Characterizing parasite generalism illuminates patterns of host-parasite associations

Abstract

The distribution of parasites among mammalian hosts is complex and represents a differential ability or opportunity to explore the host tree of life. Using data on over 1200 parasite species known to infect between 1-80 mammal host species, we characterize the generalism of parasites using standard effect sizes that relate to average host relatedness, maximum phylogenetic distance between hosts and degree of aggregation of hosts (the tendency for the hosts of a parasite to be distributed clumpily 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 generalist in the sense that the average relatedness of their mammalian host species is roughly equivalent to that expected from a random sample of mammals, with a minority associated with host sets that are more related than expected by chance. The degree of host relatedness is affected by both parasite taxonomy and transmission mode; bacteria and arthropod parasites are typically the most generalist, helminths are intermediate and viruses and protozoa are the most specialist on average. However, characterizing variation within and between groups is important; for example, viruses are more specialist than helminths on average, but the virus group has a larger range on the specialism-generalism continuum, such that the virus group contains some of the most generalist parasites, whereas helminths exhibit less variation in the degree of generalism. While not as strong as parasite taxonomy, transmission mode also influences the degree of parasite generalism, with closer relatedness of hosts exhibited for parasites that rely on close contact or vectors compared to those with environmental or complex life cycle transmission, which includes trophic transmission. Both the span of a parasite across the host tree and the tendency to infect clusters of related hosts vary among parasites. For example, all bacteria have large spans, equivalent to random host sets. However, they exhibit the least propensity to affect host clusters within this span. Instead, they appear to typically affect relatively unrelated host species. Collectively, this macroecological perspective helps to generate a broad picture of the tendency for certain types of parasite and transmission mode to associate with the acquisition of hosts, captured by metrics defining three complementary 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 ability 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 or closely 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 build on methods to generate continuous measures of generalism [7] using a detailed mammalian host phylogeny to define parasite generalism by the patterns of relatedness of the host sets of parasites. Known associations between hosts and parasites come from the global mammal parasite database [8] (*update ref*), from which this study focuses on 416 mammal species representing primates ($n = 162$), carnivores ($n = 143$) and ungulates ($n = 111$). The approach allows both comparison among parasites and calibration against a null model [9]. From each host set, we calculate the standard effect size of mean pairwise phylogenetic distance between host species, a measure that has previously been used successfully to explain parasite community similarity in primates [10]. In addition, we calculate standard effect sizes for both the maximum phylogenetic distance of host sets and a similar metric that captures the tendency for hosts to be clustered in the phylogenetic tree.

We determine that average relatedness of hosts 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]. The speciose study, which includes over 400 host species and over 1200 parasite species, enables generalism to be characterized further. All taxonomic and transmission mode groupings contain parasites that have large spans in the host tree, measured by standard effect size of maximum phylogenetic distance among hosts. Bacteria and environmentally-transmitted parasites show the lowest propensity to infect phylogenetically clustered host groups. Single host parasites ($n = 565$) are almost as common as multi-host parasites ($n = 674$) and show similar distributions among parasite taxonomic and transmission mode groupings, with the exception of viruses, which are more likely to infect multiple host species. We observe that the number of parasites known to infect a host is partly explained by the evolutionary distinctiveness of the host species.

Macroecological approaches to analyzing large datasets, such as the one presented here, provide a backdrop against which to consider individual case studies, faciltate an appreciation and understanding of robust patterns of host-parasite associations and extend the concept of parasite generalism to include phenomena such as parasites associating with multiple 'pockets' of related hosts. Further, the study lays out parasite group-level differences in generalism, including within-group variation, which helps navigate to sources of variation that may ultimately be explained by environmental , physiological and behavioral heterogeneity [11–14].

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) [8]. 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 ($n = 159$), bacteria ($n = 130$), helminths ($n = 461$), protozoa ($n = 93$) and viruses ($n = 97$) and 399 host species covering the orders Artiodactyla ($n = 105$), Carnivora ($n = 146$), Perissodactyla ($n = 10$) and Primate ($n = 138$). Transmission mode definitions were previously assigned from an extensive literature review [6] and cover close contact transmission ($n = 187$), complex life cycle transmission, including trophic transmission ($n = 176$), environmental transmission ($n = 423$) and vector-borne transmission ($n = 154$).

Additionally, the mammal phylogenetic supertree [15] 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) [16]. The latter was obtained using the R package *Picante* [17], and a null model in which the community data matrix (host-parasite occurrences) was randomized with the independent swap algorithm [9] 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 a trend of infecting hosts that are more related than expected by chance, indicated by negative standard effect sizes (ses) for mean pairwise phylogenetic distance between hosts (Fig. 1). However, while 94% of parasites exhibit negative ses values, only 13% of parasites have significantly negative ses values. There are no parasites that have significantly positive ses values. The proportion of significantly negative ses values varies across parasite types, with smallest to largest proportions of significantly negative ses values corresponding to bacteria, arthropods, helminths, protozoa and viruses (test for equality of proportions, p<0.001, Fig. S1, supplementary material). The proportion of significantly negative ses values does not vary across parasite transmission modes (test for equality of proportions, p=0.95). Across the full range of ses values each of the two covariates (parasite type and transmission mode) are predictive (GLM: parasite type $p < 2e-16$, transmission mode $p = 0.0005$). However, there is no significant interaction effect; transmission mode trends are consistent across parasite types ($p = 0.66$). The two covariates explain significant variation in z scores, with a pseudo R^2 value [18] 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 90^{th} percentile for z scores (i.e. the most generalist), the descending order of parasite groups is helminth ($n = 30$), arthropod ($n = 27$), bacteria ($n = 20$), virus ($n = 10$), protozoa ($n = 5$). For transmission mode, the 90^{th} percentile for z scores is dominated by environmental transmission ($n = 49$), then complex life cycle ($n = 17$), vector-borne ($n = 14$) and close contact ($n = 12$).

