##### Create Primate Phylogenetic Trees #####

install.packages("rotl")

install.packages("ape")

install.packages("phangorn")

library(phangorn)

library(rotl)

library(ape)

##Full Species List Tree

primates <- tnrs\_match\_names(names = "Primates")

print(primates)

primate\_tree <- tol\_subtree(ott\_id = primates$ott\_id)

plot.phylo(primate\_tree)

write.nexus(primate\_tree, file = "primate\_tree.nexus")

prunedspecies\_list <- c("Allenopithecus\_nigroviridis", "Allocebus\_trichotis", "Allochrocebus\_lhoesti", "Allochrocebus\_preussi", "Allochrocebus\_solatus", "Alouatta\_arctoidea", "Alouatta\_belzebul", "Alouatta\_caraya", "Alouatta\_discolor", "Alouatta\_juara", "Alouatta\_nigerrima", "Alouatta\_palliata", "Alouatta\_pigra", "Alouatta\_sara", "Alouatta\_seniculus", "Alouatta\_ululata", "Aotus\_azarae", "Aotus\_brumbacki", "Aotus\_griseimembra", "Aotus\_jorgehernandezi", "Aotus\_lemurinus", "Aotus\_miconax", "Aotus\_nancymaae", "Aotus\_nigriceps", "Aotus\_trivirgatus", "Aotus\_vociferans", "Aotus\_zonalis", "Arctocebus\_aureus", "Arctocebus\_calabarensis", "Ateles\_belzebuth", "Ateles\_fusciceps", "Ateles\_geoffroyi", "Ateles\_marginatus", "Ateles\_paniscus", "Avahi\_betsileo", "Avahi\_cleesei", "Avahi\_laniger", "Avahi\_meridionalis", "Avahi\_mooreorum", "Avahi\_occidentalis", "Avahi\_peyrierasi", "Avahi\_ramanantsoavani", "Avahi\_unicolor", "Brachyteles\_arachnoides", "Brachyteles\_hypoxanthus", "Cacajao\_ayresi", "Cacajao\_calvus", "Cacajao\_hosomi", "Cacajao\_melanocephalus", "Callicebus\_barbarabrownae", "Callicebus\_caquetensis", "Callicebus\_coimbrai", "Callicebus\_discolor", "Callicebus\_dubius", "Callicebus\_melanochir", "Callicebus\_nigrifrons", "Callicebus\_ornatus", "Callicebus\_personatus", "Callimico\_goeldii", "Callithrix\_aurita", "Callithrix\_flaviceps", "Callithrix\_geoffroyi", "Callithrix\_humeralifer", "Callithrix\_humilis", "Callithrix\_jacchus", "Callithrix\_kuhlii", "Callithrix\_penicillata", "Callithrix\_pygmaea", "Carlito\_syrichta", "Cebuella\_niveiventris", "Cebus\_aequatorialis", "Cebus\_albifrons", "Cebus\_brunneus", "Cebus\_capucinus", "Cebus\_cesarae", "Cebus\_cuscinus", "Cebus\_imitator", "Cebus\_kaapori", "Cebus\_leucocephalus", "Cebus\_malitiosus", "Cebus\_olivaceus", "Cebus\_unicolor", "Cebus\_versicolor", "Cebus\_yuracus", "Cephalopachus\_bancanus", "Cercocebus\_agilis", "Cercocebus\_atys", "Cercocebus\_chrysogaster", "Cercocebus\_galeritus", "Cercocebus\_sanjei", "Cercocebus\_torquatus", "Cercopithecus\_albogularis", "Cercopithecus\_ascanius", "Cercopithecus\_buettikoferi", "Cercopithecus\_lowei", "Cercopithecus\_cephus", "Cercopithecus\_denti", "Cercopithecus\_diana", "Cercopithecus\_doggetti", "Cercopithecus\_dryas", "Cercopithecus\_erythrogaster", "Cercopithecus\_hamlyni", "Cercopithecus\_kandti", "Cercopithecus\_lomamiensis", "Cercopithecus\_mitis", "Cercopithecus\_mona",

"Cercopithecus\_neglectus", "Cercopithecus\_nicticans", "Cercopithecus\_petaurista", "Cercopithecus\_pogonias", "Cercopithecus\_roloway", "Cercopithecus\_sclateri", "Cercopithecus\_wolfi", "Cheirogaleus\_andysabini", "Cheirogaleus\_crossleyi", "Cheirogaleus\_lavasoensis", "Cheirogaleus\_major", "Cheirogaleus\_medius", "Cheirogaleus\_minusculus", "Cheirogaleus\_shethi", "Cheirogaleus\_sibreei", "Cheirogaleus\_thomasi", "Cheracebus\_lucifer", "Cheracebus\_lugens", "Cheracebus\_medemi", "Cheracebus\_purinus", "Cheracebus\_regulus", "Cheracebus\_torquatus", "Chiropotes\_albinasus", "Chiropotes\_sagulata", "Chiropotes\_israelita", "Chiropotes\_satanas", "Chiropotes\_utahickae", "Chlorocebus\_aethiops", "Chlorocebus\_cynosuros", "Chlorocebus\_djamdjamensis", "Chlorocebus\_pygerythrus", "Chlorocebus\_sabaeus", "Colobus\_angolensis", "Colobus\_guereza", "Colobus\_polykomos", "Colobus\_satanas", "Colobus\_vellerosus", "Daubentonia\_madagascariensis", "Erythrocebus\_patas", "Erythrocebus\_poliophaeus", "Eulemur\_albifrons", "Eulemur\_cinereiceps", "Eulemur\_coronatus", "Eulemur\_flavifrons", "Eulemur\_fulvus", "Eulemur\_macaco", "Eulemur\_mongoz", "Eulemur\_rubriventer", "Eulemur\_rufifrons", "Eulemur\_rufus", "Eulemur\_sanfordi", "Euoticus\_elegantulus", "Euoticus\_pallidus", "Galago\_alleni", "Galago\_gabonensis", "Galago\_gallarum", "Galago\_granti", "Galago\_matschiei", "Galago\_minor", "Galago\_moholi", "Galago\_senegalensis",

"Galagoides\_kumbirensis", "Galagoides\_orinus", "Galagoides\_rondoensis", "Galagoides\_thomasi", "Galagoides\_zanzibaricus", "Gorilla\_beringei", "Gorilla\_gorilla", "Hapalemur\_alaotrensis", "Hapalemur\_aureus", "Hapalemur\_griseus", "Hapalemur\_occidentalis", "Hoolock\_hoolock", "Hoolock\_leuconedys", "Hoolock\_tianxing", "Hylobates\_agilis", "Hylobates\_albibarbis", "Hylobates\_klossii", "Hylobates\_lar", "Hylobates\_moloch", "Hylobates\_muelleri", "Hylobates\_pileatus", "Indri\_indri", "Lagothrix\_cana", "Lagothrix\_flavicauda", "Lagothrix\_lagotricha", "Lagothrix\_lugens", "Lagothrix\_poeppigii", "Lagothrix\_tschudii", "Lemur\_catta", "Leontopithecus\_caissara", "Leontopithecus\_chrysomelas", "Leontopithecus\_chrysopygus", "Leontopithecus\_rosalia", "Lepilemur\_aeeclis", "Lepilemur\_ahmansonorum", "Lepilemur\_ankaranensis", "Lepilemur\_betsileo", "Lepilemur\_dorsalis", "Lepilemur\_edwardsi", "Lepilemur\_fleuretae",

"Lepilemur\_grewcockorum", "Lepilemur\_hollandorum", "Lepilemur\_hubbardorum", "Lepilemur\_jamesorum", "Lepilemur\_leucopus", "Lepilemur\_manasamody", "Lepilemur\_microdon", "Lepilemur\_milanoii", "Lepilemur\_mitsinjoensis", "Lepilemur\_mittermeieri", "Lepilemur\_mustelinus", "Lepilemur\_otto", "Lepilemur\_petteri", "Lepilemur\_randrianasoloi", "Lepilemur\_ruficaudatus", "Lepilemur\_sahamalazensis", "Lepilemur\_scottorum", "Lepilemur\_seali", "Lepilemur\_septentrionalis", "Lepilemur\_tymerlachsoni", "Lepilemur\_wrightae", "Lophocebus\_albigena", "Lophocebus\_aterrimus", "Loris\_lydekkerianus", "Loris\_tardigradus", "Macaca\_arctoides", "Macaca\_assamensis", "Macaca\_brunnescens", "Macaca\_cyclopis", "Macaca\_fascicularis", "Macaca\_fuscata", "Macaca\_hecki", "Macaca\_leonina", "Macaca\_leucogenys", "Macaca\_maura", "Macaca\_mulatta", "Macaca\_munzala", "Macaca\_nemestrina", "Macaca\_nigra", "Macaca\_nigrescens", "Macaca\_ochreata", "Macaca\_pagensis", "Macaca\_radiata", "Macaca\_siberu", "Macaca\_silenus", "Macaca\_sinica", "Macaca\_speciosa", "Macaca\_sylvanus", "Macaca\_thibetana", "Macaca\_tonkeana", "Mandrillus\_leucophaeus", "Mandrillus\_sphinx", "Mico\_acariensis", "Mico\_argentatus", "Mico\_chrysoleucus", "Mico\_emiliae", "Mico\_humeralifer", "Mico\_intermedius", "Mico\_leucippe", "Mico\_marcai", "Mico\_mauesi", "Mico\_melanurus", "Mico\_munduruku", "Mico\_nigriceps", "Mico\_rondoni", "Mico\_saterei", "Microcebus\_arnholdi", "Microcebus\_berthae", "Microcebus\_bongolavensis", "Microcebus\_boraha", "Microcebus\_danfossi", "Microcebus\_ganzhorni", "Microcebus\_gerpi", "Microcebus\_griseorufus", "Microcebus\_jollyae", "Microcebus\_lehilahytsara", "Microcebus\_macarthurii", "Microcebus\_mamiratra", "Microcebus\_manitatra", "Microcebus\_margotmarshae", "Microcebus\_marohita", "Microcebus\_mittermeieri", "Microcebus\_murinus", "Microcebus\_myoxinus", "Microcebus\_ravelobensis", "Microcebus\_rufus", "Microcebus\_sambiranensis", "Microcebus\_simmonsi", "Microcebus\_tanosi", "Microcebus\_tavaratra", "Miopithecus\_ogouensis", "Miopithecus\_talapoin", "Mirza\_coquereli", "Mirza\_zaza", "Nasalis\_larvatus", "Nomascus\_annamensis", "Nomascus\_concolor", "Nomascus\_gabriellae", "Nomascus\_hainanus", "Nomascus\_leucogenys", "Nomascus\_nasutus", "Nomascus\_siki", "Nycticebus\_bancanus", "Nycticebus\_bengalensis", "Nycticebus\_borneanus", "Nycticebus\_coucang", "Nycticebus\_javanicus",

