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 explore host phylogeny. Using data on over 1400 parasite species that have been documented to infect between 1-81 terrestrial mammal host species, we characterize the generalism of parasites using standard effect sizes for three metrics describing host relatedness: mean pairwise phylogenetic distance (PD), maximum PD, and degree of aggregation of hosts (the tendency for the hosts of a parasite to cluster within the phylogenetic tree). We explore variation in these metrics in terms of parasite taxonomy and transmission mode. Of the multi-host parasites, the majority are generalists, with the average relatedness of their mammalian host species being roughly equivalent to that expected from a random sample of mammals. Only a minority of parasites are associated with host sets that are more related than expected by chance. However, the almost 800 parasite species associating with at least two host species exhibit pronounced trends in the degree of host relatedness, dictated by both parasite taxonomy and transmission mode; bacteria and arthropod parasites are typically the most generalist, helminths are intermediate, and viruses and protozoa are on average the most specialist. While viruses are more specialist than helminths on average, viruses show higher variance on the specialism-generalism continuum, with viruses containing some of the most generalist parasites. Transmission mode also influences the degree of parasite generalism, with closer relatedness of hosts exhibited for parasites that rely on close contact or complex life-cycle transmission involving intermediate hosts, compared to those with environmental or vector-borne transmission. Both maximum PD and degree of aggregation of hosts vary among parasites. For example, all bacteria have large maximum PD, equivalent to random host sets. However, they exhibit the least propensity to infect host clusters within this span. Instead, they appear to typically infect relatively unrelated host species. The taxonomy and transmission modes of parasites with only one known host species are broadly reflective of multi-host parasites, with rarely single-host viruses being a notable exception. Lastly, a host species' evolutionary distinctiveness is a weak predictor of the number of parasite species associated with it. Collectively, this macroecological perspective helps to generate a broad picture of how certain types of parasite and transmission mode are differentially linked with the tendency to associate with multiple host species, captured by a set of complementary metrics defining the dimensions of parasite generalism.

## 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 (or, conversely, host specificity) of parasites is quantified based on the taxonomic breadth of their host species, such as the number of host species [6–8]. Theoretical, as well as empirical, studies using such generalism metrics have provided important insights into the ecology and evolution of parasites, but lack the predictive capability regarding which host species might be subject to future emergent infectious diseases and which parasites are likely to cause them. 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 work taking the evolutionary perspective for further understanding 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 distinctive 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) and compare a diverse range of parasite taxa and transmission mode to identify factors that determine whether parasites can spread across host lineages, and in which manner. For example, parasites may exhibit patterns consistent with creeping through the host phylogeny via related hosts, taking large leaps across the host phylogeny or both.

The spread of parasites across host species has previously been shown to depend on the phylogenetic relatedness among host species [16–18]. Closely related host species not only shared common evolutionary history, and thus common parasites, until divergence in the more recent past than distantly related hosts [19,20], they also show higher similarity in their biology [21–23], including characteristics like the internal physiological structure, immune responses, and behaviors whose differences can largely limit host switching [24–26]. 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 not constrained by host phylogeny, except helminths and viruses [17]. Earlier work has also shown that many helminth species tend to have taxonomically restricted host ranges ranges [7,9,27] but suggested viruses, due to their rapid mutation rates, should more readily adapt to new hosts that are not necessarily closely related to existing hosts [28–30]. These mixed results presented in different systems invite a broader-scale investigation on the question of why some parasites can overcome the physiological and ecological barriers between distantly related host species while others are localized within a host phylogenetic tree.

In this study, we compare the degrees of phylogenetic generalism of 781 parasite species, infecting at least two host species from four mammalian orders: Artiodactyla, Carnivora, Perissodactyla and Primates to search for first-order principles governing the patterns. We consider the variation in phylogenetic generalism 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 study is unparalleled in terms of the diversity of host and parasite species considered together, and has been made possible by the newly published Global Mammal Parasite Database (GMPD) 2.0 [31], which includes an extensive record of parasite occurrence in free-ranging populations of over 400 mammal species from the focal orders, as well as 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 the picture -- 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. In addition, we test the hypothesis that patterns of low parasite species richness tend to be restricted to phylogenetically distinct host species (i.e., species without extant close relatives in the phylogeny).

