R-script

All data manipulation and analysis (including functions and packages) can be found below.

setwd("…")

library(Hmisc)

library(corrplot)

library(caper)

library(picante)

library(phytools)

library(geiger)

library(ade4)

library(spTimer)

library(vegan)

library(ggplot2)

library(betareg)

library(lme4)

library(car)

####Import data####

#With prevalence data at the population level

dadosPr<-read.csv("dados\_spp\_prev.csv",header = T, sep=";",row.names = 1)

#Test whether the mean size of host species is correlated with prevalence at the population level

#Generate prevalence at the population level

dadosPr2<-dadosPr

dadosPr2[,6:31]<-dadosPr2[,6:31]/dadosPr2[,4]

#Correlate with the mean size of the species

corrplot(cor(dadosPr2[,6:31],dadosPr2[,5]), method = "number")

#Five species of parasites show a positive correlation (r>0.5) with mean host size.

#This correlation exists for two parasites that occurred only in one host species,

#two that occurred in two host species (Oxyascaris\_oxyascaris being in

#the two largest frog species) and one parasite in three host species (less correlation)

#Test whether the proportion of infected individuals in the frog population is correlated with the mean size

#Gross data values

dadosPr2$Infected<-c(7,18,18,19,18,16,20,27,22,9,9)

dadosPr2$PropInf<-dadosPr2$Infected/dadosPr2$N

cor.test(dadosPr2$Size,dadosPr2$PropInf)

plot(dadosPr2$Size,dadosPr2$PropInf)

#Despite the high ratio (r > 0.5), there is a variation within the proportion of infected individuals per species

#that ends up taking away the significance

#Anyway, considering the values from the perspective of the parasite,

#the correlation is low, being non-existent in most parasites

#####Organizing the data#####

#Parasite Metrics Data

dadosparasito<-read.csv("dados\_parasitos.csv",header = T, row.names = 1, sep=";")

dadospara2<-dadosparasito[,c("total\_parasitos","total\_hospedeiros", "hospedeiros\_infec","prevInf","imi")]

dadospara2$prevComm<-dadospara2$hospedeiros\_infec/213

#Download the Pyron tree.

mytrees<-read.tree("amphibians.tree")

#Work first with presence and absence, using the prevalence matrix

dadosPr[,c("Arboreal","Terrestrial","Semiaquatic", "N", "Size")]->tratPr

ncol(dadosPr)

#31

dadosPr[,6:31]->sppPr

#Remove parasites without occurrence

#colSums(spp)

#spp$Phyalopteridae\_gen.3 <- NULL

#Prepare the variables for distance calculation according to Pavoine et al 2012

colnames(tratPr)

tabHabit<-prep.binary(tratPr[,1:3], col.blocks = 3)

tabSize<-data.frame(tratPr[,5])

ktab\_tr<-ktab.list.df(list(tabHabit, tabSize))

#Modified gower distance matrix

dist.func<-dist.ktab(ktab\_tr, type=c("B","Q"), option="scaledBYrange")

#scaledBYrange = either "scaledBYrange" if the quantitative variables must be scaled by their range

as.matrix(dist.func)[1:4,1:4]

is.euclid(dist.func)#false it is not an object with Euclidean distance

#Generate a functional dendrogram for creating a functional tree

dend\_func<-hclust(dist.func, method = "average", members = NULL)

dend\_func\_phy<-as.phylo(dend\_func)

plot(dend\_func\_phy)

#Compare the species name with that of the functional tree

data2<-match.phylo.data(dend\_func\_phy,sppPr)

sppPrFunc<-data2$data

sppPrFunc<-as.matrix(sppPrFunc)

#Pruning the phylogenetic tree

#Comparing the names of the phylo with that of the community

data<-match.phylo.data(mytrees,sppPr)

SppPrPhy<-data$data

mytree1<-data$phy

plot(mytree1)

#####Phylogenetic signal#####

x<-match.phylo.data(mytrees,tratPr)

sinal<-x$data

mytreesinal<-x$phy

plot(mytreesinal)

phylosignal(sinal$Size,mytreesinal, reps=1000, checkdata = F)