"Nycticebus\_kayan", "Nycticebus\_pygmaeus", "Otolemur\_crassicaudatus", "Otolemur\_garnettii", "Otolemur\_monteiri", "Pan\_paniscus", "Pan\_troglodytes", "Papio\_anubis", "Papio\_cynocephalus", "Papio\_hamadryas", "Papio\_kindae", "Papio\_papio", "Papio\_ursinus", "Paragalago\_cocos", "Paragalago\_zanzibaricus", "Perodicticus\_potto", "Phaner\_electromontis", "Phaner\_furcifer", "Phaner\_pallescens", "Phaner\_parienti", "Piliocolobus\_badius", "Piliocolobus\_bouvieri", "Piliocolobus\_epieni", "Piliocolobus\_foai", "Piliocolobus\_gordonorum", "Piliocolobus\_kirkii", "Piliocolobus\_lulindicus", "Piliocolobus\_oustaleti", "Piliocolobus\_parmentieri", "Piliocolobus\_pennantii", "Piliocolobus\_preussi", "Piliocolobus\_rufomitratus", "Piliocolobus\_semlikiensis", "Piliocolobus\_tephrosceles", "Piliocolobus\_tholloni", "Piliocolobus\_waldronae", "Pithecia\_aequatorialis", "Pithecia\_albicans", "Pithecia\_cazuzai", "Pithecia\_chrysocephala", "Pithecia\_hirsuta", "Pithecia\_inusta", "Pithecia\_irrorata", "Pithecia\_isabela", "Pithecia\_milleri", "Pithecia\_mittermeieri", "Pithecia\_monachus", "Pithecia\_napensis", "Pithecia\_pissinattii", "Pithecia\_pithecia", "Pithecia\_rylandsi", "Pithecia\_vanzolinii", "Plecturocebus\_aureipalatii", "Plecturocebus\_baptista", "Plecturocebus\_bernhardi", "Plecturocebus\_brunneus", "Plecturocebus\_cinerascens", "Plecturocebus\_cupreus", "Plecturocebus\_donacophilus", "Plecturocebus\_grovesi", "Plecturocebus\_hoffmannsi", "Plecturocebus\_miltoni", "Plecturocebus\_modestus", "Plecturocebus\_moloch", "Plecturocebus\_oenanthe", "Plecturocebus\_olallae", "Plecturocebus\_pallescens", "Plecturocebus\_stephennashi", "Plecturocebus\_toppini", "Plecturocebus\_urubambensis", "Plecturocebus\_vieirai", "Pongo\_abelii", "Pongo\_pygmaeus", "Pongo\_tapanuliensis", "Presbytis\_canicrus", "Presbytis\_chrysomelas", "Presbytis\_comata", "Presbytis\_femoralis", "Presbytis\_frontata", "Presbytis\_hosei", "Presbytis\_melalophos", "Presbytis\_natunae", "Presbytis\_pileata", "Presbytis\_potenziani", "Presbytis\_robinsoni", "Presbytis\_rubicunda", "Presbytis\_sabana", "Presbytis\_siamensis", "Presbytis\_thomasi", "Procolobus\_verus", "Prolemur\_simus", "Propithecus\_candidus", "Propithecus\_coquereli", "Propithecus\_deckenii", "Propithecus\_diadema", "Propithecus\_edwardsi", "Propithecus\_perrieri", "Propithecus\_tattersalli", "Propithecus\_verreauxi",

"Pseudopotto\_martini", "Pygathrix\_cinerea", "Pygathrix\_nemaeus", "Pygathrix\_nigripes", "Rhinopithecus\_avunculus", "Rhinopithecus\_bieti", "Rhinopithecus\_brelichi", "Rhinopithecus\_roxellana", "Rhinopithecus\_strykeri", "Rungwecebus\_kipunji", "Saguinus\_bicolor", "Saguinus\_cruzlimai", "Saguinus\_fuscicollis", "Saguinus\_geoffroyi", "Saguinus\_graellsi", "Saguinus\_imperator", "Saguinus\_inustus", "Saguinus\_labiatus", "Saguinus\_leucopus", "Saguinus\_martinsi", "Saguinus\_melanoleucus", "Saguinus\_midas", "Saguinus\_mystax", "Saguinus\_niger", "Saguinus\_nigricollis", "Saguinus\_oedipus", "Saguinus\_tripartitus", "Saguinus\_ursulus", "Saimiri\_boliviensis", "Saimiri\_cassiquiarensis", "Saimiri\_oerstedii", "Saimiri\_sciureus", "Saimiri\_ustus", "Saimiri\_vanzolinii", "Sapajus\_apella", "Sapajus\_cay", "Sapajus\_flavius", "Sapajus\_libidinosus", "Sapajus\_nigritus", "Sapajus\_xanthosternos", "Sciurocheirus\_makandensis", "Semnopithecus\_ajax", "Semnopithecus\_entellus", "Semnopithecus\_hector", "Semnopithecus\_hypoleucos", "Semnopithecus\_natunae", "Semnopithecus\_priam", "Semnopithecus\_schistaceus", "Simias\_concolor", "Symphalangus\_syndactylus", "Tarsius\_dentatus", "Tarsius\_fuscus", "Tarsius\_lariang", "Tarsius\_pelengensis", "Tarsius\_pumilus", "Tarsius\_sangirensis", "Tarsius\_spectrumgurskyae", "Tarsius\_supriatnai", "Tarsius\_tarsier", "Tarsius\_tumpara", "Tarsius\_wallacei", "Theropithecus\_gelada", "Trachypithecus\_auratus", "Trachypithecus\_barbei", "Trachypithecus\_cristatus", "Trachypithecus\_delacouri", "Trachypithecus\_ebenus", "Trachypithecus\_francoisi", "Trachypithecus\_geei", "Trachypithecus\_germaini", "Trachypithecus\_hatinhensis", "Trachypithecus\_johnii", "Trachypithecus\_laotum", "Trachypithecus\_mauritius", "Trachypithecus\_obscurus", "Trachypithecus\_phayrei", "Trachypithecus\_poliocephalus", "Trachypithecus\_selangorensis", "Trachypithecus\_shortridgei", "Trachypithecus\_vetulus", "Varecia\_rubra", "Varecia\_variegata")

prunedspecies\_match <- tnrs\_match\_names(names = prunedspecies\_list)

prunedspecies\_tree <- tol\_induced\_subtree(ott\_ids = prunedspecies\_match$ott\_id)

prunedspecies\_tree <- compute.brlen(prunedspecies\_tree, method = "Grafen")

plot.phylo(prunedspecies\_tree)

write.nexus(prunedspecies\_tree, file = "pruned\_species\_tree.nexus")

species\_list <- c("Ateles\_fusciceps", "Ateles\_geoffroyi", "Callithrix\_jacchus", "Callithrix\_penicillata",

"Cebus\_albifrons", "Cebus\_capucinus", "Cebus\_imitator", "Cebus\_olivaceus",

"Cercocebus\_atys", "Cercopithecus\_ascanius", "Cercopithecus\_mona",

"Chlorocebus\_pygerythrus", "Colobus\_guereza", "Colobus\_vellerosus", "Erythrocebus\_patas",

"Eulemur\_fulvus", "Galago\_moholi", "Gorilla\_beringei", "Gorilla\_gorilla", "Hylobates\_lar",

"Lagothrix\_lagotricha", "Lagothrix\_poeppigii", "Leontopithecus\_rosalia", "Lophocebus\_albigena",

"Macaca\_arctoides", "Macaca\_fascicularis", "Macaca\_fuscata", "Macaca\_maura", "Macaca\_mulatta",

"Macaca\_nemestrina", "Macaca\_nigra", "Macaca\_radiata", "Macaca\_silenus", "Macaca\_sylvanus",

"Macaca\_thibetana", "Macaca\_tonkeana", "Miopithecus\_talapoin", "Nasalis\_larvatus", "Pan\_paniscus",

"Pan\_troglodytes", "Papio\_anubis", "Papio\_cynocephalus", "Papio\_hamadryas", "Pongo\_abelii",

"Pongo\_pygmaeus", "Propithecus\_verreauxi", "Rhinopithecus\_roxellana", "Saguinus\_fuscicollis",

"Saguinus\_geoffroyi", "Saguinus\_mystax", "Saguinus\_oedipus", "Saimiri\_sciureus", "Sapajus\_apella",

"Sapajus\_libidinosus", "Sapajus\_nigritus", "Sapajus\_xanthosternos", "Semnopithecus\_entellus",

"Symphalangus\_syndactylus", "Theropithecus\_gelada", "Trachypithecus\_johnii")

species\_match <- tnrs\_match\_names(names = species\_list)

species\_tree <- tol\_induced\_subtree(ott\_ids = species\_match$ott\_id)

if (is.null(species\_tree$edge.length)) {

species\_tree$edge.length <- rep(1, nrow(species\_tree$edge))

}

plot.phylo(species\_tree)

write.nexus(species\_tree, file = "SSBspecies\_tree.nexus")

##Filtered SSB Species Tree

species\_list <- c("Callithrix\_jacchus", "Colobus\_vellerosus", "Gorilla\_beringei", "Gorilla\_gorilla",

"Hylobates\_lar", "Macaca\_arctoides", "Macaca\_fascicularis", "Macaca\_fuscata",

"Macaca\_mulatta", "Macaca\_nemestrina", "Macaca\_nigra", "Macaca\_sylvanus",

"Macaca\_thibetana", "Macaca\_tonkeana", "Nasalis\_larvatus", "Pan\_paniscus",

"Pan\_troglodytes", "Papio\_anubis", "Papio\_cynocephalus", "Rhinopithecus\_roxellana",

"Saimiri\_sciureus", "Sapajus\_nigritus", "Semnopithecus\_entellus")

species\_match <- tnrs\_match\_names(names = species\_list)

species\_tree <- tol\_induced\_subtree(ott\_ids = species\_match$ott\_id)

if (is.null(species\_tree$edge.length)) {

species\_tree$edge.length <- rep(1, nrow(species\_tree$edge))

}

plot.phylo(species\_tree)

write.nexus(species\_tree, file = "filteredSSBspecies\_tree.nexus")

##Phylogenetic Tree Figure

install.packages("BiocManager")

BiocManager::install("ggtree")

install.packages(c("ggplot2", "remotes", "stringr"))

remotes::install\_github("YuLab-SMU/ggtreeExtra")

library(ape)

library(dplyr)

library(stringr)

library(ggplot2)

library(ggtree)

library(readr)

full\_tree <- read.nexus("pruned\_species\_tree.nexus")

data <- read\_csv("full\_dataframe.csv") %>%

rename(label = nexus\_name)

# Assign tree tip labels

tree <- full\_tree

tree$tip.label <- full\_tree$tip.label

data\_in\_tree <- data %>% filter(label %in% tree$tip.label)

# Create named list of tips per Category

tiplist\_by\_category <- split(data\_in\_tree$label, data\_in\_tree$Category)

# Assign groups to tree tips using groupOTU

grouped\_tree <- groupOTU(tree, tiplist\_by\_category)

# Plot the grouped tree

p <- ggtree(grouped\_tree, layout = "circular", aes(color = group)) +

geom\_tree(size = 0.6) +

scale\_color\_manual(values = c(

"Strepsirrhini" = "#1b9e77",

"Platyrrhini" = "#d95f02",

"Cercopithecidae" = "#7570b3",

"Hominoidae" = "#e7298a",

"Hylobatidae" = "#66a61e",

"Tarsiiformes" = "#e6ab02"

), na.value = "grey70") +

theme(

legend.title = element\_blank(),

legend.position = "right",

plot.margin = margin(2, 2, 2, 2, "cm")

