**R Scripts: Migrant birds disperse haemosporidian parasites and affect their transmission in avian communities**

Daniela de Angeli Dutra, Antoine Filion, Alan Fecchio, Érika Martins Braga, Robert Poulin

**Spacial Correlation and Phylogenetic Signal**

library(gdata)

library(gstat)

library(sp)

library(dplyr)

Dados <- read\_excel(“dados.xlsx”)

Dados <- filter(PD7, Parasiterichness !="NA")

class(Dados)

Dados$Parasiterichness <- as.numeric(Dados$Parasiterichness)

data <- as.matrix(dist(cbind(data$Longitude, data$Latitude)))

data1 <- 1/data

diag(data1) <- 0

data1 <-ifelse(data1=="Inf",0,data1 )

data1[1:5, 1:5]

library(ape)

Moran.I(Dados$Parasiterichness, data1, na.rm = TRUE)

### repeat procedure for parasite prevalence###

library(devtools)

library(treeman)

library(ape)

library(picante)

library(adephylo)

library(ade4)

library(phylobase)

library(geiger)

library(dplyr)

allTrees <- readTree("AllBirdsHackett1.tre")

random\_trees <-sample(allTrees@treelst, size = 100)

random\_trees1 <- as(random\_trees, 'TreeMen')

random\_trees2 <- as(random\_trees1, 'multiPhylo')

str(random\_trees2)

tree <- consensus(random\_trees2)

str(tree)

tree <- as.phylo(tree)

library(phytools)

tree2 <- phylog.extract(tree, node, distance = TRUE)

library(readxl)

data1 <- read\_excel(“dados.xlsx")

x <- as.vector(data1$Species)

x1 <- as.vector((data1$Infection))

data2 = data1 %>%

dplyr::group\_by(`Species`) %>%

dplyr::summarise(n\_sites = n\_distinct(Locality),

n\_bird\_individuals = n(),

n\_infections = sum(Infection)) %>%

dplyr::arrange(-n\_bird\_individuals)

fullbirds <- as.data.frame(tree$tip.label)

mybirds <- as.data.frame(data2$Species)

class(fullbirds)

class(mybirds)

fullbirds <- as.data.frame(fullbirds)

mybirds <- as.data.frame(mybirds)

length(which(tree$tip.label%in%as.character(mybirds[,1])))

todrop<-tree$tip.label[which(tree$tip.label%in%as.character(mybirds[,1])==FALSE)]

mybirds\_tree<-drop.tip(tree,todrop)

keep.tip(tree,as.character(mybirds[,1]))

drop2 <- as.data.frame(mybirds\_tree$tip.label)

drop2 <- left\_join()

names(drop2) <- c('SpeciesTotal')

data2 <- as.data.frame(data2)

data2$n\_infections <- as.numeric(data2$n\_infections)

data2$n\_bird\_individuals <- as.numeric(data2$n\_bird\_individuals)

data2 <- data2 %>%

mutate(prevalence = n\_infections/n\_bird\_individuals)

drop3 <- drop2$SpeciesTotal

drop3 <- as.data.frame(drop3)

data3 <- select\_if(data2$Species, vars(drop2$SpeciesTotal), .predicate = TRUE, nm=NULL)

data4 <- as.tbl(data2)

data3 <- filter\_if(data4, data4$Species == c("Todirostrum\_margaritaceiventer", "Elaenia\_sp.", "Setopagis\_parvulus", "Lepidocolaptes\_wagleri", "Rupornis\_magnirostris", "Sporophila\_sp.", "Fringilla\_brissonii", "Leptotila\_sp.", "Myiothlypis\_flaveolus", "Picus\_passerinus", "Pseudopipra\_pipra", "Basileuterus\_leucomelas", "Columbina\_sp.", "Empidonomus", "Euphonia\_sp.", "Hylophilus\_sp.", "Lanius\_lictor", "Lepidocolaptes\_angustirrostris", "Myiobius\_sp.", "Picumnus\_sp.", "Rallus\_viridis", "Stenopsis\_maculicaudus", "Synallaxis\_cinereus", "Synallaxis\_sp.", "Thamnophilus\_capstratus", "tolmomyias\_flaviventris", "Tolmomyias\_flaviventris", "Venilliornis\_mixtus", "Xiphorhynchus\_picus", "Xyphocolaptes\_falcirostris"), .preserve = FALSE)