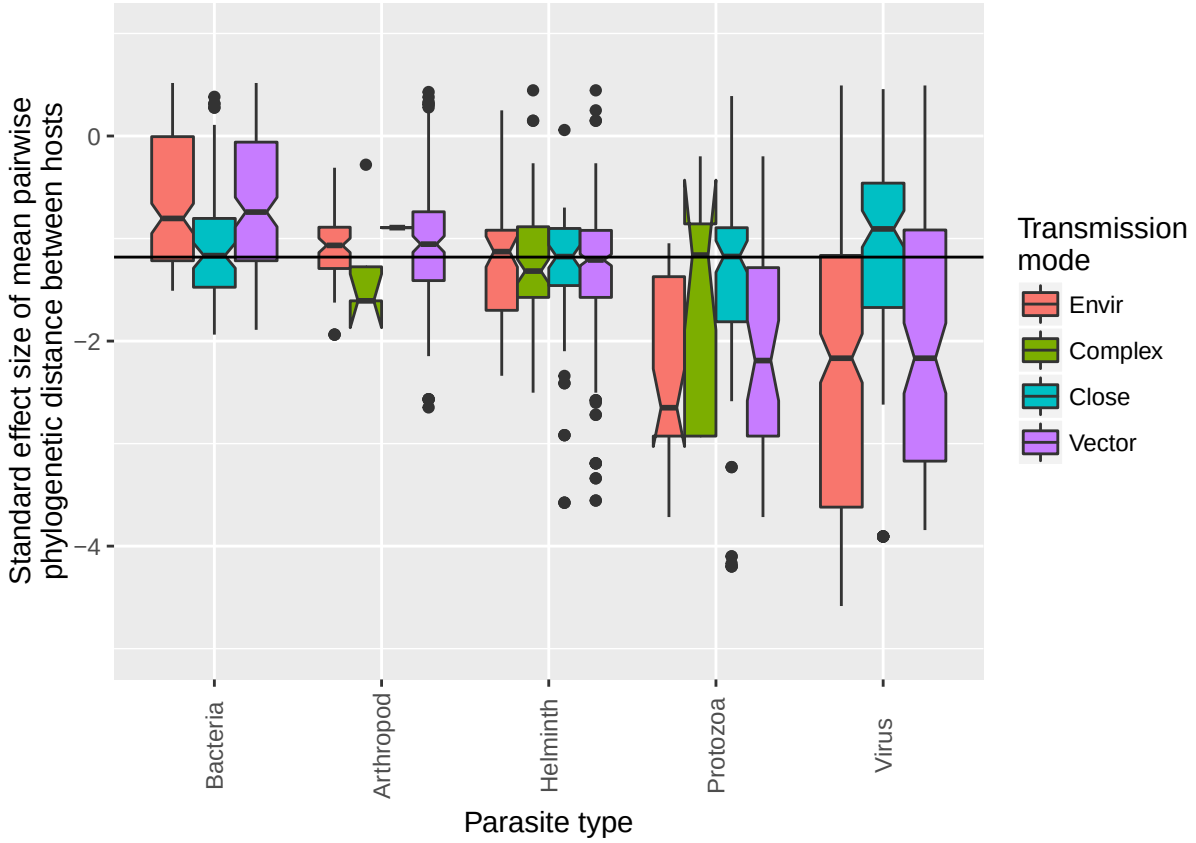


Figure 1. Boxplot of z scores in each parasite group stratified by transmission mode. Positive z scores indicate that a parasite species tends to infect host species that are more distantly related than expected by chance (under the null model). Alternatively, negative z scores indicate the opposite.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

species. Solid horizontal line indicates the global median z score.

In comparing the parasite groups that contain the greatest and fewest number of generalists (z scores>90^{th} 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.

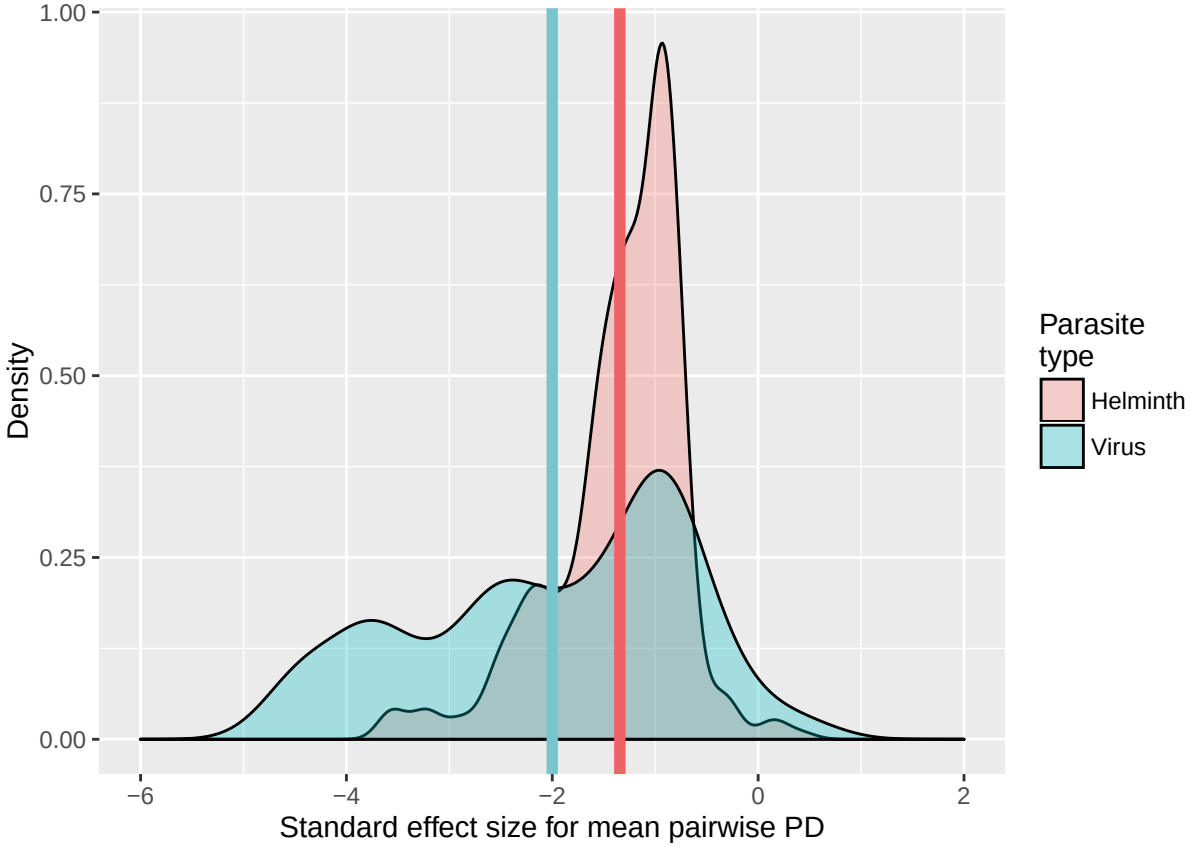


Figure 2: Probability density function of z scores for helminths (red) and viruses (blue). Helminths are more generalist on average (mean values denoted by color-coded vertical lines), but the virus group contains more extreme parasites in terms of both specialism and generalism.

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.

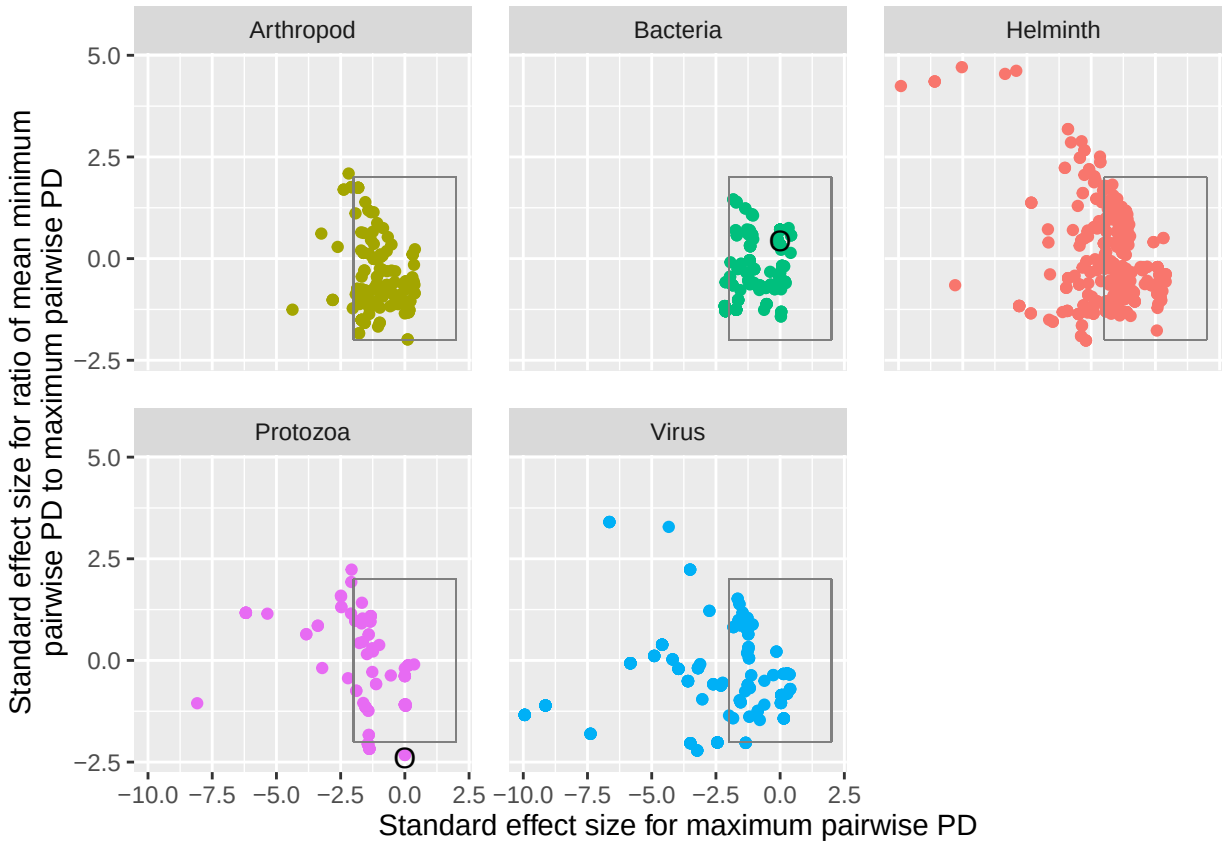


Figure 3: Correlation between x: standard effect size of maximum pairwise phylogenetic distance of a parasite's host set and y: standard effect size of the ratio of mean minimum pairwise phylogenetic distance to maximum pariwise phylogenetic distance, colored by parasite type. Circled bacterial and protozoan parasites are *Escherichia coli* and *Trypanosoma cruzi*, respectively.