## Materials and Methods

Records of parasite associations with terrestrial mammals were obtained from the GMPD [31,32]. 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 [31,32], and comprise: close contact, complex life-cycle (i.e., food-borne, with intermediate hosts transitioning parasites to their definitive hosts, including via predator-prey interactions), environmental and vector-borne transmission. The GMPD data were analyzed in R [33] to establish the number and identity of parasite species per host species and the number and identity of host species per parasite species. The resulting study has 404 terrestrial mammal species. Because Perissodactyla is under-represented, with only 10 species compared to 118, 181 and 105 for Carnivora, Primates and Artiodactyla, respectively, we combine this order with Artiodactyla to form an ungulate group consisting of 115 species.

Additionally, a complete phylogeny of mammals [34] was used to obtain the phylogenetic distance (PD) between all pairs of hosts. For the set of hosts of each parasite species, the mean pairwise PD between hosts was calculated, as was its standard effect size [35]. The latter was obtained using *picante* [36], under a null model in which the community data matrix (host-parasite associations) was randomized with the independent swap algorithm [37], maintaining species occurrence frequency and sample species richness. This measure captures the average relatedness of the host species of a given parasite species on a standardized scale for comparison. Two further standard effect size metrics were calculated on each parasite species' host set in the same way: maximum PD of any two hosts in a parasite's set, and the ratio of the mean minimum PD to the maximum PD. 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 [38] (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 (specialist) and above +1.96 (generalist) are typically significantly different from the null expectation, assuming the null model generates a normally-distributed set of scores [39]. Host evolutionary distinctiveness, measured as millions of years of evolutionary separation, was estimated directly from the mammal phylogeny [34] using the *evol.distinct* function in the R package *picante* [36], utilizing the *equal splits* function in which shared branches are apportioned equally among descendant lineages [40]. Each hosts's terminal branch length was also recorded as an additional measure of evolutionary isolation.

## Results

Across all parasites, infected hosts tend to be more related than expected by chance, indicated by negative z-scores for mean pairwise PD between hosts (Fig. 1, global median z-score=-1.1). However, while 96% of parasites exhibit negative z-scores (trending to more specialist than expected by chance), only 16% of parasites have significantly negative z-scores. No parasites were found to have significantly positive z-scores.

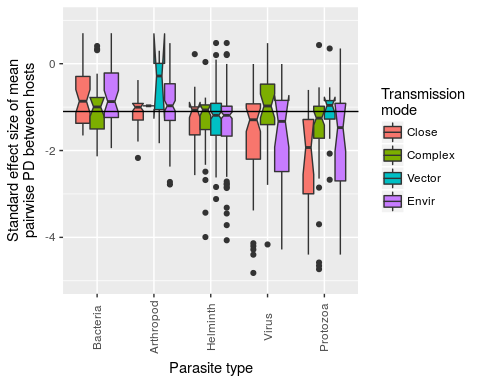


Figure 1. Boxplot of z-scores in each parasite group stratified by transmission mode. Negative z-scores indicate that a parasite species tends to infect host species that are more closely related than expected by chance (under the null model). Alternatively, positive z-scores indicate the opposite. In terms of deviation from the null model, 16% of z-scores are significantly negative (none of the positive z-scores are statistically significant). Solid horizontal line indicates the global median z-score.

The proportion of these significantly negative z-scores varies across parasite type, with smallest to largest proportions corresponding to bacteria, arthropods, helminths, viruses and protozoa (test for equality of proportions, p<0.001, Fig. 2 - top panel). The proportion of significantly negative z-scores is not significantly different across parasite transmission modes (test for equality of proportions, p=0.95, Fig. 2 - bottom panel). However, across the full range of z-scores, 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).

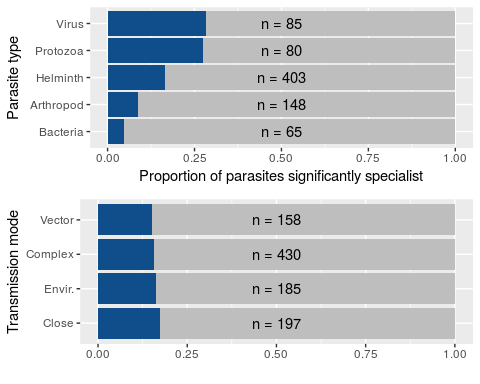


Figure 2. Proportion of parasites (blue bars) with sets of host species that are more related than expected by chance, grouped by parasite type (top panel) and transmission mode (bottom panel). Sample size () for each group is shown in gray part of bars.

