Characterizing parasite generalism illuminates patterns of host species jumps

## Abstract

The efficiency of the exploration of host-species space by parasites is both unequal and poorly understood. Using data on almost 1000 multi-host parasite species infecting terrestrial mammals, we calculate the average host relatedness for each parasite using mammal phylogenetic data. We compare this metric of parasite generalism to that obtained by suitable randomization of hosts in a null model framework. This results in a standard effect size for the mean pairwise phylogenetic distance of hosts, a continuous and comparative measure of the tendency for parasites to infect distantly related host species. We additionally explain variation in this measure of parasite generalism using data on parasite taxonomy and transmission mode. This macroecological perspective helps to generate a broad picture of the propensity for certain types of parasite and transmission mode to lead to the acquisition of novel host species. The parasite groups that are most specialist on average, are also the ones that occasionally take the largest leaps across host species barriers, highlighting the importance of defining the dimensions of parasite generalism.

## Keywords

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

## Introduction

A generalist parasite has a minimal definition: *able to infect at least two host species*. However, for an equivalent number of host species, parasites may vary in their ability to infect distantly related hosts, differentiating their potential to explore the space of host species along with their propensity to infect target species, including humans. The underlying reasons for observing relatedness of hosts of a given parasite include biogeography and co-evolutionary history [1], translating to opportunities for infection, quality resources provided by a host, and limited costs in terms of immune defense and disease-induced mortality. Mechanisms that may allow parasites to overcome host species barriers include rapid evolution and antigenic plasticity [2], novel spatial translocation of host or parasite [3], and evolution of host tolerance [4]. Consequently, it is challenging to predict the potential for parasites to acquire hosts that are distantly related to their existing hosts, in spite of the utility for understanding the distribution of parasites and predicting future zoonoses.

Previous research has defined parasite generalism in alternative ways, including satisfying the criteria of infecting humans and animals [2] as well as using host taxonomies to characterize the degree of parasite generalism [5,6]; for example, classifying parasites by their ability to infect hosts only within a species, or genus, or family, etc. Here, we make novel use of detailed mammalian host phylogeny to define parasite generalism on a continuous scale that allows both comparison among parasites and calibration against a null model [7]. We use the mean pairwise phylogenetic distance of a parasite's host set, a measure whose central tendency is independent of species richness of the set since it focuses on pairs of host species. It has previously been used successfully to explain parasite community similarity in primates [8]. We determine that this continuous measure varies with parasite type, defined both taxonomically and by transmission mode, as has been suggested in earlier studies based on categorical definitions of parasite generalism [2,6]. Differences in our assessment of parasite generalist groups and earlier studies is attributable to the definition of parasite generalism and underscores the value in looking both at general trends and extreme cases of acquiring distantly related hosts.

## Materials and Methods

*Note: cite GMPD for trans mode since cited paper only covers primates*

Parasite records of occurrence with terrestrial mammals were obtained from the global mammal parasite database (GMPD) [9]. From these records, those corresponding to parasite species that infected at least two host species and had a consistent (not host-dependent) transmission mode were retained, resulting in 940 parasite species covering arthropods (), bacteria (), helminths (), protozoa () and viruses () and 399 host species covering the orders Artiodactyla (), Carnivora (), Perissodactyla () and Primate (). Transmission mode definitions were previously assigned from an extensive literature review [6] and cover close contact transmission (), complex life cycle transmission, including trophic transmission (), environmental transmission () and vector-borne transmission ().

Additionally, the mammal phylogenetic supertree [10] was used to obtain the phylogenetic distance between all pairs of hosts. For the host set of each parasite species, the mean pairwise phylogenetic distance between hosts was calculated, along with the standard effect size of mean pairwise distance (sesMPD) [11]. The latter was obtained using the R package *Picante* [12], and a null model in which the community data matrix (host-parasite occurrences) was randomized with the independent swap algorithm [7] maintaining species occurrence frequency and sample species richness. Standard effect sizes (hereafter, z scores) were regressed onto parasite trait data (taxonomic group and transmission mode) available in GMPD using a generalized linear model (GLM, with gaussian link function). Maximum pairwise phylogenetic distance was also calculated to assess which parasites can acquire extremely distantly related hosts, albeit infrequently.