)

p

##### Phylogenetic Signal of SSB #####

library(ape)

library(geiger)

tree <- read.nexus("pruned\_species\_tree.nexus")

ssb\_data <- read.csv("full\_dataframe.csv", stringsAsFactors = FALSE)

ssb\_data <- ssb\_data[, c("nexus\_name", "ssb\_presence")]

ssb\_data$nexus\_name <- as.character(ssb\_data$nexus\_name)

ssb\_data$ssb\_presence <- as.numeric(ssb\_data$ssb\_presence)

tree <- drop.tip(tree, setdiff(tree$tip.label, ssb\_data$nexus\_name))

ssb\_data <- ssb\_data[ssb\_data$nexus\_name %in% tree$tip.label, ]

tree <- multi2di(tree)

ssb\_vector <- ssb\_data$ssb\_presence

names(ssb\_vector) <- ssb\_data$nexus\_name

ssb\_vector <- ssb\_vector + 1

fit <- fitDiscrete(tree, ssb\_vector, model = "ER", transform = "lambda")

print(fit)

##### Prepare Data for Analysis #####

install.packages("caper")

install.packages("ape")

install.packages("phylolm")

install.packages("ggplot2")

install.packages("dplyr")

install.packages("Rphylopars")

install.packages("tibble")

library(caper)

library(ape)

library(phylolm)

library(ggplot2)

library(dplyr)

library(Rphylopars)

library(tibble)

data <- read.csv("full\_dataframe.csv")

ssbdata <- read.csv("filtered\_ssb\_dataframe.csv")

tree <- read.nexus("pruned\_species\_tree.nexus")

ssbtree <- read.nexus("filteredSSBspecies\_tree.nexus")

rownames(data) <- data$nexus\_name

pruned\_tree <- drop.tip(tree, setdiff(tree$tip.label, rownames(data)))

rownames(ssbdata) <- ssbdata$nexus\_name

pruned\_ssb\_tree <- drop.tip(ssbtree, setdiff(ssbtree$tip.label, rownames(ssbdata)))

##### ENVIRONMENT #####

#Imputation - full dataset

env\_data\_full <- data %>%

filter(nexus\_name %in% pruned\_tree$tip.label) %>%

select(nexus\_name,

mean\_temp\_2019,

mean\_precip\_2019,

mean\_cloud\_2019,

mean\_biomass\_2019,

predator\_count) %>%

rename(species = nexus\_name)

env\_imputed\_full <- phylopars(trait\_data = env\_data\_full, tree = pruned\_tree)

env\_traits\_full <- env\_imputed\_full$anc\_recon[pruned\_tree$tip.label, ]

head(env\_traits\_full)

dim(env\_traits\_full)

#Imputation - subset

env\_data\_subset <- ssbdata %>%

filter(nexus\_name %in% pruned\_ssb\_tree$tip.label) %>%

select(nexus\_name,

mean\_temp\_2019,

mean\_precip\_2019,

mean\_cloud\_2019,

mean\_biomass\_2019,

predator\_count) %>%

rename(species = nexus\_name)

env\_imputed\_subset <- phylopars(trait\_data = env\_data\_subset, tree = pruned\_ssb\_tree)

env\_traits\_subset <- env\_imputed\_subset$anc\_recon[pruned\_ssb\_tree$tip.label, ]

head(env\_traits\_subset)

dim(env\_traits\_subset)

#Phylogenetic Regression - Environment and SSB Occurrence (59 species)

matched\_species <- intersect(rownames(env\_traits\_full), data$nexus\_name)

env\_traits\_matched <- env\_traits\_full[matched\_species, ]

data\_matched <- data %>%

filter(nexus\_name %in% matched\_species) %>%

mutate(ssb\_presence = as.numeric(ssb\_presence)) # Make sure it's numeric

rownames(data\_matched) <- data\_matched$nexus\_name

env\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

env\_traits\_matched

)

env\_model\_data <- env\_model\_data[!is.na(env\_model\_data$SSB\_binary), ]

env\_model\_data <- env\_model\_data %>%

mutate(across(

c(mean\_temp\_2019, mean\_precip\_2019, mean\_cloud\_2019, mean\_biomass\_2019, predator\_count),

scale

))

env\_phyloglm <- phyloglm(

SSB\_binary ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 + predator\_count,

data = env\_model\_data,

phy = pruned\_tree,

method = "logistic\_MPLE",

btol = 12

)

summary(env\_phyloglm)

#Phylogenetic Regression - Environment and SSB Prevalence (23 species)

rownames(ssbdata) <- ssbdata$nexus\_name

species\_subset <- rownames(env\_traits\_subset)

prevalence\_vec <- ssbdata[species\_subset, "ssb\_prevalence"]

env\_subset\_data <- data.frame(

ssb\_presence = prevalence\_vec,

env\_traits\_subset[species\_subset, ]

)

env\_subset\_data <- env\_subset\_data[!is.na(env\_subset\_data$ssb\_presence), ]

env\_subset\_data <- env\_subset\_data %>%

mutate(across(

c(mean\_temp\_2019, mean\_precip\_2019, mean\_cloud\_2019, mean\_biomass\_2019, predator\_count),

scale

))

final\_species <- rownames(env\_subset\_data)

final\_ssb\_tree <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, final\_species))

length(final\_ssb\_tree$tip.label)

nrow(env\_subset\_data)

env\_phylolm <- phylolm(

ssb\_presence ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 + predator\_count,

data = env\_subset\_data,

phy = pruned\_ssb\_tree,

model = "lambda"

)

summary(env\_phylolm)

##### LIFE HISTORY #####

#Imputation - full dataset

lifehist\_data\_full <- data %>%

filter(nexus\_name %in% pruned\_tree$tip.label) %>%

select(nexus\_name,

size\_dimorphism\_presence,

median\_body\_size,

body\_size\_range,

lifespan,

sex\_ratio) %>%

rename(

species = nexus\_name,

sexual\_dimorphism = size\_dimorphism\_presence,

median\_body\_mass = median\_body\_size,

bodymass\_range = body\_size\_range,

adult\_sex\_ratio = sex\_ratio

)

lifehist\_data\_full[, -1] <- lapply(lifehist\_data\_full[, -1], function(x) as.numeric(as.character(x)))

lifehist\_imputed\_full <- phylopars(trait\_data = lifehist\_data\_full, tree = pruned\_tree)

lifehist\_traits\_full <- lifehist\_imputed\_full$anc\_recon[pruned\_tree$tip.label, ]

head(lifehist\_traits\_full)

#Imputation - subset

lifehist\_data\_subset <- ssbdata %>%

filter(nexus\_name %in% pruned\_ssb\_tree$tip.label) %>%

select(nexus\_name,

size\_dimorphism\_presence,

median\_body\_size,

body\_size\_range,

lifespan,

sex\_ratio) %>%

rename(

species = nexus\_name,

sexual\_dimorphism = size\_dimorphism\_presence,

median\_body\_mass = median\_body\_size,

bodymass\_range = body\_size\_range,

adult\_sex\_ratio = sex\_ratio

)

lifehist\_data\_subset[, -1] <- lapply(lifehist\_data\_subset[, -1], function(x) as.numeric(as.character(x)))

lifehist\_imputed\_subset <- phylopars(trait\_data = lifehist\_data\_subset, tree = pruned\_ssb\_tree)

lifehist\_traits\_subset <- lifehist\_imputed\_subset$anc\_recon[pruned\_ssb\_tree$tip.label, ]

head(lifehist\_traits\_subset)

#Phylogenetic Regression - Life History and SSB Occurrence (59 species)

lifehist\_model\_data <- data.frame(

SSB\_binary = data$ssb\_presence[match(rownames(lifehist\_traits\_full), data$nexus\_name)],

lifehist\_traits\_full

)

lifehist\_model\_data[, c("median\_body\_mass", "bodymass\_range", "lifespan", "adult\_sex\_ratio")] <-

scale(lifehist\_model\_data[, c("median\_body\_mass", "bodymass\_range", "lifespan", "adult\_sex\_ratio")])

lifehist\_phyloglm <- phyloglm(

SSB\_binary ~ sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan + adult\_sex\_ratio,

data = lifehist\_model\_data,

phy = pruned\_tree,

method = "logistic\_MPLE",

btol = 16

)

summary(lifehist\_phyloglm)

#Phylogenetic Regression - Life History and SSB Prevalence (23 species)

lifehist\_scaled\_subset <- scale(lifehist\_traits\_subset)

ssb\_prevalence\_vector <- setNames(ssbdata$ssb\_prevalence, ssbdata$nexus\_name)

lifehist\_model\_data\_subset <- data.frame(

ssb\_presence = ssb\_prevalence\_vector[rownames(lifehist\_traits\_subset)],

lifehist\_scaled\_subset

)

lifehist\_model\_subset <- phylolm(

ssb\_presence ~ sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan + adult\_sex\_ratio,

data = lifehist\_model\_data\_subset,

phy = final\_ssb\_tree,

model = "lambda"

)

summary(lifehist\_model\_subset)

##### SOCIAL #####

#Imputation - full dataset

social\_data\_full <- data.frame(

species = data$nexus\_name,

median\_group\_size = data$median\_group\_size

)

social\_data\_full$median\_group\_size <- as.numeric(as.character(social\_data\_full$median\_group\_size))

social\_data\_full <- social\_data\_full %>%

filter(species %in% pruned\_tree$tip.label)

social\_imputed\_full <- phylopars(trait\_data = social\_data\_full, tree = pruned\_tree)

social\_traits\_full <- social\_imputed\_full$anc\_recon

#Imputation - subset

social\_data\_subset <- data.frame(

species = ssbdata$nexus\_name,

median\_group\_size = ssbdata$median\_group\_size

)

social\_imputed\_subset <- phylopars(trait\_data = social\_data\_subset, tree = pruned\_ssb\_tree)

social\_traits\_subset <- social\_imputed\_subset$anc\_recon

#Phylogenetic Regressions - Social and SSB Occurrence (59 species) - not used in main text due to inflate SE

social\_traits\_full\_df <- as.data.frame(social\_traits\_full[pruned\_tree$tip.label, , drop=FALSE])

data\_clean <- data[data$nexus\_name %in% rownames(social\_traits\_full\_df), ]

social\_traits\_full\_ordered <- social\_traits\_full\_df[match(data\_clean$nexus\_name, rownames(social\_traits\_full\_df)), , drop=FALSE]

social\_model\_data\_full <- data.frame(

SSB\_binary = data\_clean$ssb\_presence,

median\_group\_size = social\_traits\_full\_ordered$median\_group\_size,

group\_structure = data\_clean$group\_structure,

hierarchical\_structure = data\_clean$hierarchical\_structure,

mating\_system = data\_clean$mating\_system,

infant\_rearing\_strategy = data\_clean$infant\_rearing\_strategy

)

rownames(social\_model\_data\_full) <- data\_clean$nexus\_name

social\_model\_data\_full\_complete <- social\_model\_data\_full[complete.cases(social\_model\_data\_full), ]

social\_model\_data\_full\_complete$group\_structure <- as.factor(social\_model\_data\_full\_complete$group\_structure)

social\_model\_data\_full\_complete$hierarchical\_structure <- as.factor(social\_model\_data\_full\_complete$hierarchical\_structure)

social\_model\_data\_full\_complete$mating\_system <- as.factor(social\_model\_data\_full\_complete$mating\_system)

social\_model\_data\_full\_complete$infant\_rearing\_strategy <- as.factor(social\_model\_data\_full\_complete$infant\_rearing\_strategy)

social\_model\_data\_full\_complete$group\_structure <- droplevels(social\_model\_data\_full\_complete$group\_structure)

social\_model\_data\_full\_complete$hierarchical\_structure <- droplevels(social\_model\_data\_full\_complete$hierarchical\_structure)

social\_model\_data\_full\_complete$mating\_system <- droplevels(social\_model\_data\_full\_complete$mating\_system)

social\_model\_data\_full\_complete$infant\_rearing\_strategy <- droplevels(social\_model\_data\_full\_complete$infant\_rearing\_strategy)

species\_to\_keep <- rownames(social\_model\_data\_full\_complete)

pruned\_tree\_complete <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, species\_to\_keep))

social\_model\_data\_full\_complete$group\_size\_scaled <- scale(social\_model\_data\_full\_complete$median\_group\_size)

model\_social\_full <- phyloglm(

SSB\_binary ~ group\_size\_scaled + group\_structure + hierarchical\_structure +

mating\_system + infant\_rearing\_strategy,

data = social\_model\_data\_full\_complete,

phy = pruned\_tree\_complete,

method = "logistic\_MPLE",

btol = 40

)

summary(model\_social\_full)