Parasite type by transmission mode interactions are particularly driven by protozoa and viruses, and close contact and vector-borne transmission (Fig. S1); 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 generally qualitativey similar to the overall pattern (Figs. S2 & S3). Departures from the main pattern include the observation of several specialist bacteria infecting carnivores (Fig. S2) and the relative insensitivity of parasites infecting primates and ungulates on the patterns of specialization across transmission modes (Fig. S3).

In comparing the most abundant parasite group (helminths) with the group containing the highest proportion of specialists (viruses) it is noteable that viruses exhibit much wider variation in degree of generalism (Fig. 3), containing many extreme specialists (large, negative z-scores) but also many relative generalists (positive z-scores). By contrast, helminths to be more consistent in their degree of generalism, and contain relatively few species that are at either extreme of the generalism-specialism continuum.

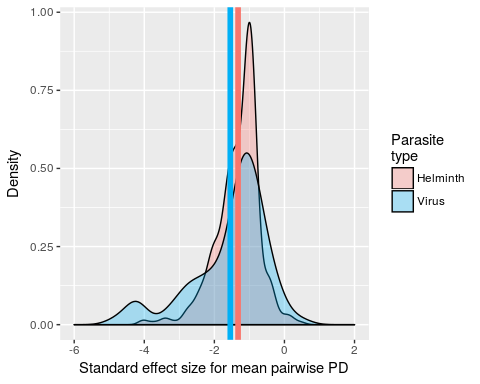


Figure 3. Probability density function of z-scores for mean pairwise PD between hosts of helminths (red, the most abundant parasite group in GMPD) and viruses (blue, the group containing the highest proportion of significantly specialist parasites). The mean z-score is somewhat higher for helminths than for viruses, perhaps reflecting a trend toward more generalism in helminths (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 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. In addition, 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 parasites have spans that are 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). Patterns of span and aggregation are similar when grouped by transmission mode (Fig. S4). An example illustration of two parasites, *Leptospira interrogans* and *Trypanosoma cruzi*, with similar span but very different aggregation patterns are provided in Fig. 5.

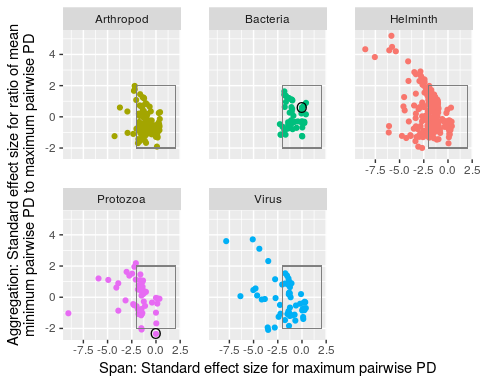


Figure 4. Scatter plot showing both span and aggregation. The x-axes provide a standard effect size of maximum pairwise PD of a parasite's host set (span) and the y-axes provide a standard effect size of the ratio of mean minimum pairwise PD to maximum pariwise PD (aggregation), colored by parasite type.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

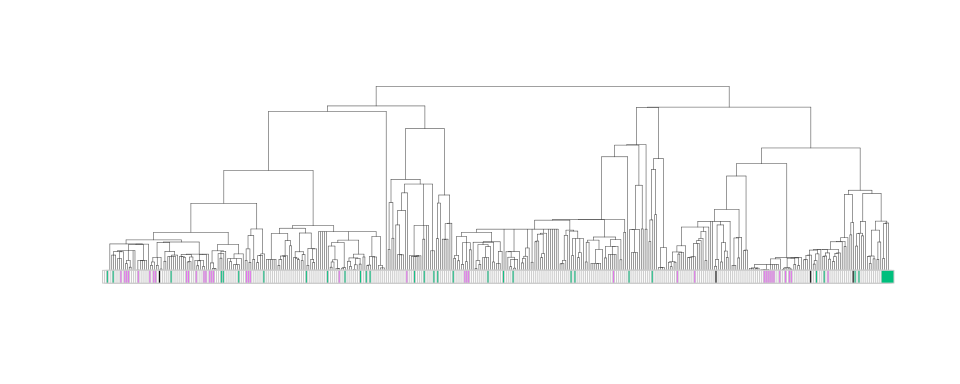


Figure 5. Examples of aggregated (purple) and random (green) distributions of parasites of similar span in the host phylogeny. The parasites are *L. interrogans* (green) and *T. cruzi* (purple), colored according to their groupings in Fig. 4. Black colors indicate hosts infected by both

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. 6). 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 (the mammal is the final, definitive host) than multi-host. Additionally, in arthropods, close transmission has a higher frequency in single-host parasites, switching to a higher frequency of environmental transmission for parasites with multiple hosts.