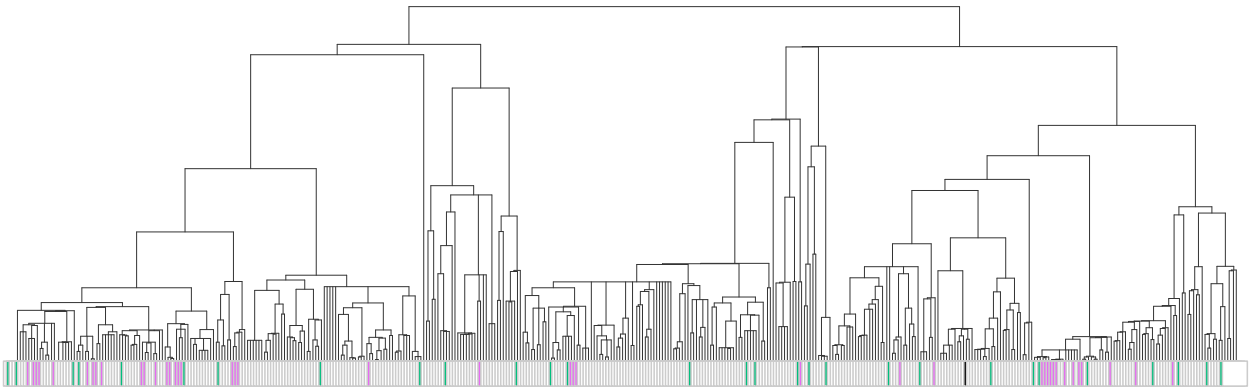
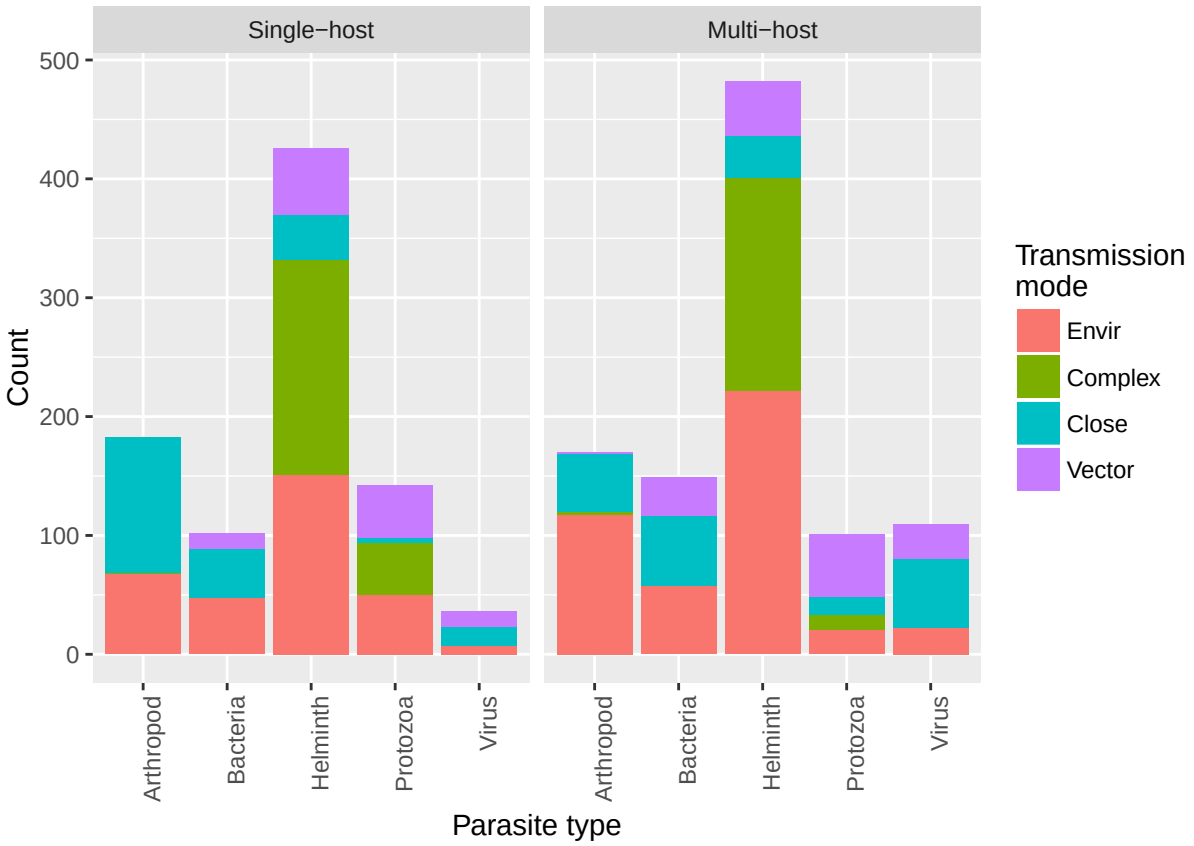
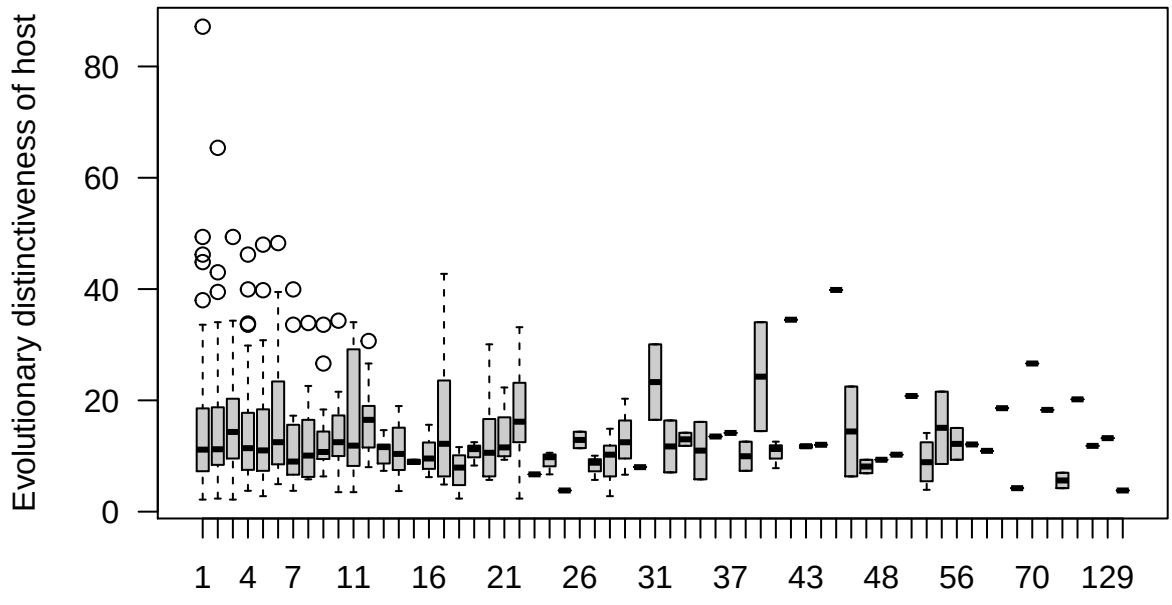