#Size has no phylogenetic signal

# K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P PIC.variance.Z

#1 0.5858464 0.09368626 0.07127529 0.9010989 1.320152

sinal$Scientific<-row.names(signal)

HabSignal <- comparative.data(mytreesinal,sinal,Scientific)

ArborealSinal<- phylo.d(HabSignal, binvar=Arboreal)

plot(ArborealSinal)

print(ArborealSinal)

#Anurans with arboreal habits have a high phylogenetic signal, even bordering on the Brownian model

#Estimated D : -6.949124

#Probability of E(D) resulting from no (random) phylogenetic structure : 0

#Probability of E(D) resulting from Brownian phylogenetic structure : 1

TerrestrialSinal<- phylo.d(HabSignal, binvar=Terrestrial)

plot(TerrestrialSinal)

print(TerrestrialSinal)

#No phylogenetic signal

#Estimated D : 1.784289

#Probability of E(D) resulting from no (random) phylogenetic structure : 0.639

#Probability of E(D) resulting from Brownian phylogenetic structure : 0.198

SemiaquaticSinal<- phylo.d(HabSignal, binvar=Semiaquatic)

plot(SemiaquaticSinal)

print(SemiaquaticSinal)

#No phylogenetic signal

#Estimated D : 3.330994

#Probability of E(D) resulting from no (random) phylogenetic structure : 0.765

#Probability of E(D) resulting from Brownian phylogenetic structure : 0.143

#Only one attribute has a phylogenetic signal, while the others have no signal

#Test whether the phylogenetic and functional distances of anurans are correlated

DC<-match.phylo.comm(mytree1,cophenetic(dend\_func\_phy))

DC2<-match.phylo.data(mytree1,DC$comm)

cor.test(as.dist(cophenetic(mytree1)),as.dist(DC2$data))

#r =0.01

plot(as.dist(cophenetic(mytree1)),as.dist(DC2$data))

mantel(as.dist(cophenetic(mytree1)),as.dist(DC2$data),method="pearson", permutations = 1000)

#Without significance, we can use

#No correlation at all

plot(as.dist(cophenetic(mytree1)),as.dist(DC2$data))

#Considering that only one variable (binomial by signal) has a phylogenetic signal, the data seem

#to be adequate to test the hypotheses of functional and phylogenetic specificity

#Now extract the mpd as specificity of each parasite

####Functional Specificity####

sppPrFunc<-t(sppPrFunc)

#Functional only with occurrence

funcStd<-ses.mpd(sppPrFunc,cophenetic(dend\_func\_phy),null.model = "taxa.labels",

abundance.weighted = F,runs=999,iterations=1000 )

#write.table(funcStd,"funcmpd.txt")

#Parasites that infect only one host do not have functional values

#We consider zero as proposed by Ellis et al (2020)

funcStd[is.na(funcStd)] <- 0

dadospara2$NHost<-funcStd$ntaxa

#dadospara2$FuncMpdOc<-funcStd$mpd.obs

dadospara2$FuncNRI<-funcStd$mpd.obs.z\*-1

####Phylogenetic specificity####

SppPrPhy<-as.matrix(SppPrPhy)

SppPrPhy<-t(SppPrPhy)

#Occurrence only

phyloStd<-ses.mpd(SppPrPhy,cophenetic(mytree1),null.model = "taxa.labels",

abundance.weighted = F,runs=999,iterations=1000 )

#write.table(phyloStd,"phylompd.txt")

phyloStd[is.na(phyloStd)] <- 0

#dadospara2$PhyloMpdOc<-phyloStd$mpd.obs

dadospara2$PhyloNRI<-phyloStd$mpd.obs.z\*-1

#Generate a correlation matrix to see some patterns

Cor<-cor(dadospara2)

corrplot(Cor, method="number",type="lower")

#There are interrogations in some variables because of the presence of NA

#Considering the prevalence at the community level, it is positively correlated with the number of host species

plot(dadospara2$prevComm,dadospara2$NHost)

#From the graph, it is possible to observe an outlier, which is a Family (Cosmocercidae) and not a species.