## Results

Across all parasites there is an overwhelming trend of infecting hosts that are more closely related than would be expected by chance, indicated by negative z scores (Fig. 1). Each of the two covariates are predictive of z scores (GLM: parasite type , transmission mode ). However, there is no significant interaction effect; transmission mode trends are consistent across parasite types (). The two covariates explain significant variation in z scores, with a pseudo value [13] of 0.78.

On average, bacteria are the group containing parasites with the most distantly related hosts, followed by arthropods, helminths, viruses and finally protozoa. There is a tendency for parasites that rely on close contact between hosts to infect closely related host species. Conversely, vector-borne parasites and environmentally transmitted parasites are typically the most generalist among a given parasite group. In terms of absolute numbers, of those parasite species in the percentile for z scores (i.e. the most generalist), the descending order of parasite groups is helminth (), arthropod (), bacteria (), virus (), protozoa (). For transmission mode, the percentile for z scores is dominated by environmental transmission (), then complex life cycle (), vector-borne () and close contact ().

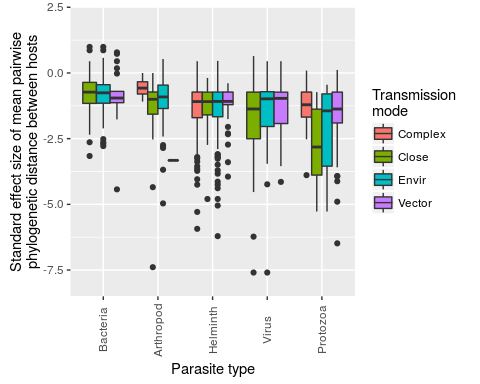


Figure 1. Boxplot of z scores in each parasite group stratified by transmission mode. Positive z scores indicate that a parasite species tends to infect host species that are more distantly related than expected by chance (under the null model). Alternatively, negative z scores indicate the opposite. None of the positive z scores are statistically significant, but the normalized measure allows compararison of generalism across parasite species.

Naturally, the maximum and mean pairwise phylogenetic distance between hosts in a parasite's set are correlated (Fig. 2). However, the ratio of mean to maximum varies across parasite groups, with lowest values typically in viruses and protozoa (Fig. 2, inset). Low ratios indicate that a given parasite's host set includes a few outlier host species, distantly related to the other host species in the set.

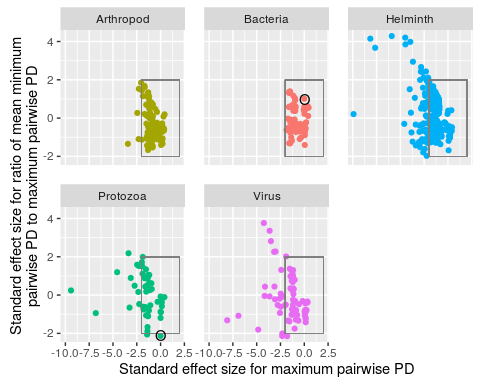


Figure 2: Main: Correlation between mean and maximum pairwise phylogenetic distance of a parasite's host set, colored by parasite type. Inset: ratios of mean to maximum aggregated by parasite type illustrating, for example, that viruses more frequently infect "outlier" hosts whose phylogenetic distance to "typical" hosts is large, resulting in a low ratio of mean to maximum phylogenetic distance.

In comparing the parasite groups that contain the greatest and fewest number of generalists (z scores> percentile; helminths and viruses, respectively) 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 are more consistent in their degree of generalism, and contain relatively few species that are extreme at either side of the generalism-specialistm continuum.

