe that carnivore-exclusive bacteria and viruses are typically more specialist than bacteria and viruses and infect other mammal orders (Fig. S4). Transmission mode has little impact on parasites exclusively associated with either primates or ungulates, whereas parasites of carnivores and those infecting multiple host orders exhibit specialism particularly when transmitted by close contact, and additionally vector-borne parasites in the case of carnivores (Fig. S5).

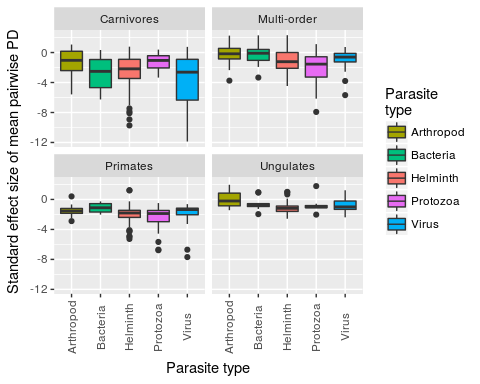


Fig. S4: The patterns for standard effect size of mean pairwise PD of all host species is broadly reflected in the patterns of specific host orders. For viruses infecting only one host order, the carnivore host group is particularly associated with specialist viruses.

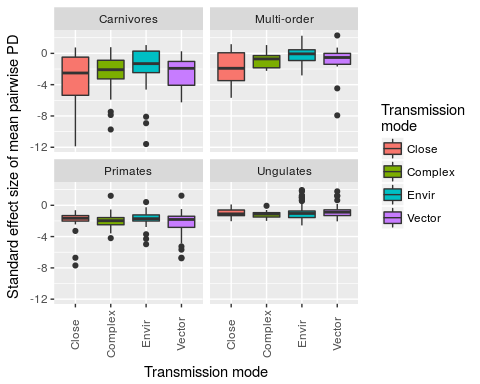


Fig. S5: Standard effect size of mean pairwise PD stratified by host order and parasite transmission mode. Parasites affecting only primates and ungulates do not exhibit differences in specialism/generalism due to parasite transmission mode. The carnivore-exclusive parasites show a similar trend to the main result (Figs. 1 & 2).

### Probability density for standard effect size of mean pairwise PD

Few parasite species exhibit a positive standard effect size for mean pairwise PD, and none of those are significantly different from values expected under the null model. The mean and variance of the effect sizes depends on parasite type. In the main text, we focus on helminths and viruses as two groups with similar mean but different variance - to highlight the importance of characterizing the tendency for groups to contain parasites with extreme standard effect sizes (Fig. 3). For completeness, here we present the density plots of the same standard effect sizes for the five parasite types together (Fig. S6).

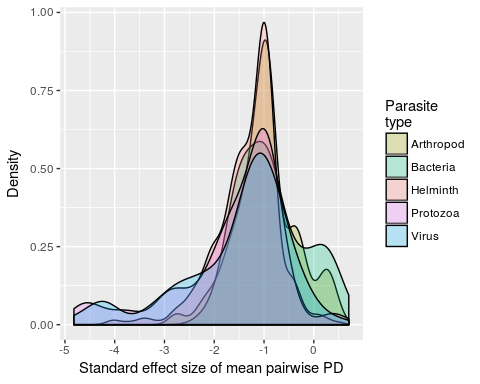


Figure S6: Density of standard effect size of mean pairwise PD by parasite type.

### Parasite transmission mode as a predictor of span and aggregation of hosts in the mammalian phylogeny

Each transmission mode group contains a majority of parasites species whose standard effect sizes for span and aggregation are not different those of equivalent random host sets (Fig. S7). All groups contain some parasites that have a smaller span than expected by chance and contain some parasites that exhibit patterns of host species phylogenetic aggregation. Patterns are relatively similar across all transmission modes.

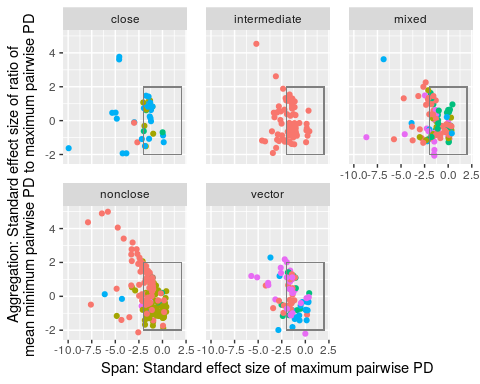


Figure S7: Standard effect sizes for span and aggregation of hosts of each parasite, grouped by parasite transmission mode (including 'mixed' for parasites with multiple transmission modes) and colored by parasite type (same color scheme as the complementary Fig. 4, main text).