#Possibly there is more than one species within this sample

#Generates a subset with only genus and species

dadospara3<-dadospara2[-c(6), ]

Cor2<-cor(dadospara3)

#View a matrix

corrplot(Cor2, method="number",type="lower")

#Community-level prevalence appears to be affected by number of hosts and functional specificity

#Now only with parasites with more than two host species

plot(dadospara3$NHost,dadospara3$prevComm)

plot(dadospara3$FuncMpdOc,dadospara3$prevComm)

####Linear models####

#write.table(dadospara3,"modelsComm.txt")

dadospara3<-read.table("modelsComm.txt", header =T, row.names = 1)

####Prevalence at community level####

hist(dadospara3$prevComm)

#Would be a gaussian lm

#We chose to use betareg as the prevalence of all parasites was calculated by the same host number

mod1<-betareg(prevComm~NHost, data=dadospara3)

plot(prevComm~NHost,data=dadospara3)

abline(lm(prevComm~NHost, data=dadospara3))

confintmod<-confint(mod1)

summary(mod1)

plot(mod1)

#Based on the distribution of waste, cook distance and laverage, the model is valid

plotMod<-ggplot(dadospara3, aes(NHost, prevComm)) + geom\_point()+

geom\_smooth(method="lm")+

xlab("Number of host species") + ylab("Prevalence") +

theme(axis.text=element\_text(size=14),axis.title=element\_text(size=14,face="bold"),

panel.background = element\_rect(fill = "transparent"), # bg of the panel

plot.background = element\_rect(fill = "transparent", color = NA), # bg of the plot

panel.grid.major = element\_blank(), # get rid of major grid

panel.grid.minor = element\_blank(), # get rid of minor grid

legend.background = element\_rect(fill = "transparent"), # get rid of legend bg

legend.box.background = element\_rect(fill = "transparent") # get rid of legend panel bg

,axis.line = element\_line(size = 1.5, linetype = "solid"))

print(plotMod)

#Using data from all parasites (including those that occur only in one host species, the prevalence at the

#community level is directly proportional to species richness.

#That is, parasites with greater host range tend to have higher prevalence

#Only with parasites with more than one host

dadospara4<-subset(dadospara3, NHost > 1)

hist(dadospara4$prevComm, breaks = 20)

corrplot(cor(dadospara4), method="number",type="lower")

plot(dadospara4$prevComm,dadospara4$FuncSESMpdOc)

dadospara4$FuncNRI<-dadospara4$FuncSESMpdOc\*-1

dadospara4$PhyloNRI<-dadospara4$PhyloSESMpdOc\*-1

#We chose to use betareg as the prevalence of all parasites was calculated by the same host number

mod2<-betareg(prevComm~NHost+FuncNRI+PhyloNRI, data=dadospara4)

cofit2<-confint(mod2)

summary(mod2)

plot(mod2)

Anova(mod2)

#Based on the distribution of waste, cook distance and laverage, the model is valid

Plotmod2<-mod2$coefficients$mean

Plotmod2<-as.data.frame(Plotmod2)

conf2<-as.data.frame(cofit2)

conf2<-conf2[c(1:4),]

Plotmod2$lower<-conf2$`2.5 %`

Plotmod2$upper<-conf2$`97.5 %`

Plotmod2$label<-row.names(conf2)

Plotmod2<-Plotmod2[-c(1),]

plotMod2 <- ggplot(data=Plotmod2, aes(x=label, y=Plotmod2, ymin=lower, ymax=upper)) +

geom\_pointrange() +

geom\_hline(yintercept=0, lty=2) + # add a dotted line at x=1 after flip

coord\_flip() + # flip coordinates (puts labels on y axis)

xlab("") + ylab("Mean (95% CI)") +

theme(axis.text=element\_text(size=14),axis.title=element\_text(size=14,face="bold"),

panel.background = element\_rect(fill = "transparent"), # bg of the panel

plot.background = element\_rect(fill = "transparent", color = NA), # bg of the plot

panel.border= element\_rect(linetype = "solid", fill = NA, size = 1),

panel.grid.major = element\_blank(), # get rid of major grid

panel.grid.minor = element\_blank(), # get rid of minor grid

legend.background = element\_rect(fill = "transparent"), # get rid of legend bg

legend.box.background = element\_rect(fill = "transparent")) # get rid of legend panel bg