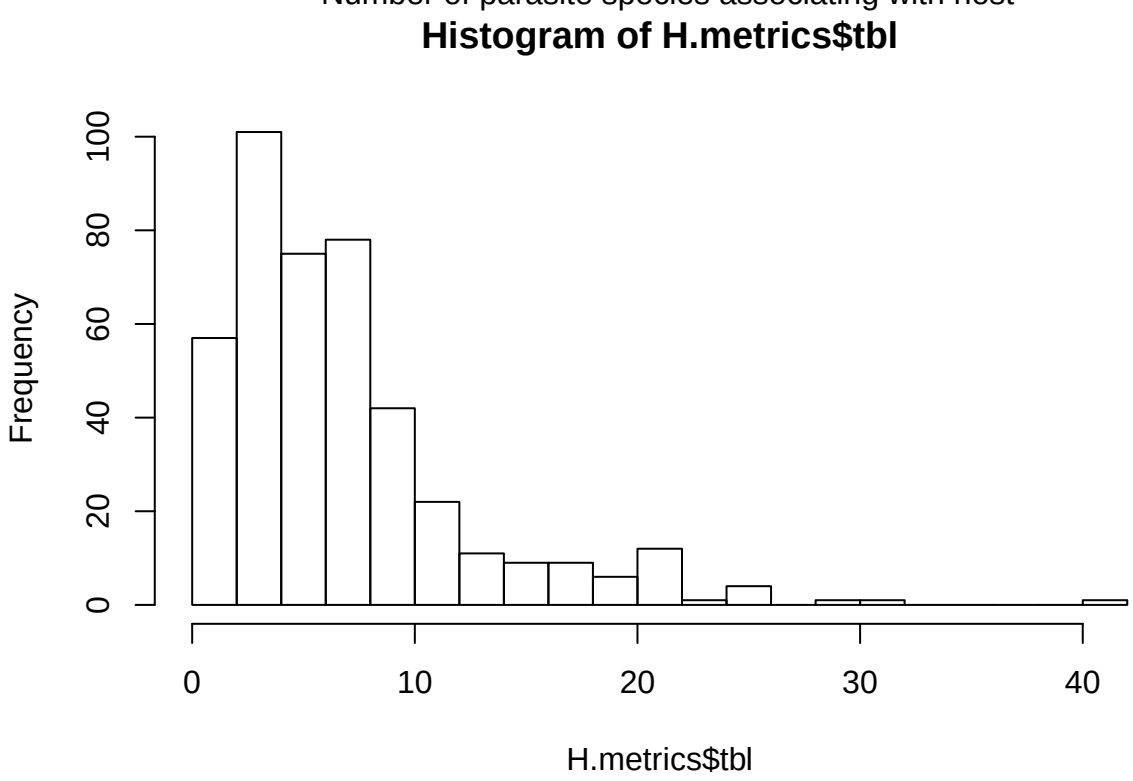
Figure 4: Examples of clumpy vs. non-clumpy distributions of parasites in host phylogenetic tree. The parasites are *Escherichia coli* and *Tryapnosoma cruzi*, colored according to their groupings in Fig. 3.

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.



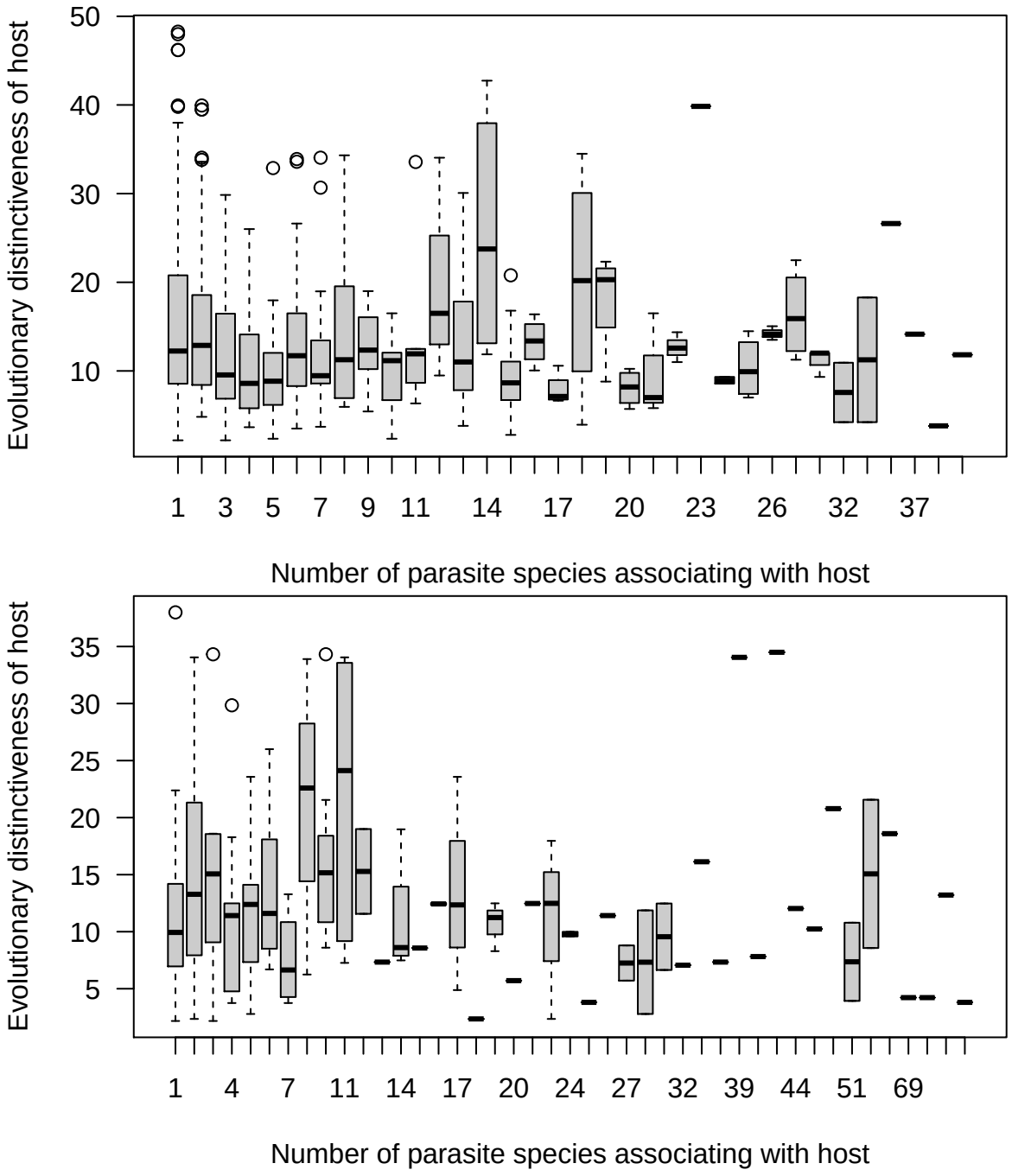
Host evolutionary distinctiveness contributes to low parasite species richness (PSR). A negative binomial generalized linear model was fitted to log transformed count data corresponding to parasite species richness. Two predictor variables were included: host evolutionary distinctiveness and number of host records in GMPD. Evolutionary distinctiveness, measured as millions of years of evolutionary change, was established from the mammal supertree data [15] using the *evol.distinct* function in R package *Picante* [17], utilizing the *equal splits* routine in which shared branches are apportioned equally among descendant lineages [19]. The number of host records in the GMPD was used to control for sampling bias. Both the number of host records and evolutionary distinctiveness were significant predictors of parasite species richness ($p < 0.001$ and $p = 0.02$, respectively). A unit increase in evolutionary distinctiveness decreased log(PSR) by 0.01, which corresponds to approximately 1 parasite fewer per million years of evolutionary distinctiveness.