#Phylogenetic Regression - Simplified Social and SSB Occurrence (59 species - group size, group structure, hierarchy) - not used in main text (exploratory)

group\_formation\_plus\_hierarchy\_model <- phyloglm(

SSB\_binary ~ median\_group\_size + group\_structure + hierarchical\_structure,

data = social\_model\_data\_full\_complete,

phy = pruned\_tree\_complete,

method = "logistic\_MPLE",

btol = 100,

log.alpha.bound = 12

)

summary(group\_formation\_plus\_hierarchy\_model)

#Phylogenetic Regression - Simplified Social and SSB Occurrence (59 species - group size, group structure, collapsed hierarchy) - used in main text

data$hierarchical\_collapsed <- as.character(data$hierarchical\_structure)

data$hierarchical\_collapsed[data$hierarchical\_collapsed %in% c("Matriarchal", "Patriarchal")] <- "High"

data$hierarchical\_collapsed[data$hierarchical\_collapsed %in% c("Egalitarian", "Dominant pair")] <- "Low"

data$hierarchical\_collapsed <- factor(data$hierarchical\_collapsed, levels = c("Low", "High"))

social\_model\_data\_full <- data.frame(

species = data$nexus\_name,

SSB\_binary = data$ssb\_presence,

median\_group\_size = as.numeric(as.character(data$median\_group\_size)),

group\_structure = factor(data$group\_structure),

hierarchical\_structure = data$hierarchical\_collapsed

)

social\_model\_data\_full\_complete <- social\_model\_data\_full %>%

filter(complete.cases(.))

social\_model\_data\_full\_complete$median\_group\_size <- scale(social\_model\_data\_full\_complete$median\_group\_size)

rownames(social\_model\_data\_full\_complete) <- social\_model\_data\_full\_complete$species

species\_to\_keep <- rownames(social\_model\_data\_full\_complete)

missing\_species <- setdiff(species\_to\_keep, pruned\_tree$tip.label)

if(length(missing\_species) > 0){

message("Species in data NOT found in tree tips: ", paste(head(missing\_species, 10), collapse = ", "))

}

pruned\_tree\_complete <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, species\_to\_keep))

if(is.null(pruned\_tree\_complete)){

stop("All tips dropped from tree after pruning. Check species names!")

}

group\_formation\_model <- phyloglm(

SSB\_binary ~ median\_group\_size + group\_structure + hierarchical\_structure,

data = social\_model\_data\_full\_complete,

phy = pruned\_tree\_complete,

method = "logistic\_MPLE",

btol = 10,

log.alpha.bound = 12

)

summary(group\_formation\_model)

#Phylogenetic Regression - Simplified Social and SSB Occurrence (59 species - group size, collapsed group structure, collapsed hierarchy) - not used in main text (exploratory)

tree\_pruned <- drop.tip(tree, setdiff(tree$tip.label, data$nexus\_name))

data\_pruned <- data[data$nexus\_name %in% tree$tip.label, ]

data\_pruned$group\_structure\_collapsed <- ifelse(data\_pruned$group\_structure %in% c("Monogamous pairs", "Solitary"),

"mono/solitary", "complex")

data\_pruned$hierarchy\_collapsed <- ifelse(data\_pruned$hierarchical\_structure %in% c("Patriarchal", "Matriarchal"),

"high", "low")

data\_pruned$group\_size\_scaled <- as.numeric(scale(data\_pruned$median\_group\_size))

data\_social <- data\_pruned[

!is.na(data\_pruned$ssb\_presence) &

!is.na(data\_pruned$group\_size\_scaled),

]

data\_social$group\_structure\_collapsed <- factor(data\_social$group\_structure\_collapsed)

data\_social$group\_structure\_collapsed <- droplevels(data\_social$group\_structure\_collapsed)

data\_social$hierarchy\_collapsed <- factor(data\_social$hierarchy\_collapsed)

data\_social$hierarchy\_collapsed <- droplevels(data\_social$hierarchy\_collapsed)

tree\_social <- drop.tip(tree, setdiff(tree$tip.label, data\_social$nexus\_name))

table(data\_social$group\_structure\_collapsed)

table(data\_social$hierarchy\_collapsed)

model\_social\_phyloglm <- phyloglm(ssb\_presence ~ group\_size\_scaled +

group\_structure\_collapsed +

hierarchy\_collapsed,

data = data\_social,

phy = tree\_social,

method = "logistic\_MPLE")

summary(model\_social\_phyloglm)

#Phylogenetic regression - Social and SSB Prevalence (23 species) - model failed due to small sample size and limited variance in predictors

social\_traits\_subset <- ssbdata[, c("nexus\_name", "median\_group\_size", "hierarchical\_structure",

"group\_structure", "mating\_system", "infant\_rearing\_strategy")]

rownames(social\_traits\_subset) <- social\_traits\_subset$nexus\_name

social\_numeric <- social\_traits\_subset[, "median\_group\_size", drop = FALSE]

social\_categorical <- social\_traits\_subset[, c("hierarchical\_structure", "group\_structure",

"mating\_system", "infant\_rearing\_strategy")]

social\_scaled\_numeric <- scale(social\_numeric)

social\_scaled\_subset <- data.frame(social\_scaled\_numeric, social\_categorical)

rownames(social\_scaled\_subset) <- rownames(social\_traits\_subset)

ssb\_prevalence\_vector <- setNames(ssbdata$ssb\_prevalence, ssbdata$nexus\_name)

social\_model\_data\_subset <- data.frame(

ssb\_presence = ssb\_prevalence\_vector[rownames(social\_scaled\_subset)],

social\_scaled\_subset

)

social\_model\_data\_subset <- na.omit(social\_model\_data\_subset)

social\_model\_data\_subset$hierarchical\_structure <- as.factor(social\_model\_data\_subset$hierarchical\_structure)

social\_model\_data\_subset$group\_structure <- as.factor(social\_model\_data\_subset$group\_structure)

social\_model\_data\_subset$mating\_system <- as.factor(social\_model\_data\_subset$mating\_system)

social\_model\_data\_subset$infant\_rearing\_strategy <- as.factor(social\_model\_data\_subset$infant\_rearing\_strategy)

social\_model\_subset <- phylolm(

ssb\_presence ~ median\_group\_size + hierarchical\_structure + group\_structure +

mating\_system + infant\_rearing\_strategy,

data = social\_model\_data\_subset,

phy = final\_ssb\_tree,

model = "lambda"

)

summary(social\_model\_subset)

#Phylogenetic Regression - Simplified Social and SSB Prevalence (23 species - group size, collapsed group structure, collapsed hierarchy) - model failed due to small sample size and limited variance in predictors

ssbdata$hierarchical\_collapsed <- as.character(ssbdata$hierarchical\_structure)

ssbdata$hierarchical\_collapsed[ssbdata$hierarchical\_collapsed %in% c("Matriarchal", "Patriarchal")] <- "High"

ssbdata$hierarchical\_collapsed[ssbdata$hierarchical\_collapsed %in% c("Egalitarian", "Dominant pair")] <- "Low"

ssbdata$hierarchical\_collapsed[is.na(ssbdata$hierarchical\_structure)] <- NA

ssbdata$hierarchical\_collapsed <- factor(ssbdata$hierarchical\_collapsed, levels = c("Low", "High"))

ssbdata$group\_structure\_collapsed <- ifelse(

is.na(ssbdata$group\_structure), NA,

ifelse(ssbdata$group\_structure %in% c("Monogamous pairs", "Solitary"),

"mono/solitary", "complex")

)

ssbdata$group\_structure\_collapsed <- factor(ssbdata$group\_structure\_collapsed, levels = c("complex", "mono/solitary"))

ssbdata$group\_size\_scaled <- scale(ssbdata$median\_group\_size)

ssbdata$log\_prevalence <- log1p(ssbdata$ssb\_prevalence)

rownames(ssbdata) <- ssbdata$nexus\_name

ssbtree\_pruned <- drop.tip(ssbtree, setdiff(ssbtree$tip.label, rownames(ssbdata)))

model\_data <- ssbdata[complete.cases(ssbdata[, c("log\_prevalence", "group\_size\_scaled",

"group\_structure\_collapsed", "hierarchical\_collapsed")]), ]

model\_prevalence <- phylolm(

log\_prevalence ~ group\_size\_scaled + group\_structure\_collapsed + hierarchical\_collapsed,

data = model\_data,

phy = drop.tip(ssbtree\_pruned, setdiff(ssbtree\_pruned$tip.label, rownames(model\_data))),

model = "lambda"

)

summary(model\_prevalence)

#####Save imputed data#####

env\_df <- as.data.frame(env\_traits\_full)

lifehist\_df <- as.data.frame(lifehist\_traits\_full)

social\_df <- as.data.frame(social\_traits\_full)

env\_df$species <- rownames(env\_df)

lifehist\_df$species <- rownames(lifehist\_df)

social\_df$species <- rownames(social\_df)

library(dplyr)

imputed\_all\_numeric <- env\_df %>%

full\_join(lifehist\_df, by = "species") %>%

full\_join(social\_df, by = "species")

write.csv(imputed\_all\_numeric, "imputed\_numeric\_traits\_full\_dataset.csv", row.names = FALSE)

##### FULL SSB OCCURENCE MODEL WITH ALL PREDICTORS ##### - not used in main text due to inflated SE

shared\_species <- Reduce(intersect, list(

rownames(env\_traits\_full),

rownames(lifehist\_traits\_full),

rownames(social\_traits\_full),

data$nexus\_name

))

env\_sub <- env\_traits\_full[shared\_species, ]

lifehist\_sub <- lifehist\_traits\_full[shared\_species, ]

social\_sub <- social\_traits\_full[shared\_species, ]

env\_scaled <- scale(env\_sub)

lifehist\_scaled <- scale(lifehist\_sub)

social\_scaled <- scale(social\_traits\_full[shared\_species, , drop = FALSE])

colnames(social\_scaled) <- "median\_group\_size"

meta\_vars <- data %>%

filter(nexus\_name %in% shared\_species) %>%

select(nexus\_name, ssb\_presence, group\_structure, hierarchical\_structure, mating\_system, infant\_rearing\_strategy) %>%

mutate(across(c(group\_structure, hierarchical\_structure, mating\_system, infant\_rearing\_strategy), factor))

rownames(meta\_vars) <- meta\_vars$nexus\_name

meta\_vars <- meta\_vars[, -which(colnames(meta\_vars) == "nexus\_name")]

combined\_df <- data.frame(

ssb\_presence = as.numeric(meta\_vars$ssb\_presence),

env\_scaled,

lifehist\_scaled,

group\_structure = meta\_vars$group\_structure,

hierarchical\_structure = meta\_vars$hierarchical\_structure,

mating\_system = meta\_vars$mating\_system,

infant\_rearing\_strategy = meta\_vars$infant\_rearing\_strategy,

group\_size = as.numeric(social\_scaled[, "median\_group\_size"])