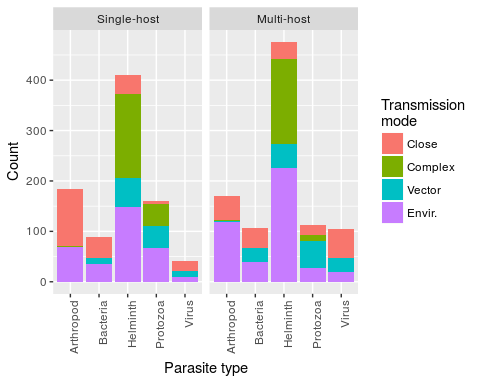


Figure 6. Number of single-host and multi-host (>1 host species) parasites as a function of parasite type and transmission mode. Parasites with more than one reported transmission mode (360/1477) are counted in each valid transmission mode category.

One potential explanation for the number of parasite species associating with a host is the evolutionary isolation of that host. To test this statistically, we first ascertained that the number of parasite species per host species is over-dispersed (variance to mean ratio ~25), meaning that a small number of host species are associated with a large number of parasite species. Accordingly, a set of negative binomial generalized linear models were constructed. 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. 7). 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 those models including 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 especially 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).

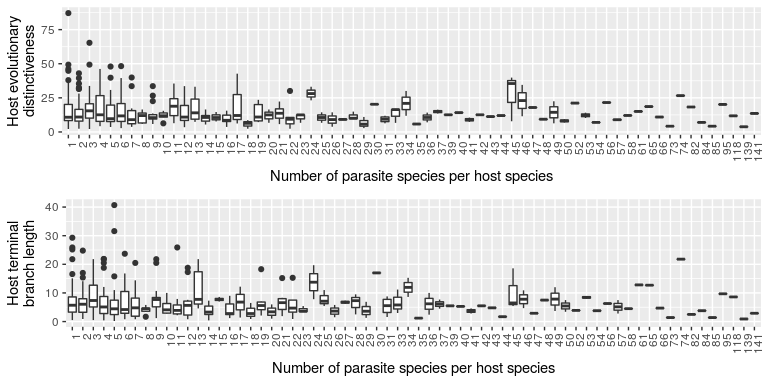


Figure 7. Box-and-whisker plots showing the relationship between number of parasite species per host species and host evolutionary distinctiveness (top panel) and host terminal branch length (bottom panel).

## 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 host species that are comparatively closely related. Of the five parasite groups, bacteria are the most generalist on average and protozoa and viruses contain the most specialist parasites. While transmission mode naturally impacts the opportunity for parasites to encounter novel host species, it was less influential in determining parasite generalism than parasite taxonomy, where the latter likely captures numerous biological traits for which data is currently lacking. 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]. Additionally, protozoan parasites that are environmentally transmitted are often more specialist than expected by chance, whereas vector-borne protozoa are typically generalist.

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 those results and the ones we present here emerge from our consideration of host phylogenetic breadth and not simply host taxonomic richness; 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 models thereby providing more robust hypotheses testing. In doing so, 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 [41], which 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. 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.

While the underlying data represent known host-parasite associations, they are not complete [42] nor are they necessarily indicative of parasite fitness. Parasites may jump between host species, establishing in each, as demonstrated by rabies virus [43]. 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 [44]. Consequently, it is difficult to infer future potential for novel host acquisition from existing data based only on observations of presence-absences. Plausibly, parasites with rapid evolution may be both good adaptors to, and explorers of, the space of host species, as is indicated by the virus group which exhibits 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. Moreover, since host species geographical ranges have previously been shown to be one of the strongest predictors of viral parasite sharing among primates [16], and are only implicitly included here via the non-independence of range overlap and phylogenetic relatedness [45], the explicit inclusion of geography is a promising line of macroecological inquiry.

Further biological interpretation of the patterns presented will require extensive accumulation and accessibility of metadata known to impact host specificity. This includes parasite organism size, mode of reproduction, and mutation rate [28,46]. Such data will give access to 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? [47]. How does virus architecture, summarized by the Baltimore classification, for example, influence host specificty directly and via the frequency of vector-borne transmission? Is helminth host specificity constrained by interspecific competition or by evolutionary history of cospeciation? Opening these kinds of questions will greatly improve our understanding of how parasite diversity is 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 [3,48]. Our study lends support to these efforts by identifying the pool of mammalian parasites and their general ability to associate with distantly related host species as well as host species that are closely related to humans.