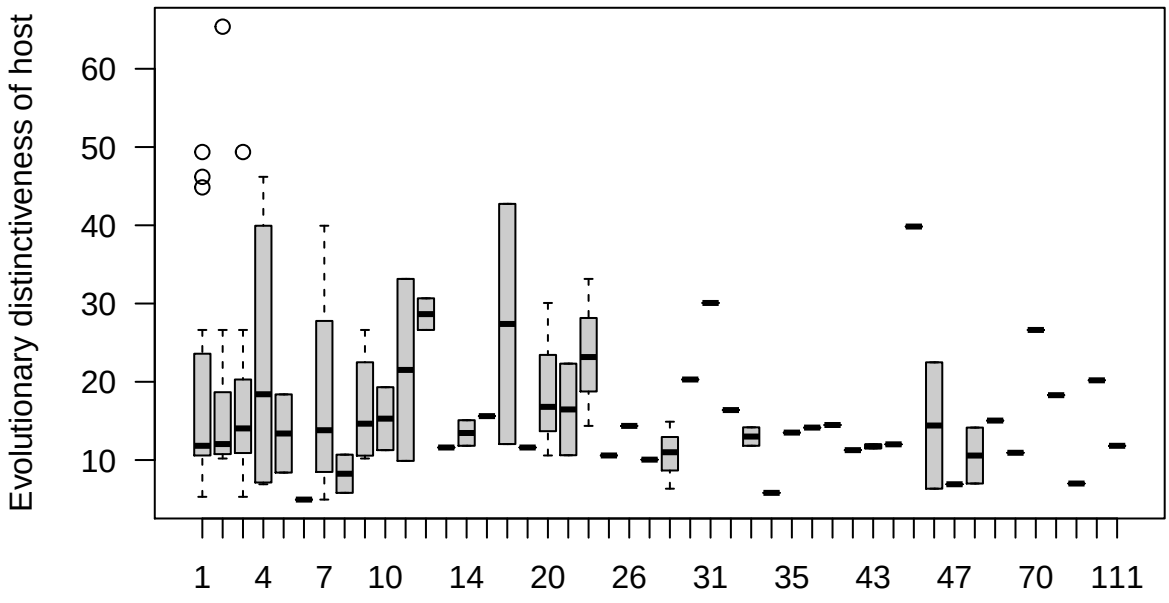


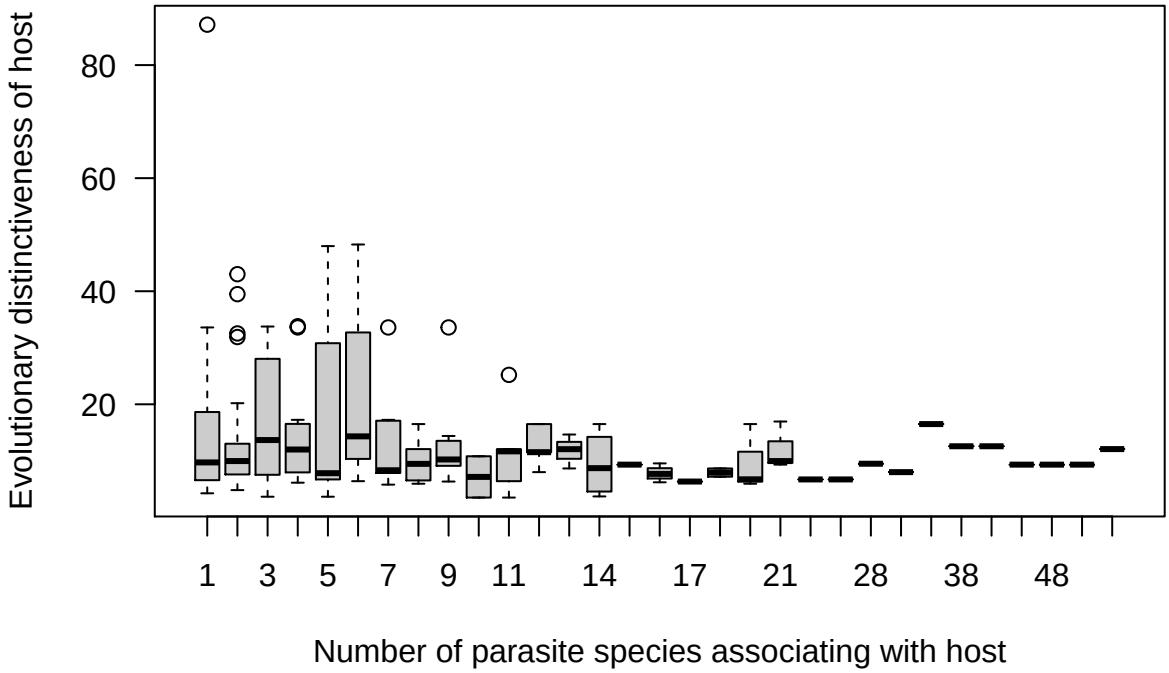



```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  H.metrics$n.para
## D = 0.51051, p-value < 2.2e-16
## alternative hypothesis: two-sided

## Loading required package: grid
```


Evolutionary distinctiveness of host
50
40
30
20
10
1
3
5
7
9
11
14
17
20
23
26
32
37
Number of parasite species associating with host
Evolutionary distinctiveness of host
35
30
25
20
15
10
5
1
4
7
11
14
17
20
24
27
32
39
44
51
69
Number of parasite species associating with host





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 [20], or maintain themselves in some host species and spillover to others, e.g. West Nile virus [21]. 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 [22], 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

Supplementary material

Looking at all multi-host parasites for single factors...

```
nri9<-nri6[nri6$ntaxa>1,]

nri9$close<-as.factor(nri9$close)
nri9$nonclose<-as.factor(nri9$nonclose)
nri9$intermediate<-as.factor(nri9$intermediate)
nri9$vector<-as.factor(nri9$vector)
nri9$para.type<-droplevels(nri9$para.type)

a1<-aov(mpd.obs.z~para.type+(close+nonclose+vector+intermediate),data=nri9)
#a1<-aov(mpd.obs.z~(close+nonclose+vector+intermediate),data=nri9)
#a1<-aov(mpd.obs.z~para.type,data=nri9)
TukeyHSD(x=a1, conf.level=0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = mpd.obs.z ~ para.type + (close + nonclose + vector + intermediate), data = nri9)
##
## $para.type
##                            diff         lwr         upr     p adj
## Bacteria-Arthropod  0.2327664  0.01068357  0.45484919 0.0345571
## Helminth-Arthropod -0.3067992 -0.50845206 -0.10514641 0.0003317
## Protozoa-Arthropod -0.6886983 -0.95451476 -0.42288175 0.0000000
```

```
## Virus-Arthropod    -0.9640829 -1.16617120 -0.76199456 0.0000000
## Helminth-Bacteria  -0.5395656 -0.73228548 -0.34684574 0.0000000
## Protozoa-Bacteria  -0.9214646 -1.18056982 -0.66235943 0.0000000
## Virus-Bacteria     -1.1968493 -1.39002477 -1.00367375 0.0000000
## Protozoa-Helminth  -0.3818990 -0.62372249 -0.14007555 0.0001670
## Virus-Helminth     -0.6572836 -0.82657589 -0.48799141 0.0000000
## Virus-Protozoa     -0.2753846 -0.51757137 -0.03319789 0.0165620
##
## $close
##           diff        lwr        upr p adj
## 1-0 -0.2508306 -0.3434853 -0.1581759 1e-07
##
## $nonclose
##           diff         lwr       upr   p adj
## 1-0 0.04310279 -0.05023757 0.1364431 0.36517
##
## $vector
##          diff       lwr       upr    p adj
## 1-0 0.2256658 0.1181006 0.3332311 4.09e-05
##
## $intermediate
##          diff         lwr       upr     p adj
## 1-0 0.1011447 -0.06901469 0.2713041 0.2438017
```