myvars <- names(data2$Species) %in% c("Todirostrum\_margaritaceiventer", "Elaenia\_sp.", "Setopagis\_parvulus", "Lepidocolaptes\_wagleri", "Rupornis\_magnirostris", "Sporophila\_sp.", "Fringilla\_brissonii", "Leptotila\_sp.", "Myiothlypis\_flaveolus", "Picus\_passerinus", "Pseudopipra\_pipra", "Basileuterus\_leucomelas", "Columbina\_sp.", "Empidonomus", "Euphonia\_sp.", "Hylophilus\_sp.", "Lanius\_lictor", "Lepidocolaptes\_angustirrostris", "Myiobius\_sp.", "Picumnus\_sp.", "Rallus\_viridis", "Stenopsis\_maculicaudus", "Synallaxis\_cinereus", "Synallaxis\_sp.", "Thamnophilus\_capstratus", "tolmomyias\_flaviventris", "Tolmomyias\_flaviventris", "Venilliornis\_mixtus", "Xiphorhynchus\_picus", "Xyphocolaptes\_falcirostris")

newdata <- data2[!mybirds2]

todrop1<-data2$Species[which(data2$Species%in%as.character(drop2[,1])==FALSE)]

print(todrop1)

data3 <- filter\_all(data2$Species == todrop1)

tree2 <- select(tree$tip.label, mybirds)

mybirds\_tree$tip.label

data3 <- left\_join(drop2, data2, by = c("SpeciesTotal" = "Species") )

phylosig(mybirds\_tree, data3$prevalence , method="lambda", test=FALSE, nsim=1000, se=NULL, start=NULL,

control=list())

### repeat the procedure for parasite richness###

**Geographical Range Calculation:**

library(brms)

library(ggplot2)

library(readxl)

library(GeoRange)

library(fossil)

library(data.table)

library(tidyverse)

library(readODS)

data1 <- read\_excel("Lineages1.xlsx")

data2 <- read\_excel("Lineages2.xlsx", sheet = "Sheet2")

class(data2)

data2 <- as.data.frame(data2)

data2 <- data2[ , c("site", "Latitude", "Longitude")]

data2 <- distinct(data2)

dados <- left\_join(data1, data2, by = c("site" = "site"))

dados <- as.data.frame(dados)

data <- create.matrix(dados, tax.name ="Lineage\_Name", locality ="site", abund = FALSE)

data3 <- t(data)

data3 <- as.data.frame(data3)

setDT(data3, keep.rownames = "site")

data4 <- left\_join(data3, data2, by = c("site" = "site"))

setcolorder(data4, c("Longitude", "Latitude"))

data4$site <- NULL

data4[data1 == 0] <- NA

data4$Longitude=as.numeric(paste(data4$Longitude))

data4$Latitude=as.numeric(paste(data4$Latitude))

data4 <- data4 %>% drop\_na(Longitude)

data4<- as.data.frame(data4)

class(data4)

Range <- GeoRange\_MultiTaxa(OccMatrix=data4,TaxaStart=3)

Range1 <- filter(Range, NLocs > 1)

write.csv(Range1, "Range.csv")

**Bayesian Model 1**

dados1 <- read\_ods("LineagesGeoRange.ods")

dados1$MST=as.numeric(paste(dados1$MST))

dados1 <- filter(dados1, NLoc > 1)

variaveis1 <- bf(MST~Host\_Status + n\_bird\_individuals + richness, family = Gamma(link = log)) #check

prior1 <- get\_prior(variaveis1, data = dados1)

prior1

dados1$Host\_Status=as.factor(paste(dados1$Host\_Status))

dados1$Host\_Status<- relevel(dados1$Host\_Status, ref="R")

levels(dados1$Host\_Status)

model2 <- brm(

MST~Host\_Status + n\_bird\_individuals + riqueza, data = dados1,

family = Gamma(link = "log"), chains = 4,

iter = 4000,

prior = c(

prior(student\_t(3, 7.3, 2.5), "Intercept"),

prior(gamma(0.01, 0.01),"shape")