)

combined\_species <- rownames(combined\_df)

combined\_tree <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, combined\_species))

model\_combined <- phyloglm(

ssb\_presence ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 +

predator\_count + sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan +

adult\_sex\_ratio + group\_size + group\_structure + hierarchical\_structure +

mating\_system + infant\_rearing\_strategy,

data = combined\_df,

phy = combined\_tree,

method = "logistic\_MPLE",

btol = 41

)

summary(model\_combined)

#Multicollinearity using VIF

library(car)

glm\_model <- glm(

ssb\_presence ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 +

predator\_count + sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan +

adult\_sex\_ratio + group\_size + group\_structure + hierarchical\_structure +

mating\_system + infant\_rearing\_strategy,

data = combined\_df,

family = binomial

)

vif(glm\_model)

#Collapsed overall SSB occurrence model - used in main text

combined\_df <- data.frame(

ssb\_presence = as.numeric(meta\_vars$ssb\_presence),

env\_scaled,

lifehist\_scaled,

group\_structure = meta\_vars$group\_structure,

hierarchical\_structure = meta\_vars$hierarchical\_structure,

mating\_system = meta\_vars$mating\_system,

infant\_rearing\_strategy = meta\_vars$infant\_rearing\_strategy,

group\_size = as.numeric(social\_scaled[, "median\_group\_size"])

)

combined\_df$group\_structure\_collapsed <- ifelse(

combined\_df$group\_structure %in% c("Multi-male-multi-female", "Single-male-multi-female", "Fission-fusion"),

"complex", "mono/solitary"

)

combined\_df$hierarchy\_collapsed <- ifelse(

combined\_df$hierarchical\_structure %in% c("Patriarchal", "Matriarchal"),

"high", "low"

)

combined\_df$mating\_system\_collapsed <- ifelse(

combined\_df$mating\_system == "Monogamous",

"monogamous", "non-monogamous"

)

combined\_df$infant\_rearing\_collapsed <- ifelse(

combined\_df$infant\_rearing\_strategy == "All-parental", "all-parental",

ifelse(combined\_df$infant\_rearing\_strategy == "Maternal", "maternal", "other-parent")

)

combined\_df$group\_structure\_collapsed <- factor(combined\_df$group\_structure\_collapsed, levels = c("mono/solitary", "complex"))

combined\_df$hierarchy\_collapsed <- factor(combined\_df$hierarchy\_collapsed, levels = c("low", "high"))

combined\_df$mating\_system\_collapsed <- factor(combined\_df$mating\_system\_collapsed, levels = c("monogamous", "non-monogamous"))

combined\_df$infant\_rearing\_collapsed <- factor(combined\_df$infant\_rearing\_collapsed, levels = c("maternal", "all-parental", "other-parent"))

vars\_to\_keep <- c(

"ssb\_presence",

"mean\_temp\_2019", "mean\_precip\_2019", "mean\_cloud\_2019", "mean\_biomass\_2019",

"predator\_count", "sexual\_dimorphism", "median\_body\_mass", "bodymass\_range", "lifespan",

"adult\_sex\_ratio", "group\_size",

"group\_structure\_collapsed", "hierarchy\_collapsed", "mating\_system\_collapsed", "infant\_rearing\_collapsed"

)

clean\_df <- combined\_df[, vars\_to\_keep] %>%

filter(complete.cases(.))

clean\_df$group\_structure\_collapsed <- droplevels(clean\_df$group\_structure\_collapsed)

clean\_df$hierarchy\_collapsed <- droplevels(clean\_df$hierarchy\_collapsed)

clean\_df$mating\_system\_collapsed <- droplevels(clean\_df$mating\_system\_collapsed)

clean\_df$infant\_rearing\_collapsed <- droplevels(clean\_df$infant\_rearing\_collapsed)

species\_final <- rownames(clean\_df)

tree\_final <- drop.tip(combined\_tree, setdiff(combined\_tree$tip.label, species\_final))

stopifnot(all(rownames(clean\_df) %in% tree\_final$tip.label))

model\_collapsed <- phyloglm(

ssb\_presence ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 +

predator\_count + sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan +

adult\_sex\_ratio + group\_size +

group\_structure\_collapsed + hierarchy\_collapsed + mating\_system\_collapsed + infant\_rearing\_collapsed,

data = clean\_df,

phy = tree\_final,

method = "logistic\_MPLE",

btol = 29

)

summary(model\_collapsed)

#VIF check for multicollinearity

glm\_collapsed <- glm(

ssb\_presence ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 +

predator\_count + sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan +

adult\_sex\_ratio + group\_size +

group\_structure\_collapsed + hierarchy\_collapsed + mating\_system\_collapsed + infant\_rearing\_collapsed,

data = clean\_df,

family = binomial

)

vif(glm\_collapsed)

##### FULL SSB PREVALENCE MODEL WITH ALL PREDICTORS - model failed due to small sample size and limited variance in predictors #####

intersected\_species <- Reduce(intersect, list(

rownames(env\_traits\_subset),

rownames(lifehist\_traits\_subset),

rownames(social\_traits\_subset),

ssbdata$nexus\_name,

pruned\_ssb\_tree$tip.label

))

length(intersected\_species)

head(intersected\_species)

env\_final <- env\_traits\_subset[intersected\_species, ]

lifehist\_final <- lifehist\_traits\_subset[intersected\_species, ]

social\_final <- social\_traits\_subset[intersected\_species, , drop = FALSE]

meta\_vars <- ssbdata %>%

filter(nexus\_name %in% intersected\_species) %>%

select(nexus\_name, group\_structure, hierarchical\_structure, mating\_system, infant\_rearing\_strategy)

meta\_vars <- column\_to\_rownames(as.data.frame(meta\_vars), var = "nexus\_name")

meta\_vars$group\_structure\_collapsed <- ifelse(

meta\_vars$group\_structure %in% c("Multi-male-multi-female", "Single-male-multi-female", "Fission-fusion"),

"complex", "mono/solitary"

)

meta\_vars$hierarchy\_collapsed <- ifelse(

meta\_vars$hierarchical\_structure %in% c("Patriarchal", "Matriarchal"),

"high", "low"

)

meta\_vars$mating\_system\_collapsed <- ifelse(

meta\_vars$mating\_system == "Monogamous", "monogamous", "non-monogamous"

)

meta\_vars$infant\_rearing\_collapsed <- ifelse(

meta\_vars$infant\_rearing\_strategy == "All-parental", "all-parental",

ifelse(meta\_vars$infant\_rearing\_strategy == "Maternal", "maternal", "other-parent")

)

meta\_vars$group\_structure\_collapsed <- factor(meta\_vars$group\_structure\_collapsed, levels = c("mono/solitary", "complex"))

meta\_vars$hierarchy\_collapsed <- factor(meta\_vars$hierarchy\_collapsed, levels = c("low", "high"))

meta\_vars$mating\_system\_collapsed <- factor(meta\_vars$mating\_system\_collapsed, levels = c("monogamous", "non-monogamous"))

meta\_vars$infant\_rearing\_collapsed <- factor(meta\_vars$infant\_rearing\_collapsed, levels = c("maternal", "all-parental", "other-parent"))

combined\_df\_prevalence <- data.frame(

ssb\_prevalence = ssbdata[match(species\_from\_data, ssbdata$nexus\_name), "ssb\_prevalence"],

env\_final,

lifehist\_final,

group\_size = as.data.frame(social\_final)[intersected\_species, "median\_group\_size"],

meta\_vars,

row.names = intersected\_species

)

combined\_df\_prevalence <- combined\_df\_prevalence[!is.na(combined\_df\_prevalence$ssb\_prevalence), ]

final\_species\_prevalence <- rownames(combined\_df\_prevalence)

final\_tree\_prevalence <- drop.tip(

pruned\_ssb\_tree,

setdiff(pruned\_ssb\_tree$tip.label, final\_species\_prevalence)

)

combined\_df\_prevalence$group\_structure\_collapsed <- factor(

ifelse(combined\_df\_prevalence$group\_structure %in% c("Multi-male–multi-female", "Single-male–multi-female", "Fission–fusion"),

"complex", "mono/solitary"),

levels = c("mono/solitary", "complex")

)

combined\_df\_prevalence$hierarchy\_collapsed <- factor(

ifelse(combined\_df\_prevalence$hierarchical\_structure %in% c("Patriarchal", "Matriarchal"),

"high", "low"),

levels = c("low", "high")

)

combined\_df\_prevalence$mating\_system\_collapsed <- factor(

ifelse(combined\_df\_prevalence$mating\_system == "Monogamous",

"monogamous", "non-monogamous"),

levels = c("monogamous", "non-monogamous")

)

combined\_df\_prevalence$infant\_rearing\_collapsed <- factor(

ifelse(combined\_df\_prevalence$infant\_rearing\_strategy == "All-parental", "all-parental",

ifelse(combined\_df\_prevalence$infant\_rearing\_strategy == "Maternal", "maternal", "other-parent")),

levels = c("maternal", "all-parental", "other-parent")

)

model\_prevalence\_full <- phylolm(

ssb\_prevalence ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 +

predator\_count + sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan +

adult\_sex\_ratio + group\_size +

group\_structure\_collapsed + hierarchy\_collapsed + mating\_system\_collapsed + infant\_rearing\_collapsed,

data = combined\_df\_prevalence,

phy = final\_tree\_prevalence,

model = "lambda"

)

summary(model\_prevalence\_full)

##### STRUCTURAL EQUATION MODELLING #####

#SEM - Data cleaning

species\_all <- Reduce(intersect, list(

rownames(env\_traits\_full),

rownames(lifehist\_traits\_full),

rownames(social\_model\_data\_full\_complete)

))

env\_all <- env\_traits\_full[species\_all, ]

lifehist\_all <- lifehist\_traits\_full[species\_all, ]

social\_all <- social\_model\_data\_full\_complete[species\_all, ]

ssb\_vector <- data$ssb\_presence[match(species\_all, data$nexus\_name)]

sem\_data <- data.frame(

SSB = ssb\_vector,

env\_all,

lifehist\_all,

group\_size = social\_all$group\_size\_scaled,

group\_structure = droplevels(factor(social\_all$group\_structure)),

hierarchical\_structure = droplevels(factor(social\_all$hierarchical\_structure)),

mating\_system = droplevels(factor(social\_all$mating\_system)),

infant\_rearing\_strategy = droplevels(factor(social\_all$infant\_rearing\_strategy)),

species = species\_all

)

sem\_data <- sem\_data[complete.cases(sem\_data), ]

sem\_data <- sem\_data[sem\_data$mating\_system != "", ]

sem\_data$mating\_system <- droplevels(sem\_data$mating\_system)

sem\_data$SSB <- as.numeric(sem\_data$SSB)

sem\_data <- as.data.frame(lapply(sem\_data, function(x) {

if (is.numeric(x)) unname(as.numeric(x)) else x

}))

rownames(sem\_data) <- sem\_data$species

sem\_data$species <- NULL

sem\_tree <- ape::drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, rownames(sem\_data)))

sem\_data <- sem\_data[rownames(sem\_data) %in% sem\_tree$tip.label, ]

sem\_tree <- ape::drop.tip(sem\_tree, setdiff(sem\_tree$tip.label, rownames(sem\_data)))

sem\_data[] <- lapply(sem\_data, function(x) {

if (is.logical(x)) as.numeric(x) else x

})