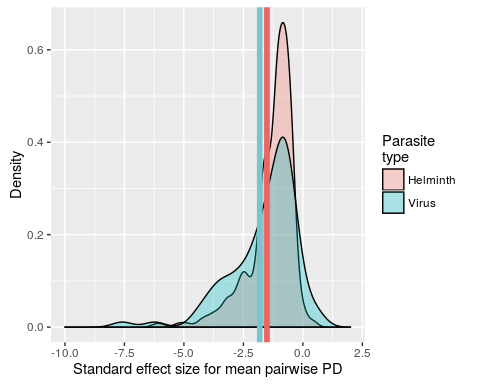
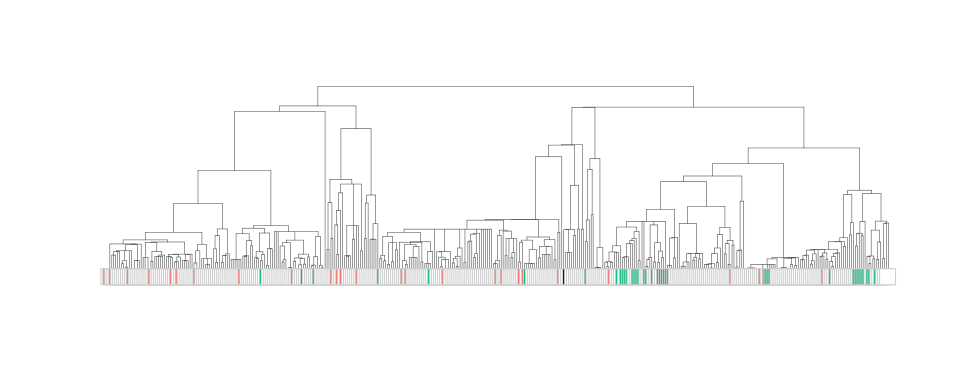


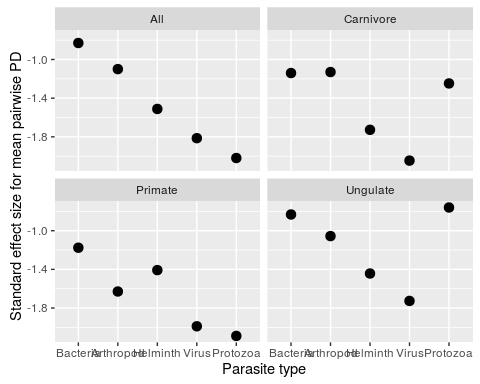
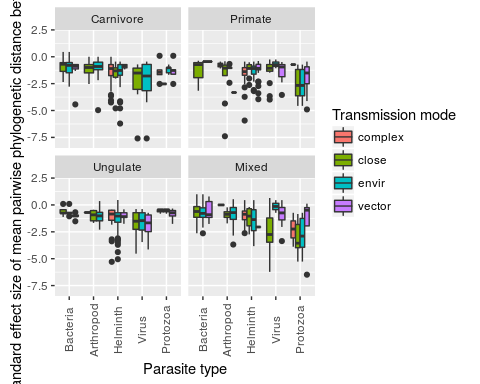
Figure 3: Probability density function of z scores for helminths (red) and viruses (blue). Helminths are more generalist on average (mean values denoted by color-coded vertical lines), but the virus group contains more extreme parasites in terms of both specialism and generalism.

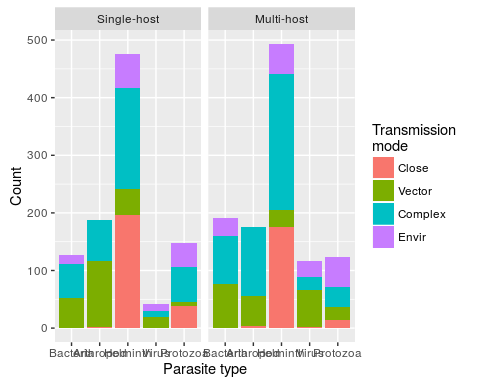
Examples of clumpy vs. non-clumpy distributions of parasites in host phylogenetic tree



The main analysis is centered on parasite species that infect at least two host species. However, there may be important information on parasite specialism revealed by studying those parasite species that are only known to infect one host species among the host orders studied. We consider these parasite species and examine the parasite type and transmission mode, along with the terminal branch length of the host, a simple measure of evolutionary distinctiveness.