))

summary(model2)

plot(model2, N = 4, ask =FALSE)

a <- theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),

panel.background = element\_blank(), axis.line = element\_line(colour = "black"))

plot <- plot(conditional\_effects(model2), points = FALSE, theme = a)

plot2 <- plot$Host\_Status + labs(x = "Host Status", y = "Geographical Range (km)") #caption = "Plasmodium")

+ geom\_boxplot(aes(color = Host\_Status))

plot2

**Bayesian Model 2:**

library(tidyverse)

library(ggplot2)

library(brms)

library(readxl)

library(ggpubr)

dados1 <- read\_excel("dados.xlsx", sheet = "Objetivo2b>=10",

col\_types = c("text", "text", "numeric",

"numeric", "numeric", "numeric",

"numeric", "numeric", "numeric",

"numeric", "text", "numeric", "numeric",

"numeric", "numeric", "numeric",

"numeric", "numeric", "numeric", "numeric",

"numeric"))

dados1 <- filter(dados1, Parasiterichness != "NA")

hist(dados2$Pos, breaks = 100)

hist(dados1$Parasiterichness)

hist(dados1$RiquezadeHospedeiros)

hist(dados1$migrantindividuals)

hist(dados1$Temp)

hist(dados1$Prec)

hist(dados1$n\_migrants)

dados1$Species = gsub(' ', '\_', dados1$Species)

dados1$Species=as.factor(paste(dados1$Species))

haemo\_species <- dados1$Species

myphy <- readRDS("C:/Users/danid/Documents/Lab Poulin/PhD/PhD/Dados/PhD/myphy.rds")

library(ape)

matches2 <-match(haemo\_species, myphy$tip.label)

matches2 <-na.omit(matches2)

haemo\_tree <-drop.tip(myphy, myphy$tip.label[-matches2])

sptest=as.factor(haemo\_tree$tip.label)

sp=as.data.frame(sptest)

dados2=dados1 %>%

filter(Species %in% sp$sptest)

inv.phylo <- MCMCglmm::inverseA(haemo\_tree, nodes = "TIPS", scale = TRUE)

A <- solve(inv.phylo$Ainv)

rownames(A) <- rownames(inv.phylo$Ainv)

dados2$phylo <- dados2$Species

teste <- brm(Pos~Totalsample + Parasiterichness + (1|Bioma) + (1|Locality) + (1|phylo),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000, cov\_ranef = list(phylo = A))

summary(teste)

teste2 <- brm(Pos~ Totalsample + Prec + (1|Bioma) + (1|Locality) + (1|phylo),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000, cov\_ranef = list(phylo = A))

summary(teste2)

teste3 <- brm(Pos~ Totalsample + Temp + (1|Bioma) + (1|Locality) + (1|phylo),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000, cov\_ranef = list(phylo = A))

summary(teste3)

teste4 <- brm(Pos~ Totalsample + n\_migrants + (1|Bioma) + (1|Locality) + (1|phylo),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000, cov\_ranef = list(phylo = A))

summary(teste4)

variaveis <- bf(Pos~ Totalsample +migrantindividuals + Parasiterichness +

(1|Bioma) + (1|Locality) + (1|phylo), family = negbinomial())

prior <- get\_prior(variaveis, data = dados2)

prior

model <- brm(Pos~ Totalsample + migrantindividuals + migrantspecies + (1|Bioma) + (1|Locality) + (1|phylo),

data = dados2,

family = negbinomial(),

chain = 4, iter = 4000,

cov\_ranef = list(phylo = A),

prior = c(

prior(student\_t(3, 1.1, 2.5), "Intercept"),

prior(student\_t(3, 0, 2.5), "sd" ),

prior(gamma(0.01, 0.01), "shape")

)