### Observed phylogenetic metrics of each parasite type

The null modeling approach first compares observed to simulated values for each phylogenetic measure of specialism, before applying standard effect size calculations. These raw values presented by parasite group show that phylogenetic measurements are generally in numerical agreement across parasite types, with helminth and protozoan parasites exhibiting slightly lower values, which can reflect both barriers to host-parasite associations and the particular phylogenetic structure of each parasites host set.

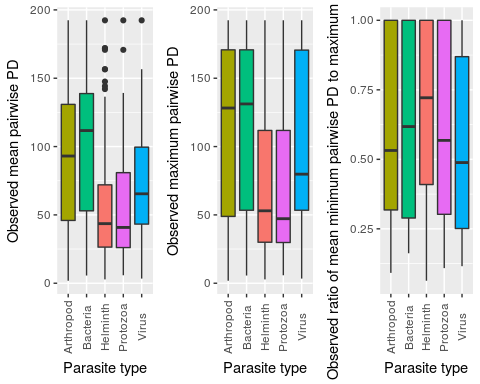


Figure S9: Raw values of observed phylogenetic metrics by parasite type.

### Parasite species richness and host phylogenetic isolation

Both evolutionary distinctiveness and terminal branch length of each host species are calculated relative to their parasite species richness to explore the extent to which host-parasite association patterns are driven by phylogenetic isolation of certain host species (Fig. S9).

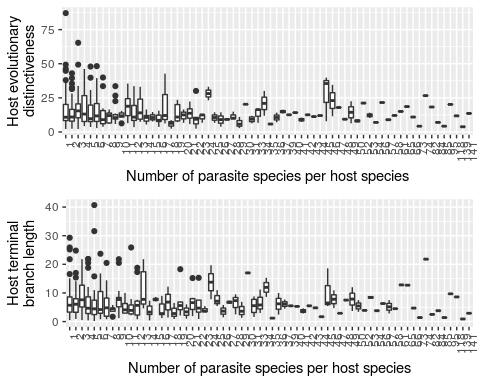


Figure S9: Box-and-whisker plots showing the relationship between number of parasite species per host species and host evolutionary distinctiveness (top panel) and host terminal branch length (bottom panel).

From a set of negative binomial generalized linear models, the main effect in these models, evolutionary distinctivenes (ed) or terminal branch length (tbl) is only significant for ed when used without controlling for the number of host records (main effects not significant in other models). In this model, the estimate for the evolutionary distinctiveness (ed) covariate is -0.015644. Given the log link function, this may be interpreted in the following way. For each unit change in ed (measured in millions of years) there is a reduction in parasite species richness by exp(-0.015644)=0.984~1 parasite.

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

##   
## Call:  
## glm.nb(formula = n.para ~ ed, data = H.metrics, maxit = 100,   
## link = log, init.theta = 0.7734391846)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5574 -1.2126 -0.6165 0.1408 3.2741   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.807557 0.099748 28.147 < 2e-16 \*\*\*  
## ed -0.015640 0.005512 -2.837 0.00455 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(0.7734) family taken to be 1)  
##   
## Null deviance: 456.21 on 403 degrees of freedom  
## Residual deviance: 450.04 on 402 degrees of freedom  
## AIC: 2910.4  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 0.7734   
## Std. Err.: 0.0517   
##   
## 2 x log-likelihood: -2904.3920

##   
## Call:  
## glm.nb(formula = n.para ~ tbl, data = H.metrics, maxit = 100,   
## link = log, init.theta = 0.7674990804)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5154 -1.1934 -0.6343 0.1481 3.1468   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.69931 0.09226 29.26 <2e-16 \*\*\*  
## tbl -0.01695 0.01040 -1.63 0.103   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(0.7675) family taken to be 1)  
##   
## Null deviance: 453.04 on 403 degrees of freedom  
## Residual deviance: 450.69 on 402 degrees of freedom  
## AIC: 2914.2  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 0.7675   
## Std. Err.: 0.0512   
##   
## 2 x log-likelihood: -2908.1740