## References

## Supplementary material

### Parasite taxonomy, transmission mode and their interaction explain significant variation in mean pairwise PD of the sets of hosts of each parasite

#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$para.type 4 60.8 15.192 27.550 < 2e-16 \*\*\*  
## nri.flat$tmode 3 6.5 2.168 3.931 0.00838 \*\*   
## nri.flat$para.type:nri.flat$tmode 10 13.7 1.375 2.493 0.00591 \*\*   
## Residuals 952 525.0 0.551   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Table S1: ANOVA testing differences in z-scores according to parasite taxonomic group, transmission mode and their interaction.

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

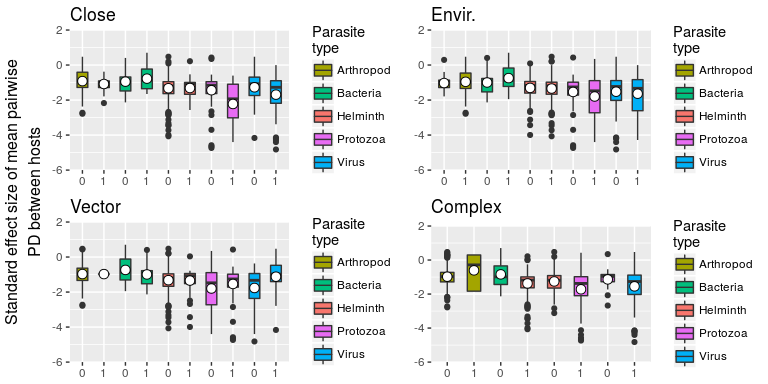


Figure S1: Standard effect size of mean pairwise PD between hosts stratified by parasite type (colored bars) and presence (1) or absence (0) of each transmission mode (named subplots), indicating interactions between parasite type (particularly protozoa and viruses) and transmission mode (particularly 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 in a similar way to the main text, except that random mammalian host species are selected 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 constrcuted from 5 randomly selected carnivores and 5 randomly selected primates. The plot separates parasites that infect hosts of multiple orders, carnivores only, primates only and ungulates only.

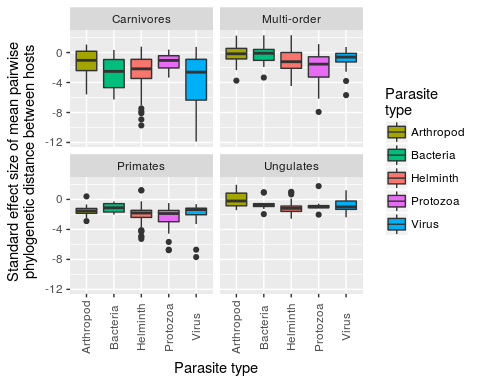


Fig. S2: The patterns for standard effect size of mean pairwise PD of all host species is broadly reflected in the patterns of specific host orders. For viruses infecting only one host order, the carnivore host group is particularly associated with specialist viruses.

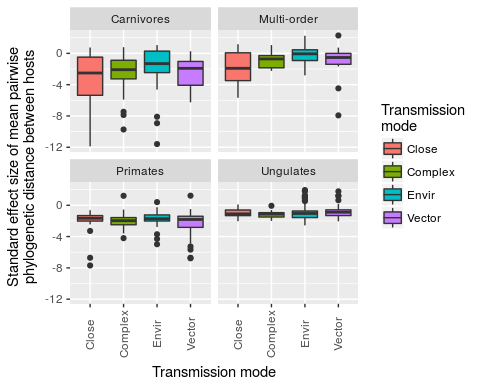


Fig. S3: Standard effect size for mean pairwise PD stratified by host order and parasite transmission mode. Parasites affecting only primates and ungulates do not exhibit differences in specialism/generalism due to parasite transmission mode. The carnivore-exclusive parasites show a similar trend to the main result (Figs. 1 & 2).