)

summary(model)

plot(model, N = 4, ask = FALSE)

plot1 <- plot(conditional\_effects(model), points = FALSE, theme = a)

a <- theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),

panel.background = element\_blank(), axis.line = element\_line(colour = "black"))

plot2 <- plot1$migrantindividuals + labs(x = "Proportion of migrant individuals", y = "Number of Infections") + coord\_cartesian(ylim = c(0, 15)) #caption = "Plasmodium")

+ geom\_boxplot(aes(color = migrantindividuals))

plot2

res1 <- residuals(model)

res1 <- as.data.frame(res1)

dados2$residuals <- res1$Estimate

dados2$residuals <- as.numeric(dados2$residuals)

dados2a <- as.matrix(dist(cbind(dados2$Longitude, dados2$Latitude)))

dados2b <- 1/dados2a

diag(dados2b) <- 0

dados2b <-ifelse(dados2b=="Inf",0,dados2b )

dados2b[1:5, 1:5]

Moran.I(dados2$residuals, dados2b, na.rm = TRUE)

**Bayesian Model 3**

library(tidyverse)

library(ggplot2)

library(brms)

library(readxl)

library(ggpubr)

dados1 <- read\_excel("PD7.xlsx", sheet = "Objetivo 2.txt",

col\_types = c("text", "numeric", "numeric",

"numeric", "numeric", "numeric",

"numeric", "numeric", "numeric",

"numeric", "numeric", "text", "numeric",

"numeric", "numeric", "numeric",

"numeric", "numeric", "numeric", "numeric",

"numeric",

"numeric"))

dados1 <- filter(dados1, Parasiterichness != "NA")

hist(dados1$Parasiterichness)

hist(dados1$Prevalence)

hist(dados1$Hostrichness)

hist(dados1$migrantspecies)

hist(dados1$Temp)

hist(dados1$Prec)

hist(dados1$n\_migrants)

teste <- brm(Parasiterichness~ Totalsample + Prevalencia + (1|Bioma) + (1|Locality),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000)

summary(teste)

teste2 <- brm(Parasiterichness~ Totalsample + log1p(Temp) + (1|Bioma) + (1|Locality),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000)

summary(teste2)

teste3 <- brm(Parasiterichness~ Totalsample + Prec + (1|Bioma) + (1|Locality),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000)

summary(teste3)

teste4 <- brm(Parasiterichness~ Totalsample + Hostrichness + (1|Bioma) + (1|Locality),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000)

summary(teste4)

variaveis <- bf(Parasiterichness~ Totalsample + migrantspecies + migrantindividuals + Hostrichness + (1|Bioma) + (1|Locality), family = negbinomial())

prior <- get\_prior(variaveis, data = dados1)

prior

model <- brm(Parasiterichness~ Totalsample + migrantindividuals + migrantspecies + Hostrichness + (1|Bioma) + (1|Locality),

data = dados1,

family = negbinomial(),

chain = 4, iter = 4000,

prior = c(

prior(student\_t(3, 1.1, 2.5), "Intercept"),

prior(student\_t(3, 0, 2.5), "sd" ),

prior(gamma(0.01, 0.01), "shape")

)

)

summary(model)

plot(model, N = 4, ask = FALSE)

plot1 <- plot(conditional\_effects(model), points = FALSE, theme = a)

a <- theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),

panel.background = element\_blank(), axis.line = element\_line(colour = "black"))

plot2 <- plot1$migrantindividuals + labs(x = "Proportion of migrant individuals", y = "Number of Infections") + coord\_cartesian(ylim = c(0, 15)) #caption = "Plasmodium")

+ geom\_boxplot(aes(color = migrantindividuals))

plot2

res1 <- residuals(model)

res1 <- as.data.frame(res1)

dados1$residuals <- res1$Estimate

dados1$residuals <- as.numeric(dados1$residuals)

dados1a <- as.matrix(dist(cbind(dados1$Longitude, dados1$Latitude)))

dados1b <- 1/dados1a

diag(dados1b) <- 0

dados1b <-ifelse(dados1b=="Inf",0,dados1b )

dados1b[1:5, 1:5]

Moran.I(dados1$residuals, dados1b, na.rm = TRUE)