Breaking up host species by host order allows us to check if results on generalism are driven by the non-monophyletic structure of a tree with three host orders.





## Discussion

We have shown that, in general, parasites are relatively specialist. However, among the almost one thousand parasite species considered here, some infect host species that are comparatively distantly related. Of the five parasite groups, bacteria are the most generalist on average and helminths represent the group with the largest number of parasite species in the overall highest 10% of z scores (though they are also the most represented group). Conversely, protozoa are the most specialist on average, and have the fewest number of species in the overall highest 10% of z scores. After protozoa, viruses are the next most specialist group.

Previous research based on taxonomic definitions of generalism have tended to suggest that viruses and protozoa are relatively generalist and helminths relatively specialist [6]. The differences between those results and the ones we present here lie in the definition of generalism. For example, while viruses typically infect closely related hosts (shown by large negative z scores), they are also the group that occasionally acquires considerably distantly related hosts (low ratio in mean to maximum host phylogenetic pairwise distances). While other parasite groups may typically contain species whose hosts are more distantly related, they do not show this same propensity to make leaps across large host species barriers. A taxonomic definition of generalism such as the ability to infect hosts of different orders [6], will do well in detecting this potential, whereas looking at the standard effect size of mean host phylogenetic pairwise distances complements this as a continuous, standardized measure that facilitates comparison across parasite species and with null models.

While the underlying data represent known host-parasite occurrences, they are not necessarily indicative of parasite fitness. Parasites may jump between host species, establishing in each, e.g. Rabies virus [14], or maintain themselves in some host species and spillover to others, e.g. West Nile virus [15]. Consequently, it is difficult to infer future potential for novel host acquisition from existing data. 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.

Transmission mode naturally impacts the opportunity to encounter novel host species. Within each parasite group, contact based transmission typically results in a higher degree of specialism, whereas vector-borne and environmental transmission are associated with relative generalism, consistent with previous research [6]. Host species geographical ranges have previously been shown to be one of the strongest predictors of viral parasite sharing among primates [16], and is only implicitly included here due to the non-independence of range overlap and phylogenetic relatedness [1]. Explicit inclusion of geography is a promising line of macroecological inquiry.

## References

1. Poulin, R. & Keeney, D. B. 2008 Host specificity under molecular and experimental scrutiny. *Trends in parasitology* **24**, 24–8.

2. Woolhouse, M. E. J. 2002 Population biology of emerging and re-emerging pathogens. *Trends Microbiol Trends Microbiol* **10**, S3–S7.

3. Farrell, M. J., Berrang-Ford, L. & Davies, T. J. 2013 The study of parasite sharing for surveillance of zoonotic diseases. *Environmental Research Letters* **8**, 015036.

4. Raberg, L., Graham, A. L. & Read, A. F. 2009 Decomposing health: tolerance and resistance to parasites in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 37–49.

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

6. Pedersen, A. B., Altizer, S., Poss, M., Cunningham, A. A. & Nunn, C. L. 2005 Patterns of host specificity and transmission among parasites of wild primates. *International journal for parasitology* **35**, 647–57.

7. Gotelli, N. J. 2000 Null model analysis of species co-occurrence patterns. *Ecology* **81**, 2606–2621.

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

9. Nunn, C. L. & Altizer, S. M. 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.

10. Bininda-Emonds, O. R. P. et al. 2007 The delayed rise of present-day mammals. *Nature* **446**, 507–12.

11. Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. 2002 Phylogenies and community ecology. *Annual Review of Ecology and Systematics* **33**, 475–505.

12. Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P. & Webb, C. O. 2010 Picante: {R} tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464.

13. Faraway, J. J. 2005 *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. CRC Press.

14. Holmes, E. C. 2009 *The Evolution and Emergence of RNA Viruses*. OUP Oxford.

15. Reisen, W. K. 2013 Ecology of West Nile virus in North America. *Viruses* **5**, 2079–105.

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