##   
## Call:  
## glm.nb(formula = n.para ~ ed + n.records, data = H.metrics, maxit = 100,   
## link = log, init.theta = 1.118445548)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.6063 -1.0833 -0.4941 0.3550 2.3913   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.167e+00 8.842e-02 24.503 <2e-16 \*\*\*  
## ed -8.565e-03 4.754e-03 -1.802 0.0716 .   
## n.records 1.701e-03 8.567e-05 19.859 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.1184) family taken to be 1)  
##   
## Null deviance: 633.01 on 403 degrees of freedom  
## Residual deviance: 426.68 on 401 degrees of freedom  
## AIC: 2746.8  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.1184   
## Std. Err.: 0.0814   
##   
## 2 x log-likelihood: -2738.7840

##   
## Call:  
## glm.nb(formula = n.para ~ tbl + n.records, data = H.metrics,   
## maxit = 100, link = log, init.theta = 1.113795854)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.5950 -1.0695 -0.4886 0.3613 2.2107   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.096e+00 8.165e-02 25.675 <2e-16 \*\*\*  
## tbl -8.006e-03 8.931e-03 -0.896 0.37   
## n.records 1.707e-03 8.573e-05 19.912 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.1138) family taken to be 1)  
##   
## Null deviance: 630.72 on 403 degrees of freedom  
## Residual deviance: 427.08 on 401 degrees of freedom  
## AIC: 2748.7  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.1138   
## Std. Err.: 0.0810   
##   
## 2 x log-likelihood: -2740.7090

Table S2: Set of negative binomial generalized linear models used to explore the effect of evolutionary distinctiveness (ed), terminal branch length (tbl), and number of host species records in the GMPD (n.records) on the number of parasite species per host species, using the glm.nb formula in the MASS package in R.

1. Cleaveland S, Laurenson MK, Taylor LH. 2001 Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences* **356**, 991–999. (doi:[10.1098/rstb.2001.0889](https://doi.org/10.1098/rstb.2001.0889))

2. Cleaveland S, Hess GR, Dobson AP, Laurenson MK, McCallum HI, Roberts MG, Woodroffe R. 2002 The role of pathogens in biological conservation. In *The ecology of wildlife diseases* (eds PJ Hudson, A Rizzoli, BT Grenfell, H Heesterbeek, AP Dobson), pp. 139–150. Oxford, UK: Oxford University Press.

3. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008 Global trends in emerging infectious diseases. *Nature* **451**, 990–3.

4. Fuller T, Bensch S, Müller I, Novembre J, Pérez-Tris J, Ricklefs R, Smith T, Waldenström J. 2012 The ecology of emerging infectious diseases in migratory birds: an assessment of the role of climate change and priorities for future research. *EcoHealth* **9**, 80–88. (doi:[10.1007/s10393-012-0750-1](https://doi.org/10.1007/s10393-012-0750-1))

5. Kreuder Johnson C *et al.* 2015 Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Scientific Reports* **5**, 14830. (doi:[10.1038/srep14830](https://doi.org/10.1038/srep14830))

6. Poulin R. 1992 Determinants of host-specificity in parasites of freshwater fishes. *International Journal for Parasitology* **22**, 753–758. (doi:[10.1016/0020-7519(92)90124-4](https://doi.org/10.1016/0020-7519(92)90124-4))

7. Pedersen AB, Altizer S, Poss M, Cunningham AA, Nunn CL. 2005 Patterns of host specificity and transmission among parasites of wild primates. *International journal for parasitology* **35**, 647–57.

8. Rohde K. 1980 Host specificity indices of parasites and their application. *Experientia* **36**, 1369–1371. (doi:[10.1007/bf01960103](https://doi.org/10.1007/bf01960103))

9. Poulin R, Mouillot D. 2003 Parasite specialization from a phylogenetic perspective: a new index of host specificity. *Parasitology* **126**, 473–480.

10. Faith DP. 1992 Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1–10.

11. Purvis A, Hector A. 2000 Getting the measure of biodiversity. *Nature* **405**, 212–219.

12. Krasnov BR, Mouillot D, Shenbrot GI, Khokhlova IS, Poulin R. 2004 Geographical variation in host specificity of fleas (Siphonaptera) parasitic on small mammals: the influence of phylogeny and local environmental conditions. *Ecography* **27**, 787–797. (doi:[10.1111/J.0906-7590.2004.04015.X](https://doi.org/10.1111/J.0906-7590.2004.04015.X))

13. Poulin R, Krasnov BR, Mouillot D. 2011 Host specificity in phylogenetic and geographic space. *Trends in parasitology* **27**, 355–61.

14. Stephens PR *et al.* 2016 The macroecology of infectious diseases: a new perspective on global-scale drivers of pathogen distributions and impacts. *Ecology Letters* **19**, 1159–1171.