#Check for multicollinearity

library(car)

vif\_model <- lm(

SSB ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 + predator\_count +

lifespan + sexual\_dimorphism + median\_body\_mass + bodymass\_range + adult\_sex\_ratio +

group\_size + group\_structure + hierarchical\_structure + mating\_system + infant\_rearing\_strategy,

data = sem\_data

)

vif\_scores <- vif(vif\_model)

print(vif\_scores)

#EFA

install.packages("psych")

install.packages("GPArotation")

library(psych)

library(GPArotation)

efa\_predictors <- sem\_data[, sapply(sem\_data, is.numeric)]

efa\_predictors <- efa\_predictors[, colnames(efa\_predictors) != "SSB"]

fa.parallel(efa\_predictors, fa = "fa")

efa\_result <- fa(efa\_predictors, nfactors = 3, rotate = "oblimin", fm = "ml")

print(efa\_result, cut = 0.3)

#Check categorical

table(sem\_data$group\_structure)

table(sem\_data$hierarchical\_structure)

table(sem\_data$mating\_system)

table(sem\_data$infant\_rearing\_strategy)

install.packages("rcompanion")

library(rcompanion)

group\_structure <- as.factor(sem\_data$group\_structure)

hierarchical\_structure <- as.factor(sem\_data$hierarchical\_structure)

mating\_system <- as.factor(sem\_data$mating\_system)

infant\_rearing <- as.factor(sem\_data$infant\_rearing\_strategy)

cramer\_gs\_hs <- cramerV(group\_structure, hierarchical\_structure)

cramer\_gs\_ms <- cramerV(group\_structure, mating\_system)

cramer\_gs\_ir <- cramerV(group\_structure, infant\_rearing)

cramer\_hs\_ms <- cramerV(hierarchical\_structure, mating\_system)

cramer\_hs\_ir <- cramerV(hierarchical\_structure, infant\_rearing)

cramer\_ms\_ir <- cramerV(mating\_system, infant\_rearing)

cat("Cramér's V between group\_structure and hierarchical\_structure:", cramer\_gs\_hs, "\n")

cat("Cramér's V between group\_structure and mating\_system:", cramer\_gs\_ms, "\n")

cat("Cramér's V between group\_structure and infant\_rearing\_strategy:", cramer\_gs\_ir, "\n")

cat("Cramér's V between hierarchical\_structure and mating\_system:", cramer\_hs\_ms, "\n")

cat("Cramér's V between hierarchical\_structure and infant\_rearing\_strategy:", cramer\_hs\_ir, "\n")

cat("Cramér's V between mating\_system and infant\_rearing\_strategy:", cramer\_ms\_ir, "\n")

#Filtered and layered SEM (removing temp, biomass, lifespan, mating system)

library(brms)

group\_vals <- social\_model\_data\_full\_complete[rownames(sem\_data), "group\_structure"]

group\_clean <- as.character(group\_vals)

group\_clean[group\_clean == "Fission-fusion"] <- "Multi-male-multi-female"

group\_factor <- factor(group\_clean,

levels = c("Solitary", "Monogamous pairs", "Single-male-multi-female", "Multi-male-multi-female"),

ordered = TRUE

)

sem\_data$group\_structure <- as.numeric(group\_factor)

sem\_data$hierarchical\_structure <- factor(

as.character(sem\_data$hierarchical\_structure),

levels = c("Egalitarian", "Dominant pair", "Matriarchal", "Patriarchal"),

ordered = TRUE

)

sem\_data$hierarchical\_structure <- as.numeric(sem\_data$hierarchical\_structure)

sem\_data$infant\_rearing\_strategy <- factor(

as.character(sem\_data$infant\_rearing\_strategy),

levels = c("Allo-parenting", "Shared parental care", "Paternal", "Maternal"),

ordered = TRUE

)

sem\_data$infant\_rearing\_strategy <- as.numeric(sem\_data$infant\_rearing\_strategy)

sem\_data <- sem\_data[complete.cases(sem\_data), ]

brms\_layered\_binary\_ssb <-

bf(SSB ~ group\_size + group\_structure + hierarchical\_structure +

infant\_rearing\_strategy +

sexual\_dimorphism + median\_body\_mass +

bodymass\_range + adult\_sex\_ratio +

mean\_precip\_2019 + mean\_cloud\_2019 + predator\_count,

family = bernoulli(link = "logit")) + # Binary outcome

bf(group\_size ~ sexual\_dimorphism + median\_body\_mass +

bodymass\_range + adult\_sex\_ratio) +

bf(group\_structure ~ sexual\_dimorphism + median\_body\_mass +

bodymass\_range + adult\_sex\_ratio) +

bf(hierarchical\_structure ~ sexual\_dimorphism + median\_body\_mass +

bodymass\_range + adult\_sex\_ratio) +

bf(infant\_rearing\_strategy ~ sexual\_dimorphism + median\_body\_mass +

bodymass\_range + adult\_sex\_ratio) +

bf(sexual\_dimorphism ~ mean\_precip\_2019 + mean\_cloud\_2019 + predator\_count) +

bf(median\_body\_mass ~ mean\_precip\_2019 + mean\_cloud\_2019 + predator\_count) +

bf(bodymass\_range ~ mean\_precip\_2019 + mean\_cloud\_2019 + predator\_count) +

bf(adult\_sex\_ratio ~ mean\_precip\_2019 + mean\_cloud\_2019 + predator\_count) +

set\_rescor(FALSE)

fit\_binary <- brm(brms\_layered\_binary\_ssb, data = sem\_data,

chains = 4, iter = 4000, cores = 4, seed = 123)

summary(fit\_binary)

install.packages("openxlsx")

library(openxlsx)

sem\_summary <- as.data.frame(summary(fit\_binary)$fixed)

sem\_summary$term <- rownames(sem\_summary)

sem\_summary <- sem\_summary %>%

mutate(

Response = sub("^(.\*?)\_.\*", "\\1", term),

Predictor = sub("^.\*?\_(.\*)", "\\1", term)

) %>%

select(Response, Predictor,

Estimate = Estimate,

SE = Est.Error,

`2.5% CI` = `l-95% CI`,

`97.5% CI` = `u-95% CI`,

`Bulk ESS` = Bulk\_ESS,

`Tail ESS` = Tail\_ESS)

write.xlsx(sem\_summary, "supplementary\_SEM\_results.xlsx", rowNames = FALSE)

#Model fit check

install.packages("loo")

library(loo)

loo\_fit <- loo(fit\_binary)

print(loo\_fit)

bayes\_R2(fit\_binary, resp = "SSB")

pp\_check(fit\_binary, resp = "SSB", type = "dens\_overlay")

##### PRIMATE GROUPS #####

#Cercopithecidae - Environment

data\_cercopithecidae <- data %>%

filter(Category == "Cercopithecidae") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence)) # Ensure numeric

matched\_species <- intersect(rownames(env\_traits\_full), data\_cercopithecidae$nexus\_name)

env\_traits\_matched <- env\_traits\_full[matched\_species, ]

data\_matched <- data\_cercopithecidae %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

env\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

env\_traits\_matched

)

env\_model\_data <- env\_model\_data[!is.na(env\_model\_data$SSB\_binary), ]

env\_model\_data <- env\_model\_data %>%

mutate(across(

c(mean\_temp\_2019, mean\_precip\_2019, mean\_cloud\_2019, mean\_biomass\_2019, predator\_count),

scale

))

pruned\_tree\_cercopithecidae <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, matched\_species))

env\_phyloglm\_cercopithecidae <- phyloglm(

SSB\_binary ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 + predator\_count,

data = env\_model\_data,

phy = pruned\_tree\_cercopithecidae,

method = "logistic\_MPLE"

)

summary(env\_phyloglm\_cercopithecidae)

#Cercopithecidae - Life History

data\_cercopithecidae <- data %>%

filter(Category == "Cercopithecidae") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

matched\_species <- intersect(rownames(lifehist\_traits\_full), data\_cercopithecidae$nexus\_name)

lifehist\_traits\_matched <- lifehist\_traits\_full[matched\_species, ]

data\_matched <- data\_cercopithecidae %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

lifehist\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

lifehist\_traits\_matched

)

lifehist\_model\_data <- lifehist\_model\_data[!is.na(lifehist\_model\_data$SSB\_binary), ]

lifehist\_model\_data <- lifehist\_model\_data %>%

mutate(across(

c(median\_body\_mass, bodymass\_range, lifespan, adult\_sex\_ratio),

scale

))

pruned\_tree\_cercopithecidae <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, matched\_species))

lifehist\_phyloglm\_cercopithecidae <- phyloglm(

SSB\_binary ~ sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan + adult\_sex\_ratio,

data = lifehist\_model\_data,

phy = pruned\_tree\_cercopithecidae,

method = "logistic\_MPLE",

btol = 20, log.alpha.bound = 6

)

summary(lifehist\_phyloglm\_cercopithecidae)

#Cercopithecidae - Social - model failed due to small sample size and limited variance in predictors

data\_cercopithecidae <- data %>%

filter(Category == "Cercopithecidae") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

data\_cercopithecidae <- data\_cercopithecidae %>%

mutate(

hierarchy\_collapsed = ifelse(

tolower(hierarchical\_structure) %in% c("patriarchal", "matriarchal"),

"high", "low"

),

hierarchy\_collapsed = factor(hierarchy\_collapsed, levels = c("low", "high"))

)

matched\_species <- intersect(rownames(social\_traits\_full), data\_cercopithecidae$nexus\_name)

social\_traits\_matched <- social\_traits\_full[matched\_species, , drop = FALSE]

data\_matched <- data\_cercopithecidae %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

social\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

group\_structure = data\_matched[matched\_species, "group\_structure"],

hierarchy\_collapsed = data\_matched[matched\_species, "hierarchy\_collapsed"],

group\_size = social\_traits\_matched[matched\_species, "median\_group\_size"]

)

social\_model\_data <- social\_model\_data %>%

filter(complete.cases(.)) %>%

mutate(group\_size\_scaled = scale(group\_size)) %>%

select(SSB\_binary, group\_size\_scaled, group\_structure, hierarchy\_collapsed)

pruned\_tree\_cercopithecidae <- drop.tip(

pruned\_tree,

setdiff(pruned\_tree$tip.label, rownames(social\_model\_data))

)

social\_phyloglm\_cercopithecidae <- phyloglm(

SSB\_binary ~ group\_size\_scaled + group\_structure + hierarchy\_collapsed,

data = social\_model\_data,

phy = pruned\_tree\_cercopithecidae,

method = "logistic\_MPLE",

btol = 20

)

summary(social\_phyloglm\_cercopithecidae)