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

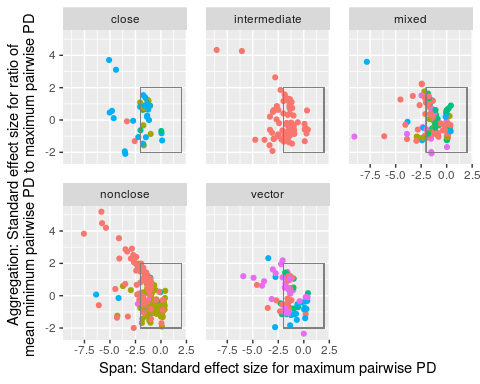


Figure S4: Standard effect sizes for span and aggregation of hosts of each parasite, grouped by parasite transmission mode (including 'mixed' for parasites with multiple transmission modes) and colored by parasite type (same color scheme as the complementary Fig. 4, main text).

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. 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 groups.

### Summary of negative binomial GLMs with predictor variables selected from: evolutionary distinctiveness (ed), terminal branch length (tbl) and number of host records in GMPD (n.records), and response variable number of parasite species per host species (n.para)

##   
## Call:  
## glm.nb(formula = n.para ~ ed, data = H.metrics, maxit = 100,   
## link = log, init.theta = 0.7739623914)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5580 -1.2130 -0.6168 0.1408 3.2747   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.807799 0.099716 28.158 < 2e-16 \*\*\*  
## ed -0.015644 0.005511 -2.839 0.00453 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(0.774) family taken to be 1)  
##   
## Null deviance: 456.20 on 403 degrees of freedom  
## Residual deviance: 450.03 on 402 degrees of freedom  
## AIC: 2910.6  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 0.7740   
## Std. Err.: 0.0518   
##   
## 2 x log-likelihood: -2904.6190

##   
## Call:  
## glm.nb(formula = n.para ~ tbl, data = H.metrics, maxit = 100,   
## link = log, init.theta = 0.7680005892)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5158 -1.1939 -0.6345 0.1480 3.1476   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.69938 0.09223 29.268 <2e-16 \*\*\*  
## tbl -0.01693 0.01040 -1.628 0.103   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(0.768) family taken to be 1)  
##   
## Null deviance: 453.02 on 403 degrees of freedom  
## Residual deviance: 450.67 on 402 degrees of freedom  
## AIC: 2914.4  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 0.7680   
## Std. Err.: 0.0513   
##   
## 2 x log-likelihood: -2908.4110

##   
## Call:  
## glm.nb(formula = n.para ~ ed + n.records, data = H.metrics, maxit = 100,   
## link = log, init.theta = 1.119337296)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.6062 -1.0840 -0.4946 0.3549 2.3919   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.167e+00 8.839e-02 24.519 <2e-16 \*\*\*  
## ed -8.572e-03 4.752e-03 -1.804 0.0713 .   
## n.records 1.700e-03 8.563e-05 19.855 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.1193) family taken to be 1)  
##   
## Null deviance: 633.06 on 403 degrees of freedom  
## Residual deviance: 426.68 on 401 degrees of freedom  
## AIC: 2747  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.1193   
## Std. Err.: 0.0814   
##   
## 2 x log-likelihood: -2738.9930

##   
## Call:  
## glm.nb(formula = n.para ~ tbl + n.records, data = H.metrics,   
## maxit = 100, link = log, init.theta = 1.11465791)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.5947 -1.0701 -0.4893 0.3612 2.2110   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.0966825 0.0816206 25.688 <2e-16 \*\*\*  
## tbl -0.0079740 0.0089273 -0.893 0.372   
## n.records 0.0017060 0.0000857 19.907 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.1147) family taken to be 1)  
##   
## Null deviance: 630.75 on 403 degrees of freedom  
## Residual deviance: 427.09 on 401 degrees of freedom  
## AIC: 2748.9  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.1147   
## Std. Err.: 0.0810   
##   
## 2 x log-likelihood: -2740.9290

Table S2: Set of generlized linear negative binomial 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 (n.para), using the glm.nb formula in the MASS package in R.

The main effect in these models (ed or tbl) is only significant for ed when used without controlling for the number of host records. In this case thspredictor variable is only significant for evolutionary disctinctiveness without controlling for number of host records. In this model, the estimate for the evolutionary distinctiveness (ed) covariate is -0.015644. Given the log link function, this may be interpreted as: 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.

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. 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))

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

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

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

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

24. 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.

25. 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))

26. 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))

27. 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))

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

29. 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–+.

30. 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))

31. 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))

32. 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.

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

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

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

36. 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.

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

38. 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**.

39. 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))

40. 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))

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

42. 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))

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

44. 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))

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

46. 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))

47. 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))

48. 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.