15. Cooper N, Griffin R, Franz M, Omotayo M, Nunn CL, Fryxell J. 2012 Phylogenetic host specificity and understanding parasite sharing in primates. *Ecology letters* **15**, 1370–7.

16. Davies TJ, Pedersen AB. 2008 Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proceedings. Biological sciences / The Royal Society* **275**, 1695–701.

17. Huang S, Bininda-Emonds ORP, Stephens PR, Gittleman JL, Altizer S. 2014 Phylogenetically related and ecologically similar carnivores harbour similar parasite assemblages. *Journal of Animal Ecology* **83**, 671–680. (doi:[10.1111/1365-2656.12160](https://doi.org/10.1111/1365-2656.12160))

18. Poulin R. 2010 Decay of similarity with host phylogenetic distance in parasite faunas. *Parasitology* **137**, 733–741. (doi:[doi:10.1017/S0031182009991491](https://doi.org/doi:10.1017/S0031182009991491))

19. Presley SJ, Dallas T, Klingbeil BT, Willig MR. 2015 Phylogenetic signals in host-parasite associations for Neotropical bats and Nearctic desert rodents. *Biological Journal of the Linnean Society* **116**, 312–327. (doi:[10.1111/bij.12601](https://doi.org/10.1111/bij.12601))

20. Ricklefs RE, Fallon SM. 2002 Diversification and host switching in avian malaria parasites. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 885–892. (doi:[10.1098/rspb.2001.1940](https://doi.org/10.1098/rspb.2001.1940))

21. Page RDM. 2003 *Tangled trees: phylogeny, cospeciation, and coevolution*. University of Chicago Press.

22. Harvey PH. 1996 Phylogenies for ecologists. *Journal of Animal Ecology* **65**, 255–263.

23. Freckleton RP, Harvey PH, Pagel M. 2002 Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist* **160**, 712.

24. Harvey PH, Pagel MD. 1991 *The comparative method in evolutionary biology*. Oxford University Press Oxford.

25. Pfennig DW. 2000 Effect of predator-prey phylogenetic similarity on the fitness consequences of predation: a trade-off between nutrition and disease? *The American Naturalist* **155**, 335–345.

26. Longdon B, Hadfield JD, Webster CL, Obbard DJ, Jiggins FM. 2011 Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathog* **7**, e1002260. (doi:[10.1371/journal.ppat.1002260](https://doi.org/10.1371/journal.ppat.1002260))

27. De Vienne DM, Hood ME, Giraud T. 2009 Phylogenetic determinants of potential host shifts in fungal pathogens. *Journal of Evolutionary Biology* **22**, 2532–2541. (doi:[10.1111/j.1420-9101.2009.01878.x](https://doi.org/10.1111/j.1420-9101.2009.01878.x))

28. Rosas-Valdez R, León GP-P de. 2010 Patterns of host specificity among the helminth parasite fauna of freshwater siluriforms: testing the biogeographical core parasite fauna hypothesis. *Journal of Parasitology* **97**, 361–363. (doi:[10.1645/ge-2541.1](https://doi.org/10.1645/ge-2541.1))

29. Woolhouse MEJ. 2001 Population Biology of Multihost Pathogens. *Science* **292**, 1109–1112. (doi:[10.1126/science.1059026](https://doi.org/10.1126/science.1059026))

30. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. 2008 Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol R Microbiol Mol Biol R* **72**, 457–+.

31. Elena SF, Froissart R. 2010 New experimental and theoretical approaches towards the understanding of the emergence of viral infections. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 1867–1869. (doi:[10.1098/rstb.2010.0088](https://doi.org/10.1098/rstb.2010.0088))

32. Stephens PR *et al.* 2017 Global Mammal Parasite Database version 2.0. *Ecology* (doi:[10.1002/ecy.1799](https://doi.org/10.1002/ecy.1799))

33. Nunn CL, Altizer SM. 2005 The global mammal parasite database: An online resource for infectious disease records in wild primates. *Evolutionary Anthropology: Issues, News, and Reviews* **14**, 1–2.