#Platyrrhini - Environment

data\_platyrrhini <- data %>%

filter(Category == "Platyrrhini") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence)) # Ensure numeric

matched\_species <- intersect(rownames(env\_traits\_full), data\_platyrrhini$nexus\_name)

env\_traits\_matched <- env\_traits\_full[matched\_species, ]

data\_matched <- data\_platyrrhini %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

env\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

env\_traits\_matched

)

env\_model\_data <- env\_model\_data[!is.na(env\_model\_data$SSB\_binary), ]

env\_model\_data <- env\_model\_data %>%

mutate(across(

c(mean\_temp\_2019, mean\_precip\_2019, mean\_cloud\_2019, mean\_biomass\_2019, predator\_count),

scale

))

pruned\_tree\_platyrrhini <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, matched\_species))

env\_phyloglm\_platyrrhini <- phyloglm(

SSB\_binary ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 + predator\_count,

data = env\_model\_data,

phy = pruned\_tree\_platyrrhini,

method = "logistic\_MPLE",

btol = 43

)

summary(env\_phyloglm\_platyrrhini)

#Platyrrhini - Life History

data\_platyrrhini <- data %>%

filter(Category == "Platyrrhini") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence)) # Ensure numeric

matched\_species <- intersect(rownames(lifehist\_traits\_full), data\_platyrrhini$nexus\_name)

lifehist\_traits\_matched <- lifehist\_traits\_full[matched\_species, ]

data\_matched <- data\_platyrrhini %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

lifehist\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

lifehist\_traits\_matched

)

lifehist\_model\_data <- lifehist\_model\_data[!is.na(lifehist\_model\_data$SSB\_binary), ]

lifehist\_model\_data <- lifehist\_model\_data %>%

mutate(across(

c(median\_body\_mass, bodymass\_range, lifespan, adult\_sex\_ratio),

scale

))

pruned\_tree\_platyrrhini <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, matched\_species))

lifehist\_phyloglm\_platyrrhini <- phyloglm(

SSB\_binary ~ sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan + adult\_sex\_ratio,

data = lifehist\_model\_data,

phy = pruned\_tree\_platyrrhini,

method = "logistic\_MPLE"

)

summary(lifehist\_phyloglm\_platyrrhini)

#Platyrrhini - Social

data\_platyrrhini <- data %>%

filter(Category == "Platyrrhini") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

data\_platyrrhini <- data\_platyrrhini %>%

mutate(

hierarchy\_collapsed = ifelse(

tolower(hierarchical\_structure) %in% c("patriarchal", "matriarchal"),

"high", "low"

),

hierarchy\_collapsed = factor(hierarchy\_collapsed, levels = c("low", "high"))

)

matched\_species <- intersect(rownames(social\_traits\_full), data\_platyrrhini$nexus\_name)

social\_traits\_matched <- social\_traits\_full[matched\_species, , drop = FALSE]

data\_matched <- data\_platyrrhini %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

social\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

group\_structure = data\_matched[matched\_species, "group\_structure"],

hierarchy\_collapsed = data\_matched[matched\_species, "hierarchy\_collapsed"],

group\_size = social\_traits\_matched[matched\_species, "median\_group\_size"]

)

social\_model\_data <- social\_model\_data %>%

filter(complete.cases(.)) %>%

mutate(group\_size\_scaled = scale(group\_size)) %>%

select(SSB\_binary, group\_size\_scaled, group\_structure, hierarchy\_collapsed)

pruned\_tree\_platyrrhini <- drop.tip(

pruned\_tree,

setdiff(pruned\_tree$tip.label, rownames(social\_model\_data))

)

social\_phyloglm\_platyrrhini <- phyloglm(

SSB\_binary ~ group\_size\_scaled + group\_structure + hierarchy\_collapsed,

data = social\_model\_data,

phy = pruned\_tree\_platyrrhini,

method = "logistic\_MPLE",

btol = 20

)

summary(social\_phyloglm\_platyrrhini)

#Apes (Hominoidae and Hylobatidae) - Environment

data\_apes <- data %>%

filter(Category %in% c("Hominoidae", "Hylobatidae")) %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

matched\_species <- intersect(rownames(env\_traits\_full), data\_apes$nexus\_name)

env\_traits\_matched <- env\_traits\_full[matched\_species, ]

data\_matched <- data\_apes %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

env\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

env\_traits\_matched

)

env\_model\_data <- env\_model\_data[!is.na(env\_model\_data$SSB\_binary), ]

env\_model\_data <- env\_model\_data %>%

mutate(across(

c(mean\_temp\_2019, mean\_precip\_2019, mean\_cloud\_2019, mean\_biomass\_2019, predator\_count),

scale

))

pruned\_tree\_apes <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, matched\_species))

env\_phyloglm\_apes <- phyloglm(

SSB\_binary ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 + predator\_count,

data = env\_model\_data,

phy = pruned\_tree\_apes,

method = "logistic\_MPLE",

btol = 12

)

summary(env\_phyloglm\_apes)

#Apes (Hominoidae and Hylobatidae) - Life History

data\_apes <- data %>%

filter(Category %in% c("Hominoidae", "Hylobatidae")) %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

matched\_species <- intersect(rownames(lifehist\_traits\_full), data\_apes$nexus\_name)

lifehist\_traits\_matched <- lifehist\_traits\_full[matched\_species, ]

data\_matched <- data\_apes %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

lifehist\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

lifehist\_traits\_matched

)

lifehist\_model\_data <- lifehist\_model\_data[!is.na(lifehist\_model\_data$SSB\_binary), ]

lifehist\_model\_data <- lifehist\_model\_data %>%

mutate(across(

c(median\_body\_mass, bodymass\_range, lifespan, adult\_sex\_ratio),

scale

))

pruned\_tree\_apes <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, matched\_species))

lifehist\_phyloglm\_apes <- phyloglm(

SSB\_binary ~ sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan + adult\_sex\_ratio,

data = lifehist\_model\_data,

phy = pruned\_tree\_apes,

method = "logistic\_MPLE",

btol = 40, log.alpha.bound = 12

)

summary(lifehist\_phyloglm\_apes)

#Apes (Hominoidae and Hylobatidae) - Social - model failed due to small sample size and limited variance in predictors

data\_apes <- data %>%

filter(Category %in% c("Hominoidae", "Hylobatidae")) %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

data\_apes <- data\_apes %>%

mutate(

hierarchy\_collapsed = ifelse(

tolower(hierarchical\_structure) %in% c("patriarchal", "matriarchal"),

"high", "low"

),

hierarchy\_collapsed = factor(hierarchy\_collapsed, levels = c("low", "high"))

)

matched\_species <- intersect(rownames(social\_traits\_full), data\_apes$nexus\_name)

social\_traits\_matched <- social\_traits\_full[matched\_species, , drop = FALSE]

data\_matched <- data\_apes %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

social\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

group\_structure = data\_matched[matched\_species, "group\_structure"],

hierarchy\_collapsed = data\_matched[matched\_species, "hierarchy\_collapsed"],

group\_size = social\_traits\_matched[matched\_species, "median\_group\_size"]

)

social\_model\_data <- social\_model\_data %>%

filter(complete.cases(.)) %>%

mutate(group\_size\_scaled = scale(group\_size)) %>%

select(SSB\_binary, group\_size\_scaled, group\_structure, hierarchy\_collapsed)

pruned\_tree\_apes <- drop.tip(

pruned\_tree,

setdiff(pruned\_tree$tip.label, rownames(social\_model\_data))

)

social\_model\_data <- social\_model\_data %>%

mutate(

group\_structure = factor(group\_structure),

hierarchy\_collapsed = droplevels(hierarchy\_collapsed)

)

social\_phyloglm\_apes <- phyloglm(

SSB\_binary ~ group\_size\_scaled + group\_structure + hierarchy\_collapsed,

data = social\_model\_data,

phy = pruned\_tree\_apes,

method = "logistic\_MPLE",

btol = 20

)

summary(social\_phyloglm\_apes)

#Strepsirrhini - Environment

data\_streps <- data %>%

filter(Category == "Strepsirrhini") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

matched\_species <- intersect(rownames(env\_traits\_full), data\_streps$nexus\_name)

env\_traits\_matched <- env\_traits\_full[matched\_species, ]

data\_matched <- data\_streps %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

env\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

env\_traits\_matched

)

env\_model\_data <- env\_model\_data[!is.na(env\_model\_data$SSB\_binary), ]

env\_model\_data <- env\_model\_data %>%

mutate(across(

c(mean\_temp\_2019, mean\_precip\_2019, mean\_cloud\_2019, mean\_biomass\_2019, predator\_count),

scale

))

pruned\_tree\_streps <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, matched\_species))

env\_phyloglm\_streps <- phyloglm(

SSB\_binary ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 + mean\_biomass\_2019 + predator\_count,

data = env\_model\_data,

phy = pruned\_tree\_streps,

method = "logistic\_MPLE",

btol = 35

)

summary(env\_phyloglm\_streps)

#Strepsirrhini - Life History

data\_streps <- data %>%

filter(Category == "Strepsirrhini") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

matched\_species <- intersect(rownames(lifehist\_traits\_full), data\_streps$nexus\_name)

lifehist\_traits\_matched <- lifehist\_traits\_full[matched\_species, ]

data\_matched <- data\_streps %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

lifehist\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

lifehist\_traits\_matched

)

lifehist\_model\_data <- lifehist\_model\_data[!is.na(lifehist\_model\_data$SSB\_binary), ]

lifehist\_model\_data <- lifehist\_model\_data %>%

mutate(across(

c(median\_body\_mass, bodymass\_range, lifespan, adult\_sex\_ratio),

scale

))

pruned\_tree\_streps <- drop.tip(pruned\_tree, setdiff(pruned\_tree$tip.label, matched\_species))

lifehist\_phyloglm\_streps <- phyloglm(

SSB\_binary ~ sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan + adult\_sex\_ratio,

data = lifehist\_model\_data,

phy = pruned\_tree\_streps,

method = "logistic\_MPLE",

btol = 30, log.alpha.bound = 12

)

summary(lifehist\_phyloglm\_streps)

#Strepsirrhini - Social

data\_strepsirrhini <- data %>%

filter(Category == "Strepsirrhini") %>%

mutate(ssb\_presence = as.numeric(ssb\_presence))

data\_strepsirrhini <- data\_strepsirrhini %>%

mutate(

hierarchy\_collapsed = ifelse(

tolower(hierarchical\_structure) %in% c("patriarchal", "matriarchal"),

"high", "low"

),

hierarchy\_collapsed = factor(hierarchy\_collapsed, levels = c("low", "high"))

)

matched\_species <- intersect(rownames(social\_traits\_full), data\_strepsirrhini$nexus\_name)

social\_traits\_matched <- social\_traits\_full[matched\_species, , drop = FALSE]

data\_matched <- data\_strepsirrhini %>%

filter(nexus\_name %in% matched\_species)

rownames(data\_matched) <- data\_matched$nexus\_name

social\_model\_data <- data.frame(

SSB\_binary = data\_matched[matched\_species, "ssb\_presence"],

group\_structure = data\_matched[matched\_species, "group\_structure"],

hierarchy\_collapsed = data\_matched[matched\_species, "hierarchy\_collapsed"],

group\_size = social\_traits\_matched[matched\_species, "median\_group\_size"]

)

social\_model\_data <- social\_model\_data %>%

filter(complete.cases(.)) %>%

mutate(

group\_size\_scaled = scale(group\_size),

group\_structure = factor(group\_structure),

hierarchy\_collapsed = droplevels(hierarchy\_collapsed)

