Supplemental Information

The ecology of zoonotic parasites in the Carnivora

Barbara A. Han1, Adrian A. Castellanos1, John Paul Schmidt2,3, Ilya R. Fischhoff1, John M. Drake2,3

1 Cary Institute of Ecosystem Studies Box AB Millbrook, NY 12545

2 Odum School of Ecology, University of Georgia 140 E. Green Street Athens, GA 30602

3 Center for the Ecology of Infectious Diseases, University of Georgia 203 D.W. Brooks Drive Athens, GA 30602

Correspondence: hanb@caryinstitute.org (B.A. Han).

**Box S1. Methods for phylogenetic clustering of parasitic transmission modes described in “Zoonotic hosts among the Carnivora.”**

To determine whether parasite transmission modes clustered within the carnivore tree, we used the phyloclust function in the R package RRphylo [(1)](https://paperpile.com/c/dodVGH/MO8v) with R version 4.0.0 (2). Our phylogenetic tree data were the 10,000 birth-death node-dated trees subsetted to the Carnivora from [(3)](https://paperpile.com/c/dodVGH/sw2V). Data for parasite associations with each carnivore species was obtained from a previous study (4) and updated according to associations found in the Global Infectious Diseases and Epidemiology Network (GIDEON; www.gideononline.com) and the literature. Transmission modes (used as our “state” data for the clustering analysis) for each parasite were obtained from the Global Mammal Parasite Database ([(5)](https://paperpile.com/c/dodVGH/sfbq); see Table S1 for these data). We report the mean *p* values of 10,000 runs for each transmission mode in the main text. R code to perform this analysis is given below:

library(RRphylo)

library(ape)

library(tidyverse)

TREES <- read.nexus(file = "tree-pruner-b0c4fb63-d1aa-4771-9271-04e573ccf185/output.nex") #This is a set of pruned trees from http://vertlife.org/phylosubsets/ as described above

#carnivoreGMPD read in below is Table S1 described below

GMPD <- read.csv("carnivoreGMPD.csv", header = T) %>%

mutate(species = gsub(" ", "\_", species)) %>%

group\_by(species) %>%

summarise(

close = as.character(as.numeric(sum(close, na.rm = T) > 0)),

nonclose = as.character(as.numeric(sum(nonclose, na.rm = T) > 0)),

vector = as.character(as.numeric(sum(vector, na.rm = T) > 0)),

intermediate = as.character(as.numeric(sum(intermediate, na.rm = T) > 0)))

setdiff(GMPD$species, TREES[[1]]$tip.label)

EXTRA <- cbind.data.frame(species = setdiff(TREES[[1]]$tip.label, GMPD$species),

close = "0",

nonclose = "0",

vector = "0",

intermediate = "0")

GMPD <- rbind(GMPD, EXTRA)

STATE <- GMPD$close

names(STATE) <- GMPD$species

CLUSC <- sapply(TREES, function(i) phyloclust(tree = i, state = STATE, focal = "1", nsim = 100)$p)

STATE <- GMPD$nonclose

names(STATE) <- GMPD$species

CLUSN <- sapply(TREES, function(i) phyloclust(tree = i, state = STATE, focal = "1", nsim = 100)$p)

STATE <- GMPD$vector

names(STATE) <- GMPD$species

CLUSV <- sapply(TREES, function(i) phyloclust(tree = i, state = STATE, focal = "1", nsim = 100)$p)

STATE <- GMPD$intermediate

names(STATE) <- GMPD$species

CLUSI <- sapply(TREES, function(i) phyloclust(tree = i, state = STATE, focal = "1", nsim = 100)$p)

**Box S2. Methods for boosted regression trees reanalysis described in “Traits related to omnivory distinguish zoonotic hosts from non-hosts.”**

We used the gbm R package [(5)](https://paperpile.com/c/dodVGH/JKT4) in R version 4.0.0 (2) to investigate the relationship between previously published species trait data [(6; Table S3)](https://paperpile.com/c/dodVGH/Irl2) and zoonotic host status data [(7, Tables S1 and S3)](https://paperpile.com/c/dodVGH/VvIZ). Briefly, zoonotic host status and the number of zoonoses associated with each species were determined from the literature and cross-checked against GIDEON. Trait data were compiled from various existing databases or calculated from spatial data layers (Table S4). We used a hurdle model process that first fit a classification model to determine zoonotic host status followed by a regression model of all non-zero data regarding the number of zoonoses associated with each species ((8); see provided R code below or in RMarkdown format at https://doi.org/10.25390/caryinstitute.c.5351444.v1 for more information). We used a Bernoulli error distribution for the classification model and a Poisson error distribution for modeling abundance of zoonoses. Model parameterization was done using a grid search of all appropriate parameters. This parameter search was limited to learning rates of 0.0001, 0.001, and 0.01, maximum depth of 2, 3, and 4, and number of minimum observations per node of 2, 3, 4, and 5. The parameter combination with the best deviance curve and highest evaluation statistics was evaluated using 10 bootstrap iterations. The number of zoonoses model showed consistently low pseudo R2 from this bootstrap evaluation, so we only report results from the classification model of zoonotic host status.

**Supplemental References**

S1. [Castiglione S, Tesone G, Piccolo M, Melchionna M, Mondanaro A, Serio C, Di Febbraro M, Raia P. 2018. A new method for testing evolutionary rate variation and shifts in phenotypic evolution. Methods Ecol Evol 9:974–983.](http://paperpile.com/b/dodVGH/MO8v)

S2. R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

S3. [Upham NS, Esselstyn JA, Jetz W. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol 17:e3000494.](http://paperpile.com/b/dodVGH/sw2V)

S4. [Stephens PR, Pappalardo P, Huang S, Byers JE, Farrell MJ, Gehman A, Ghai RR, Haas SE, Han B, Park AW, Schmidt JP, Altizer S, Ezenwa VO, Nunn CL. 2017. Global Mammal Parasite Database version 2.0. Ecology 98:1476.](http://paperpile.com/b/dodVGH/sfbq)

S5. [Greenwell B, Boehmke B, Cunningham J, GBM Developers. 2020. gbm: Generalized Boosted Regression Models. R package version 2.1.8. Comprehensive R Archive Network (CRAN).](http://paperpile.com/b/dodVGH/JKT4)

S6. [Fischhoff IR, Castellanos AA, Rodrigues JPGLM, Varsani A, Han BA. 2021. Predicting the zoonotic capacity of mammal species for SARS-CoV-2. bioRxiv 2021.02.18.431844.](http://paperpile.com/b/dodVGH/Irl2)

S7. [Han BA, Kramer AM, Drake JM. 2016. Global Patterns of Zoonotic Disease in Mammals. Trends Parasitol 32:565–577.](http://paperpile.com/b/dodVGH/VvIZ)

S8.   Cragg, J.G., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39:829-844.