34. R Core Team. 2016 R: A Language and Environment for Statistical Computing.

35. Bininda-Emonds ORP *et al.* 2007 The delayed rise of present-day mammals. *Nature* **446**, 507–12.

36. Fritz SA, Bininda-Emonds ORP, Purvis A. 2009 Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology Letters* **12**, 538–549. (doi:[10.1111/j.1461-0248.2009.01307.x](https://doi.org/10.1111/j.1461-0248.2009.01307.x))

37. Swenson NG, Ackerly D, Kembel S, Zimmerman J, Zimmerman J. 2009 Phylogenetic Resolution and Quantifying the Phylogenetic Diversity and Dispersion of Communities. *PLoS ONE* **4**, e4390. (doi:[10.1371/journal.pone.0004390](https://doi.org/10.1371/journal.pone.0004390))

38. Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. 2002 Phylogenies and community ecology. *Annual Review of Ecology and Systematics* **33**, 475–505.

39. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010 Picante: {R} tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464.

40. Gotelli NJ. 2000 Null model analysis of species co-occurrence patterns. *Ecology* **81**, 2606–2621.

41. Cooper N, Rodríguez J, Purvis A. 2008 A common tendency for phylogenetic overdispersion in mammalian assemblages. *Proceedings of the Royal Society of London B: Biological Sciences* **275**.

42. Ulrich W, Gotelli NJ. 2013 Pattern detection in null model analysis. *Oikos* **122**, 2–18. (doi:[10.1111/j.1600-0706.2012.20325.x](https://doi.org/10.1111/j.1600-0706.2012.20325.x))

43. Redding DW, Mooers AØ. 2006 Incorporating Evolutionary Measures into Conservation Prioritization. *Conservation Biology* **20**, 1670–1678. (doi:[10.1111/j.1523-1739.2006.00555.x](https://doi.org/10.1111/j.1523-1739.2006.00555.x))

44. Hugot JP. 1999 Primates and their pinworm parasites: the Cameron hypothesis revisited. *Systematic biology* **48**, 523–46.

45. Hopkins ME, Nunn CL. 2007 A global gap analysis of infectious agents in wild primates. *Diversity and Distributions* **13**, 561–572. (doi:[10.1111/j.1472-4642.2007.00364.x](https://doi.org/10.1111/j.1472-4642.2007.00364.x))

46. Holmes EC. 2009 *The Evolution and Emergence of RNA Viruses*. OUP Oxford. See <https://books.google.com/books?id=79LwAAAAMAAJ{\&}pgis=1>.

47. Mackay IM, Arden KE. 2015 MERS coronavirus: diagnostics, epidemiology and transmission. *Virology Journal* **12**, 222. (doi:[10.1186/s12985-015-0439-5](https://doi.org/10.1186/s12985-015-0439-5))

48. Han BA, Kramer AM, Drake JM. 2016 Global Patterns of Zoonotic Disease in Mammals. *Trends in Parasitology* **32**, 565–577.

49. Dallas T, Park AW, Drake JM. 2017 Predicting cryptic links in host-parasite networks. *PLOS Computational Biology* **13**, e1005557. (doi:[10.1371/journal.pcbi.1005557](https://doi.org/10.1371/journal.pcbi.1005557))

50. Dallas T, Park AW, Drake JM. 2017 Predictability of helminth parasite host range using information on geography, host traits and parasite community structure. *Parasitology* **144**, 200–205. (doi:[10.1017/S0031182016001608](https://doi.org/10.1017/S0031182016001608))

51. Wolfe ND, Dunavan CP, Diamond J. 2007 Origins of major human infectious diseases. *Nature* **447**, 279–283. (doi:[10.1038/nature05775](https://doi.org/10.1038/nature05775))

52. Poulin R, Keeney DB. 2008 Host specificity under molecular and experimental scrutiny. *Trends in parasitology* **24**, 24–8.

53. Poulin R, Morand S. 2000 The Diversity of Parasites. *The Quarterly Review of Biology* **75**, 277–293. (doi:[10.1086/393500](https://doi.org/10.1086/393500))

54. Wilson BA, Salyers AA. 2003 Is the evolution of bacterial pathogens an out-of-body experience? *Trends in Microbiology* **11**, 347–350. (doi:[10.1016/S0966-842X(03)00179-3](https://doi.org/10.1016/S0966-842X(03)00179-3))

55. Mouillot D, R. Krasnov B, I. Shenbrot G, J. Gaston K, Poulin R. 2006 Conservatism of host specificity in parasites. *Ecography* **29**, 596–602. (doi:[10.1111/j.0906-7590.2006.04507.x](https://doi.org/10.1111/j.0906-7590.2006.04507.x))

56. Taylor LH, Latham SM, Woolhouse MEJ. 2001 Risk Factors for Human Disease Emergence. *Source: Philosophical Transactions: Biological Sciences Population Biology of Emerging and Re-emerging Pathogens* **356**, 983–989.