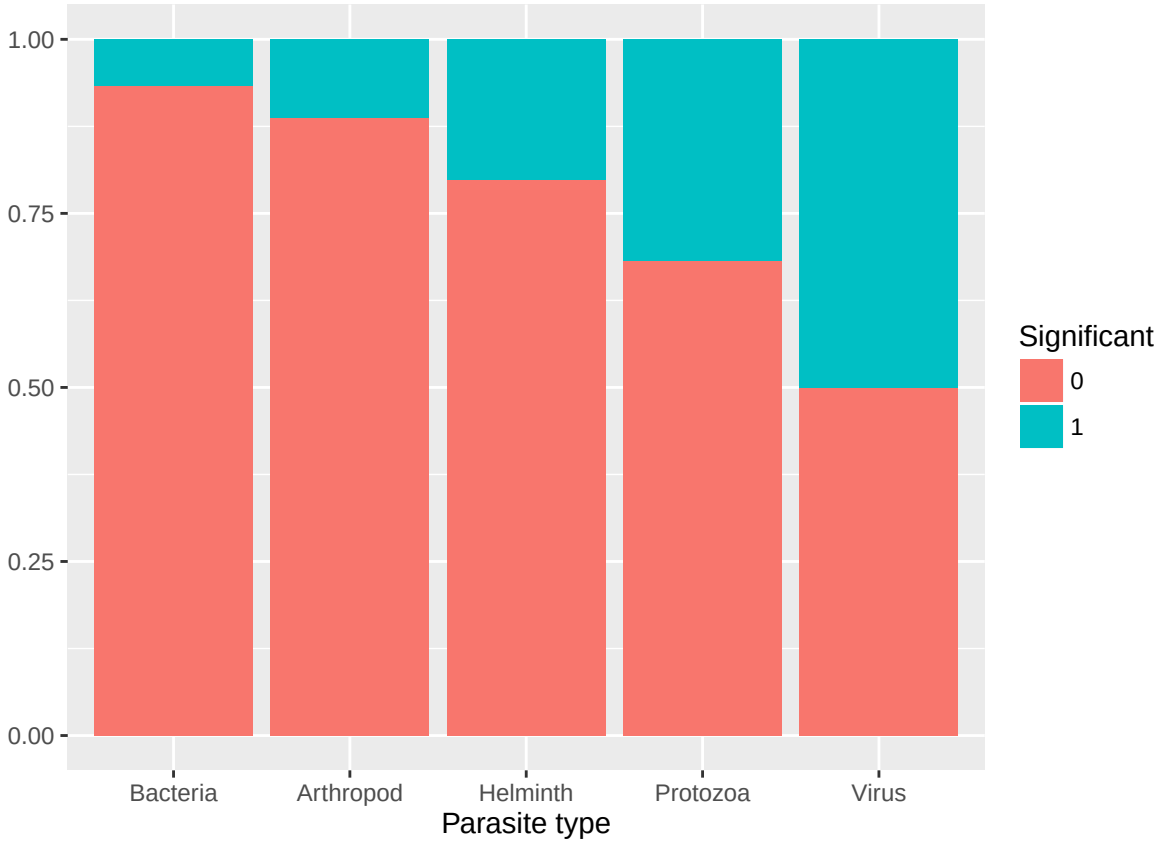
Look at proportion of significant, neg. z scores by parasite type

```
sig.z<-NULL #note all significant z scores are negative (checked independently)
for (i in unique(nri9$para.type)){
  tmp<-subset(nri9,para.type==i)
  sig.z<-rbind(sig.z,c(i,dim(tmp)[1],length(which(tmp$mpd.obs.p<0.05))))
}
sig.z<-as.data.frame(sig.z)
names(sig.z)<-c("para.type","n","x")
sig.z$n<-as.integer(as.character(sig.z$n))
sig.z$x<-as.integer(as.character(sig.z$x))
sig.z$y<-sig.z$n-sig.z$x
prop.test(sig.z$x,sig.z$n)
```

```
##
##  5-sample test for equality of proportions without continuity
##  correction
##
## data:  sig.z$x out of sig.z$n
## X-squared = 206.19, df = 4, p-value < 2.2e-16
## alternative hypothesis: two.sided
## sample estimates:
##     prop 1     prop 2     prop 3     prop 4     prop 5
## 0.06666667 0.20243902 0.11312217 0.50123457 0.31818182
```

```
sig<-NULL
for (i in 1:dim(sig.z)[1]){
  for(j in 1:sig.z[i,"x"]){
    sig<-rbind(sig,c(as.character(sig.z$para.type[i]),1))
```

```
  }
  for(j in 1:sig.z[i,"y"]){
    sig<-rbind(sig,c(as.character(sig.z$para.type[i]),0))
  }
}
sig<-as.data.frame(sig)
names(sig)<-c("para.type","sig")
sig$para.type<-factor(sig$para.type,levels=c("Bacteria","Arthropod","Helminth","Protozoa","Virus"))
ggplot(sig,aes(x=para.type,fill=sig))+geom_bar(position="fill")+scale_fill_discrete(name="Significant")+
```



Look at proportion of significant, neg. z scores by transmission mode

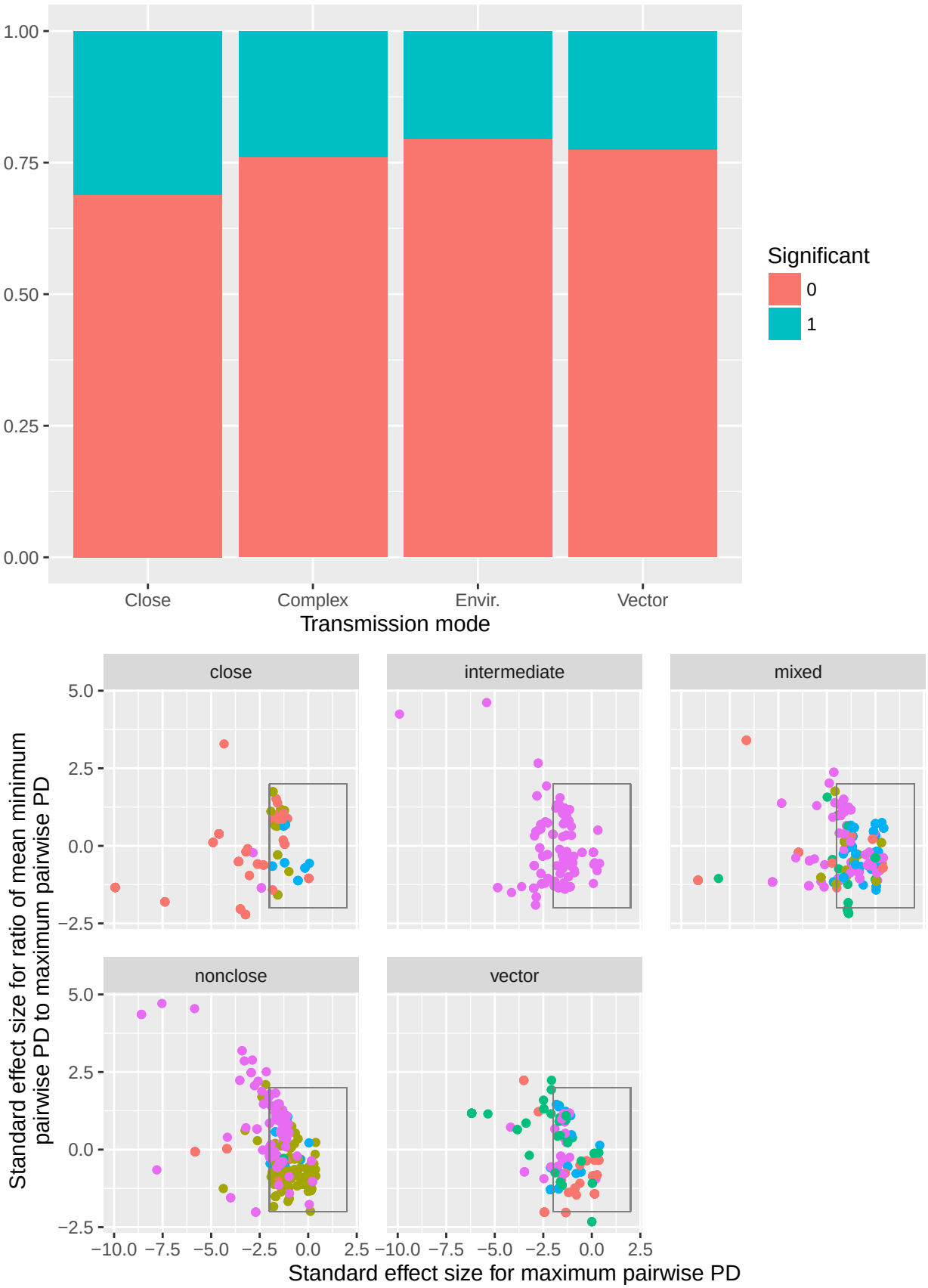
```
sig.z<-NULL #note all significant z scores are negative (checked independently)
for (i in unique(nri.flat$tmode)){
  tmp<-subset(nri.flat,tmode==i)
  sig.z<-rbind(sig.z,c(i,dim(tmp)[1],length(which(tmp$mpd.obs.p<0.05))))
}
sig.z<-as.data.frame(sig.z)
names(sig.z)<-c("tmode","n","x")
sig.z$n<-as.integer(as.character(sig.z$n))
sig.z$x<-as.integer(as.character(sig.z$x))
sig.z$y<-sig.z$n-sig.z$x
prop.test(sig.z$x,sig.z$n)
```

```
##
##  4-sample test for equality of proportions without continuity
```