#,axis.line = element\_line(size = 1, linetype = "solid"))

#theme\_bw() # use a white background

print(plotMod2)

#We found the same results when considering only species with more than two hosts, reinforcing that functional

#specificity and number of host species are important for the prevalence at the community level

####Mean intensity of infection####

####Mean intensity of community infection####

dadospara3$limi<-log10(dadospara3$imi)

hist(dadospara3$limi, breaks = 20)

shapiro.test(dadospara3$limi)

modImi<-lm(limi~NHost, data=dadospara3)

#é uma skewed gaussian, mas vamos de modelo linear normal mesmo...

plot(limi~NHost, data=dadospara3)

abline(modImi)

cofitImi<-confint(modImi)

plot(modImi)

Anova(modImi)

#Based on the distribution of waste, cook distance and laverage, the model is still accepted

plotModII<-ggplot(dadospara3, aes(NHost, limi)) + geom\_point() +

geom\_point()+

#geom\_boxplot(fatten = 2,lwd=1.25) +

#geom\_jitter(aes(color = Parasite),size=3.5,

# position = position\_jitter(width = 0.25)) +

geom\_smooth(method = "lm")+

xlab("Number of host species") + ylab("Mean Intensity of Infection") +

theme(axis.text=element\_text(size=14),axis.title=element\_text(size=14,face="bold"),

panel.background = element\_rect(fill = "transparent"), # bg of the panel

plot.background = element\_rect(fill = "transparent", color = NA), # bg of the plot

panel.grid.major = element\_blank(), # get rid of major grid

panel.grid.minor = element\_blank(), # get rid of minor grid

legend.background = element\_rect(fill = "transparent"), # get rid of legend bg

legend.box.background = element\_rect(fill = "transparent") # get rid of legend panel bg

,axis.line = element\_line(size = 1.5, linetype = "solid"))

print(plotModII)

#Based on these data and analysis, the mean infection intensity of the community is

#not affected by parasite specificity

#Only positive

dadospara4$limi<-log10(dadospara4$imi)

hist(dadospara4$limi)

#poisson

modImIp<-glm(limi~NHost+FuncNRI+PhyloNRI, data=dadospara4, family = "quasipoisson")

cofitImIp<-confint(modImIp)

summary(modImIp)

plot(modImIp)

Anova(modImIp)

#Based on the distribution of waste, cook distance and laverage, the model is valid

PlotmodImIp<-modImIp$coefficients

PlotmodImIp<-as.data.frame(PlotmodImIp)

confImIp<-as.data.frame(cofitImIp)

PlotmodImIp$lower<-confImIp$`2.5 %`

PlotmodImIp$upper<-confImIp$`97.5 %`

PlotmodImIp$label<-row.names(confImIp)

PlotmodImIp<-PlotmodImIp[-c(1),]

plotModImIp<- ggplot(data=PlotmodImIp, aes(x=label, y=PlotmodImIp, ymin=lower, ymax=upper)) +

geom\_pointrange() +

geom\_hline(yintercept=0, lty=2) + # add a dotted line at x=1 after flip

coord\_flip() + # flip coordinates (puts labels on y axis)

xlab("") + ylab("Mean (95% CI)") +

theme(axis.text=element\_text(size=14),axis.title=element\_text(size=14,face="bold"),

panel.background = element\_rect(fill = "transparent"), # bg of the panel

plot.background = element\_rect(fill = "transparent", color = NA), # bg of the plot

panel.border= element\_rect(linetype = "solid", fill = NA, size = 1),

panel.grid.major = element\_blank(), # get rid of major grid

panel.grid.minor = element\_blank(), # get rid of minor grid

legend.background = element\_rect(fill = "transparent"), # get rid of legend bg

legend.box.background = element\_rect(fill = "transparent")) # get rid of legend panel bg