) %>%

select(SSB\_binary, group\_size\_scaled, group\_structure, hierarchy\_collapsed)

pruned\_tree\_strepsirrhini <- drop.tip(

pruned\_tree,

setdiff(pruned\_tree$tip.label, rownames(social\_model\_data))

)

social\_phyloglm\_strepsirrhini <- phyloglm(

SSB\_binary ~ group\_size\_scaled + group\_structure + hierarchy\_collapsed,

data = social\_model\_data,

phy = pruned\_tree\_strepsirrhini,

method = "logistic\_MPLE",

btol = 50

)

summary(social\_phyloglm\_strepsirrhini)

##### SEX DIFFERENCES IN SSB #####

#Male SSB and environment

ssbdata$logMale <- log1p(ssbdata$male\_ssb / ssbdata$sampling\_hours)

ssbdata$logFemale <- log1p(ssbdata$female\_ssb / ssbdata$sampling\_hours)

log\_male\_vec <- setNames(ssbdata$logMale, ssbdata$nexus\_name)

env\_male\_df <- data.frame(

male\_ssb = log\_male\_vec[rownames(env\_traits\_subset)],

env\_traits\_subset

)

env\_male\_df <- env\_male\_df[complete.cases(env\_male\_df), ]

env\_male\_df[, c("mean\_temp\_2019", "mean\_precip\_2019", "mean\_cloud\_2019",

"mean\_biomass\_2019", "predator\_count")] <-

scale(env\_male\_df[, c("mean\_temp\_2019", "mean\_precip\_2019", "mean\_cloud\_2019",

"mean\_biomass\_2019", "predator\_count")])

tree\_env\_male <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, rownames(env\_male\_df)))

model\_male\_env <- phylolm(

male\_ssb ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 +

mean\_biomass\_2019 + predator\_count,

data = env\_male\_df,

phy = tree\_env\_male,

model = "lambda"

)

summary(model\_male\_env)

#Male SSB and life history

log\_male\_vec <- setNames(ssbdata$logMale, ssbdata$nexus\_name)

lifehist\_male\_df <- data.frame(

male\_ssb = log\_male\_vec[rownames(lifehist\_traits\_subset)],

lifehist\_traits\_subset

)

lifehist\_male\_df <- lifehist\_male\_df[complete.cases(lifehist\_male\_df), ]

lifehist\_male\_df[, c("median\_body\_mass", "bodymass\_range", "lifespan", "adult\_sex\_ratio")] <-

scale(lifehist\_male\_df[, c("median\_body\_mass", "bodymass\_range", "lifespan", "adult\_sex\_ratio")])

tree\_lifehist\_male <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, rownames(lifehist\_male\_df)))

model\_male\_lifehist <- phylolm(

male\_ssb ~ sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan + adult\_sex\_ratio,

data = lifehist\_male\_df,

phy = tree\_lifehist\_male,

model = "lambda"

)

summary(model\_male\_lifehist)

#Male SSB and social - model failed due to small sample size and limited variance in predictors

log\_male\_vec <- setNames(ssbdata$logMale, ssbdata$nexus\_name)

social\_male\_df <- data.frame(

male\_ssb = log\_male\_vec[rownames(social\_traits\_subset)],

social\_traits\_subset

)

social\_male\_df <- social\_male\_df[complete.cases(social\_male\_df), ]

social\_male\_df[, c("median\_group\_size")] <-

scale(social\_male\_df[, c("median\_group\_size")])

tree\_social\_male <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, rownames(social\_male\_df)))

model\_male\_social <- phylolm(

male\_ssb ~ group\_structure + hierarchical\_structure + mating\_system + infant\_rearing\_strategy + median\_group\_size,

data = social\_male\_df,

phy = tree\_social\_male,

model = "lambda"

)

summary(model\_male\_social)

#Male SSB and simplified social - model failed due to small sample size and limited variance in predictors

log\_male\_vec <- setNames(ssbdata$logMale, ssbdata$nexus\_name)

social\_male\_df <- data.frame(

male\_ssb = log\_male\_vec[rownames(social\_traits\_subset)],

group\_structure = social\_traits\_subset$group\_structure,

hierarchical\_structure = social\_traits\_subset$hierarchical\_structure,

median\_group\_size = social\_traits\_subset$median\_group\_size

)

social\_male\_df$hierarchical\_structure <- as.character(social\_male\_df$hierarchical\_structure)

social\_male\_df$hierarchical\_structure[social\_male\_df$hierarchical\_structure %in% c("egalitarian", "dominant pair")] <- "low"

social\_male\_df$hierarchical\_structure[social\_male\_df$hierarchical\_structure %in% c("matriarchal", "patriarchal")] <- "high"

social\_male\_df$hierarchical\_structure <- as.factor(social\_male\_df$hierarchical\_structure)

social\_male\_df <- social\_male\_df[complete.cases(social\_male\_df), ]

social\_male\_df$median\_group\_size <- scale(social\_male\_df$median\_group\_size)

tree\_social\_male <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, rownames(social\_male\_df)))

model\_male\_social\_simple <- phylolm(

male\_ssb ~ group\_structure + hierarchical\_structure + median\_group\_size,

data = social\_male\_df,

phy = tree\_social\_male,

model = "lambda"

)

summary(model\_male\_social\_simple)

#Female SSB and environment

ssbdata$logFemale <- log1p(ssbdata$female\_ssb / ssbdata$sampling\_hours)

log\_female\_vec <- setNames(ssbdata$logFemale, ssbdata$nexus\_name)

env\_female\_df <- data.frame(

female\_ssb = log\_female\_vec[rownames(env\_traits\_subset)],

env\_traits\_subset

)

env\_female\_df <- env\_female\_df[complete.cases(env\_female\_df), ]

env\_female\_df[, c("mean\_temp\_2019", "mean\_precip\_2019", "mean\_cloud\_2019",

"mean\_biomass\_2019", "predator\_count")] <-

scale(env\_female\_df[, c("mean\_temp\_2019", "mean\_precip\_2019", "mean\_cloud\_2019",

"mean\_biomass\_2019", "predator\_count")])

tree\_env\_female <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, rownames(env\_female\_df)))

model\_female\_env <- phylolm(

female\_ssb ~ mean\_temp\_2019 + mean\_precip\_2019 + mean\_cloud\_2019 +

mean\_biomass\_2019 + predator\_count,

data = env\_female\_df,

phy = tree\_env\_female,

model = "lambda"

)

summary(model\_female\_env)

#Female SSB and life history

log\_female\_vec <- setNames(ssbdata$logFemale, ssbdata$nexus\_name)

lifehist\_female\_df <- data.frame(

female\_ssb = log\_female\_vec[rownames(lifehist\_traits\_subset)],

lifehist\_traits\_subset

)

lifehist\_female\_df <- lifehist\_female\_df[complete.cases(lifehist\_female\_df), ]

lifehist\_female\_df[, c("median\_body\_mass", "bodymass\_range", "lifespan", "adult\_sex\_ratio")] <-

scale(lifehist\_female\_df[, c("median\_body\_mass", "bodymass\_range", "lifespan", "adult\_sex\_ratio")])

tree\_lifehist\_female <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, rownames(lifehist\_female\_df)))

model\_female\_lifehist <- phylolm(

female\_ssb ~ sexual\_dimorphism + median\_body\_mass + bodymass\_range + lifespan + adult\_sex\_ratio,

data = lifehist\_female\_df,

phy = tree\_lifehist\_female,

model = "lambda"

)

summary(model\_female\_lifehist)

#Female SSB and social - model failed due to small sample size and limited variance in predictors

log\_female\_vec <- setNames(ssbdata$logFemale, ssbdata$nexus\_name)

social\_female\_df <- data.frame(

female\_ssb = log\_female\_vec[rownames(social\_traits\_subset)],

social\_traits\_subset

)

social\_female\_df <- social\_female\_df[complete.cases(social\_female\_df), ]

social\_female\_df[, c("median\_group\_size")] <-

scale(social\_female\_df[, c("median\_group\_size")])

tree\_social\_female <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, rownames(social\_female\_df)))

model\_female\_social <- phylolm(

female\_ssb ~ group\_structure + hierarchical\_structure + mating\_system + infant\_rearing\_strategy + median\_group\_size,

data = social\_female\_df,

phy = tree\_social\_female,

model = "lambda"

)

summary(model\_female\_social)

#Female SSB and simplified social - model failed due to small sample size and limited variance in predictors

log\_female\_vec <- setNames(ssbdata$logFemale, ssbdata$nexus\_name)

social\_female\_df <- data.frame(

female\_ssb = log\_female\_vec[rownames(social\_traits\_subset)],

group\_structure = social\_traits\_subset$group\_structure,

hierarchical\_structure = social\_traits\_subset$hierarchical\_structure,

median\_group\_size = social\_traits\_subset$median\_group\_size

)

social\_female\_df$hierarchical\_structure <- as.character(social\_female\_df$hierarchical\_structure)

social\_female\_df$hierarchical\_structure[social\_female\_df$hierarchical\_structure %in% c("egalitarian", "dominant pair")] <- "low"

social\_female\_df$hierarchical\_structure[social\_female\_df$hierarchical\_structure %in% c("matriarchal", "patriarchal")] <- "high"

social\_female\_df$hierarchical\_structure <- as.factor(social\_female\_df$hierarchical\_structure)

social\_female\_df <- social\_female\_df[complete.cases(social\_female\_df), ]

social\_female\_df$median\_group\_size <- scale(social\_female\_df$median\_group\_size)

tree\_social\_female <- drop.tip(pruned\_ssb\_tree, setdiff(pruned\_ssb\_tree$tip.label, rownames(social\_female\_df)))

model\_female\_social\_simple <- phylolm(

female\_ssb ~ group\_structure + hierarchical\_structure + median\_group\_size,

data = social\_female\_df,

phy = tree\_social\_female,

model = "lambda"

)

summary(model\_female\_social\_simple)

##### Species Similarity Checks #####

install.packages("dplyr")

install.packages("vegan")

install.packages("ggplot2")

library(dplyr)

library(vegan)

library(ggplot2)

library(tidyr)

df <- read.csv("Species list comparison.csv")

Species\_Search1 <- df$sexual\_behaviour\_search

Mentions\_1 <- df$sexual\_behavious\_mentions

Species\_Search2 <- df$ssb\_search

Mentions\_2 <- df$ssb\_mentions

#Jaccard Similarity Index

df <- df %>%

mutate(Presence\_1 = ifelse(is.na(Mentions\_1), 0, ifelse(Mentions\_1 > 0, 1, 0)),

Presence\_2 = ifelse(is.na(Mentions\_2), 0, ifelse(Mentions\_2 > 0, 1, 0)))

intersection <- sum(df$Presence\_1 & df$Presence\_2)

union <- sum(df$Presence\_1 | df$Presence\_2)

jaccard\_similarity <- intersection / union

print(paste("Jaccard similarity index:", round(jaccard\_similarity, 3)))

#Fisher Test

contingency\_table <- table(df$Presence\_1, df$Presence\_2)

fisher\_test <- fisher.test(contingency\_table)

print(fisher\_test)

#Spearman's rank correlation

spearman\_corr <- cor.test(Mentions\_1, Mentions\_2, method = "spearman")

print(spearman\_corr)