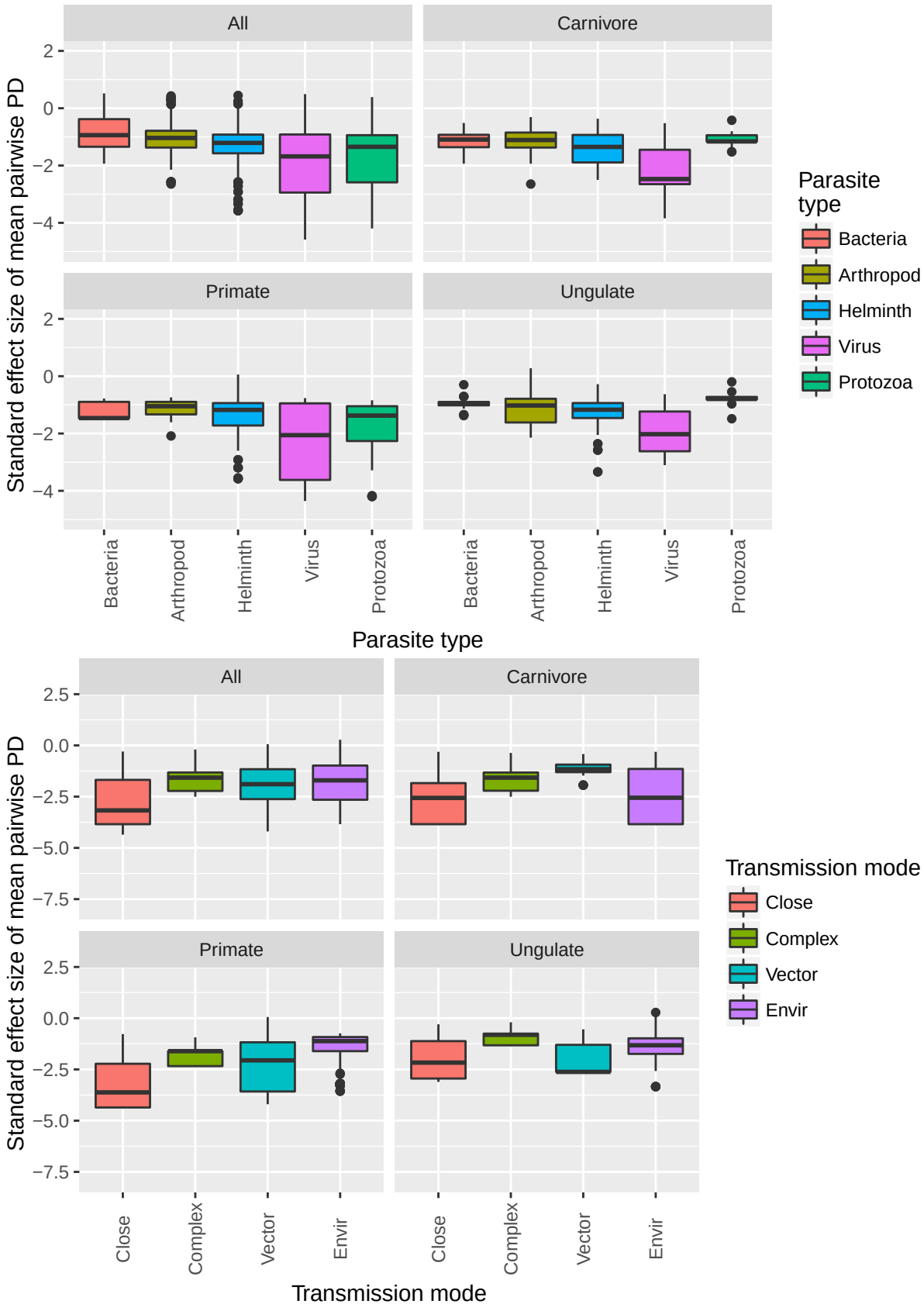
```
##  correction
##
## data:  sig.z$x out of sig.z$n
## X-squared = 22.72, df = 3, p-value = 4.618e-05
## alternative hypothesis: two.sided
## sample estimates:
##    prop 1    prop 2    prop 3    prop 4
## 0.3105096 0.2045728 0.2244898 0.2389381
```

```
sig<-NULL
for (i in 1:dim(sig.z)[1]){
  for(j in 1:sig.z[i,"x"]){
    sig<-rbind(sig,c(as.character(sig.z$tmode[i]),1))
  }
  for(j in 1:sig.z[i,"y"]){
    sig<-rbind(sig,c(as.character(sig.z$tmode[i]),0))
  }
}
sig<-as.data.frame(sig)
names(sig)<-c("tmode","sig")
sig$tmode<-as.character(sig$tmode)
sig$tmode[which(sig$tmode=="close")]<-"Close"
sig$tmode[which(sig$tmode=="intermediate")]<-"Complex"
sig$tmode[which(sig$tmode=="nonclose")]<-"Envir."
sig$tmode[which(sig$tmode=="vector")]<-"Vector"

#sig$para.type<-factor(sig$para.type,levels=c("Bacteria","Arthropod","Helminth","Protozoa","Virus"))
ggplot(sig,aes(x=tmode,fill=sig))+geom_bar(position="fill")+xlab("Transmission mode")+ylab("")+scale_fi
```


1.00
0.75
0.50
0.25
0.00
Close
Complex
Envir.
Vector
Transmission mode
Significant
0
1
close
intermediate
mixed
nonclose
vector
5.0
2.5
0.0
−2.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
Standard effect size for ratio of mean minimum pairwise PD to maximum pairwise PD
Standard effect size for maximum pairwise PD

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.