#,axis.line = element\_line(size = 1, linetype = "solid"))

#theme\_bw() # use a white background

print(plotModImIp)

####Now population-level models####

####Make the mixed model as proposed by Ellis et al 2020####

dadosSPPrev<-read.csv("dados\_spp\_prevEllis.csv",header = T, sep=";")

head(dadosSPPrev)

names(dadosSPPrev)[names(dadosSPPrev) == "ï..Parasite"] <- "Parasite"

dadosPrev1<-subset(dadosSPPrev, Prevalence > 0)

hist(dadosPrev1$Prevalence)

modPrevM<-glmer(cbind(Infected,Uninf)~NHost+(1|Host)+(1|Parasite),data=dadosPrev1,

family = binomial)

Anova(modPrevM)

summary(modPrevM)

plot(modPrevM)

confitPrevM<-confint(modPrevM)

#plot

#Host species richness is inversely proportional to the prevalence of infected host populations

plotPrevPop <- ggplot(dadosPrev1, aes(x = NHost, y = Prevalence)) +

geom\_point()+

geom\_smooth(method = "glm", formula=y ~ logit(x))+

xlab("Number of host species") + ylab("Prevalence") +

theme(axis.text=element\_text(size=14),axis.title=element\_text(size=14,face="bold"),

panel.background = element\_rect(fill = "transparent"), # bg of the panel

plot.background = element\_rect(fill = "transparent", color = NA), # bg of the plot

panel.grid.major = element\_blank(), # get rid of major grid

panel.grid.minor = element\_blank(), # get rid of minor grid

legend.background = element\_rect(fill = "transparent"), # get rid of legend bg

legend.box.background = element\_rect(fill = "transparent") # get rid of legend panel bg

,axis.line = element\_line(size = 1.5, linetype = "solid"))

print(plotPrevPop)

#Only with parasites that occur in 2 or more species

dadosPrev2<-subset(dadosPrev1,NHost>1)

corPrev2<-cor(dadosPrev2[,-c(1:2)])

corrplot(corPrev2,method="number",type="lower")

dadosPrev2$FuncNRI<-dadosPrev2$FuncSESMpdOc\*-1

dadosPrev2$PhyloNRI<-dadosPrev2$PhyloSESMpdOc\*-1

#Only with parasites with more than two hosts

modPrevM2<-glmer(cbind(Infected,Uninf)~NHost+FuncNRI+PhyloNRI+(1|Parasite)+(1|Host),data=dadosPrev2,

family = "binomial")

summary(modPrevM2)

plot(modPrevM2)

Anova(modPrevM2)

confitPrevM2<-confint(modPrevM2)

#plot

PlotmodPrevM2<-modPrevM2@beta

PlotmodPrevM2<-as.data.frame(PlotmodPrevM2)

confPrevM2<-as.data.frame(confitPrevM2)

confPrevM2<-confPrevM2[-c(1:2),]

PlotmodPrevM2$lower<-confPrevM2$`2.5 %`

PlotmodPrevM2$upper<-confPrevM2$`97.5 %`

PlotmodPrevM2$label<-row.names(confPrevM2)

PlotmodPrevM2<-PlotmodPrevM2[-c(1),]

plotModPrevM2 <- ggplot(data=PlotmodPrevM2, aes(x=label, y=PlotmodPrevM2, ymin=lower, ymax=upper)) +

geom\_pointrange() +

geom\_hline(yintercept=0, lty=2) + # add a dotted line at x=1 after flip

coord\_flip() + # flip coordinates (puts labels on y axis)

xlab("") + ylab("Mean (95% CI)") +

theme(axis.text=element\_text(size=14),axis.title=element\_text(size=14,face="bold"),

panel.background = element\_rect(fill = "transparent"), # bg of the panel

plot.background = element\_rect(fill = "transparent", color = NA), # bg of the plot

panel.border= element\_rect(linetype = "solid", fill = NA, size = 1),

panel.grid.major = element\_blank(), # get rid of major grid

panel.grid.minor = element\_blank(), # get rid of minor grid

legend.background = element\_rect(fill = "transparent"), # get rid of legend bg

legend.box.background = element\_rect(fill = "transparent")) # get rid of legend panel bg