All
Carnivore
Primate
Ungulate
Standard effect size of mean pairwise PD
2
0
−2
−4
Bacteria
Arthropod
Helminth
Virus
Protozoa
Parasite type
Parasite type
Bacteria
Arthropod
Helminth
Virus
Protozoa
2.5
0.0
−2.5
−5.0
−7.5
Close
Complex
Vector
Envir
Transmission mode
Transmission mode
Close
Complex
Vector
Envir

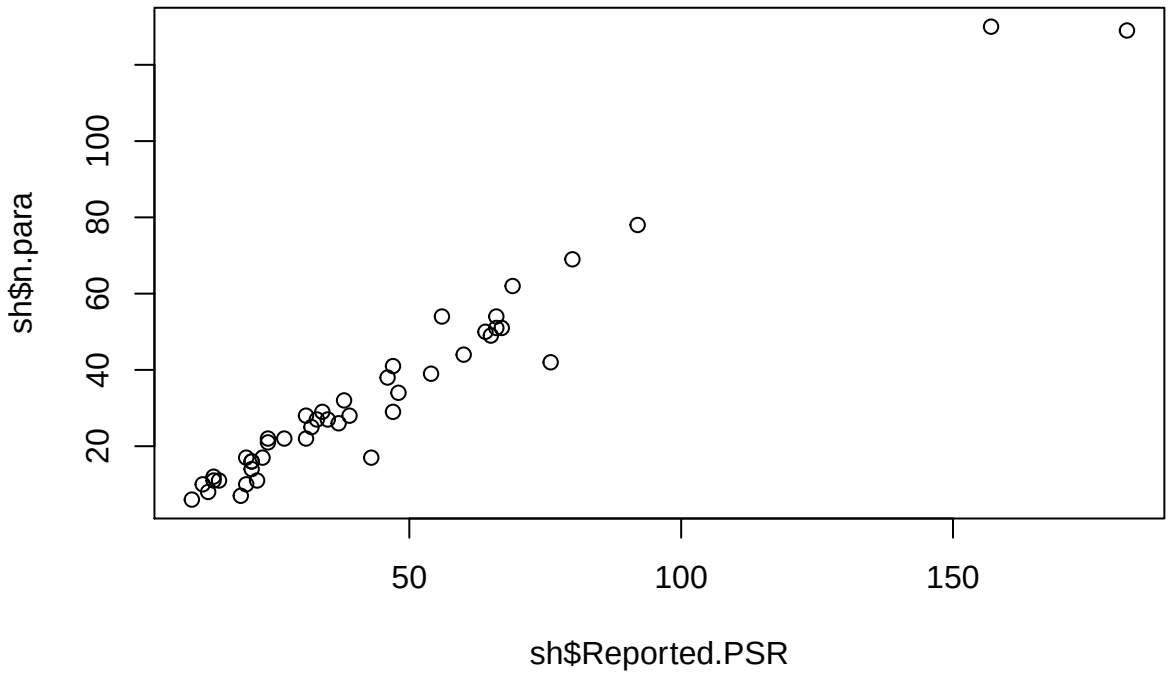
Examination of (multi-host) parasites with unique transmission modes shows there is not interaction between parasite type and transmission mode

```
idx.1<-which(nri6$ntaxa>1)
idx.2<-which(nri6$n.modes==1)
idx<-intersect(idx.1,idx.2)
u<-nri6[idx,]
u$tmode<-"tbd"
for (i in 1:dim(u)[1]){
  if (u$close[i]==1){u$tmode[i]<-"close"}
  if (u$nonclose[i]==1){u$tmode[i]<-"nonclose"}
  if (u$intermediate[i]==1){u$tmode[i]<-"intermediate"}
  if (u$vector[i]==1){u$tmode[i]<-"vector"}
}
u$tmode<-as.factor(u$tmode)
u$para.type<-droplevels(u$para.type)
a<-aov(mpd.obs.z~para.type*tmode,data=u)
summary(a)
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## para.type          4  153.9   38.46  51.287  < 2e-16 ***
## tmode              3   51.5   17.16  22.881 2.38e-14 ***
## para.type:tmode    9   51.7    5.74   7.658 5.96e-11 ***
## Residuals       1046  784.5    0.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparison with Shan's paper

```
sh<-read.csv("huang_carnPSR.csv",header=T)# Table S1
#look by host order
d.carn<-subset(d,Group=="carnivores")
H.metrics<-as.data.frame(table(d.carn$HostCorrectedName))
names(H.metrics)<-c("host.name","n.records")
H.metrics$n.para<-0
for (i in 1:dim(H.metrics)[1]){
  p.set<-unique(d.carn$ParasiteCorrectedName[which(d.carn$HostCorrectedName==H.metrics$host.name[i])])
  H.metrics$n.para[i]<-length(p.set)
}
library(picante)
mtree2[[1]]$tip.label<-orig.tip.label
my.ed<-evol.distinct(mtree2[[1]],type="equal.splits")
names(my.ed)<-c("host.name","ed")
H.metrics<-merge(H.metrics,my.ed)
H.metrics$singleton<-ifelse(H.metrics$n.para==1,1,0)
H.order.binom.tmp<-subset(d.carn,select=c("HostCorrectedName","Group"))
H.order.binom.tmp<-H.order.binom.tmp[!duplicated(H.order.binom.tmp),]
names(H.order.binom.tmp)<-c("host.name","order")
H.metrics<-merge(H.metrics,H.order.binom.tmp)
H.metrics$host.name<-gsub("_"," ",H.metrics$host.name)
names(sh)[which(names(sh)=="Host.species")]<-"host.name"
sh<-merge(sh,H.metrics,by="host.name")
plot(sh$Reported.PSR,sh$n.para)
```




```
tmp<-lm(sh$Reported.PSR~sh$ed)
#use terminal branch length
terms<-mtree$edge[,2]<=Ntip(mtree)
terminal.edges<-as.data.frame(cbind(mtree$tip.label[terms],mtree$edge.length[terms]))
names(terminal.edges)<-c("host.name","tbl")
terminal.edges$tbl<-as.numeric(as.character(terminal.edges$tbl))
terminal.edges$host.name<-gsub("_"," ",terminal.edges$host.name)
sh<-merge(sh,terminal.edges,by="host.name")
```

test phylo

```
data(bird.families)
summary(bird.families)
```

```
##
## Phylogenetic tree: bird.families
##
##   Number of tips: 137
##   Number of nodes: 135
##   Branch lengths:
##     mean: 7.413653
##     variance: 38.49237
##     distribution summary:
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##    0.10    1.15    9.10   11.50   27.00
##   No root edge.
##   First ten tip labels: Struthionidae
##                         Rheidae
##                         Casuariidae
##                         Apterygidae
##                         Tinamidae
```

```
##                         Cracidae
##                         Megapodiidae
##                         Phasianidae
##                         Numididae
##                         Odontophoridae
##   No node labels.
```

```
Ntip(bird.families)
```

```
## [1] 137
```

```
Nnode(bird.families)
```

```
## [1] 135
```

```
Nedge(bird.families)
```

```
## [1] 271
```


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

8. 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.

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

10. 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.

11. Ezenwa, V. O. 2004 Host social behavior and parasitic infection: a multifactorial approach. *Behavioral Ecology* **15**, 446–454. (doi:10.1093/beheco/arh028)

12. Lee, K. A., Wikelski, M., Robinson, W. D., Robinson, T. R. & Klasing, K. C. 2008 Constitutive immune defences correlate with life-history variables in tropical birds. *Journal of Animal Ecology* **77**, 356–363.

(doi:10.1111/j.1365-2656.2007.01347.x)

13. Poulin, R. & Forbes, M. R. 2012 Meta-analysis and research on host–parasite interactions: past and future. *Evolutionary Ecology* **26**, 1169–1185. (doi:10.1007/s10682-011-9544-0)

14. Stein, A., Gerstner, K. & Kreft, H. 2014 Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters* **17**, 866–880. (doi:10.1111/ele.12277)

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

16. 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.

17. 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.

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

19. Redding, D. W. & Mooers, A. Ø. 2006 Incorporating Evolutionary Measures into Conservation Prioritization. *Conservation Biology* **20**, 1670–1678. (doi:10.1111/j.1523-1739.2006.00555.x)

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

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

22. 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.