#,axis.line = element\_line(size = 1, linetype = "solid"))

#theme\_bw() # use a white background

print(plotModPrevM2)

####Mean Infection Intensity per population####

hist(dadosPrev1$Limi)

modPrevIMI<-lmer(Limi~NHost+(1|Parasite)+(1|Host),data=dadosPrev1)

summary(modPrevIMI)

Anova(modPrevIMI)

#p = 0.0348

plot(modPrevIMI)

confitPrevIMI<-confint(modPrevIMI)

#plot

plotPrevPop <- ggplot(dadosPrev1, aes(x = NHost, y = Limi)) +

geom\_point()+

geom\_smooth(method = "lm")+

xlab("Number of host species") + ylab("Mean intensity of infection (Log)") +

theme(axis.text=element\_text(size=14),axis.title=element\_text(size=14,face="bold"),

panel.background = element\_rect(fill = "transparent"), # bg of the panel

plot.background = element\_rect(fill = "transparent", color = NA), # bg of the plot

panel.grid.major = element\_blank(), # get rid of major grid

panel.grid.minor = element\_blank(), # get rid of minor grid

legend.background = element\_rect(fill = "transparent"), # get rid of legend bg

legend.box.background = element\_rect(fill = "transparent") # get rid of legend panel bg

,axis.line = element\_line(size = 1.5, linetype = "solid"))

print(plotPrevPop)

#Mean infection intensity at the population level for parasites with two or more host species

#Mean Infection Intensity per population

hist(dadosPrev2$Limi)

modPrevIMI2<-lmer(Limi~NHost+FuncNRI+PhyloNRI+(1|Parasite)+(1|Host),data=dadosPrev2)

summary(modPrevIMI2)

plot(modPrevIMI2)

confitPrevIMI2<-confint(modPrevIMI2)

Anova(modPrevIMI2)

#plot

PlotmodPrevIMI2<-modPrevIMI2@beta

PlotmodPrevIMI2<-as.data.frame(PlotmodPrevIMI2)

confPrevIMI2<-as.data.frame(confitPrevIMI2)

confPrevIMI2<-confPrevIMI2[-c(1:3),]

PlotmodPrevIMI2$lower<-confPrevIMI2$`2.5 %`

PlotmodPrevIMI2$upper<-confPrevIMI2$`97.5 %`

PlotmodPrevIMI2$label<-row.names(confPrevIMI2)

PlotmodPrevIMI2<-PlotmodPrevIMI2[-c(1),]

PlotmodPrevIMI2<-as.data.frame(PlotmodPrevIMI2)

plotModPrevIMI2<- ggplot(data=PlotmodPrevIMI2, aes(x=label, y=PlotmodPrevIMI2, ymin=lower, ymax=upper)) +

geom\_pointrange() +

geom\_hline(yintercept=0, lty=2) + # add a dotted line at x=1 after flip

coord\_flip() + # flip coordinates (puts labels on y axis)

xlab("") + ylab("Mean (95% CI)") +

theme(axis.text=element\_text(size=14),axis.title=element\_text(size=14,face="bold"),

panel.background = element\_rect(fill = "transparent"), # bg of the panel

plot.background = element\_rect(fill = "transparent", color = NA), # bg of the plot

panel.border= element\_rect(linetype = "solid", fill = NA, size = 1),

panel.grid.major = element\_blank(), # get rid of major grid

panel.grid.minor = element\_blank(), # get rid of minor grid

legend.background = element\_rect(fill = "transparent"), # get rid of legend bg

legend.box.background = element\_rect(fill = "transparent")) # get rid of legend panel bg

#,axis.line = element\_line(size = 1, linetype = "solid"))

#theme\_bw() # use a white background

print(plotModPrevIMI2)