Overview over phylogenetic mixed models in Stoffel et al. (2018)

This document provides an overview over the different models in the paper with sufficient detail. To fully reproduce the results, please look at <https://github.com/mastoffel/>, where you can find the GitHub repository with the complete analysis pipeline.

library(MCMCglmm)  
library(ape)  
library(phytools)  
library(readr)  
library(dplyr)

## Data

1. Phylogeny of the 29 species.
2. All other data for modeling in all\_stats

tree\_final <- read.tree("../data/raw/phylogeny/29\_species\_10ktrees.tre")  
all\_stats <- as.data.frame(read\_csv("../data/processed/all\_stats\_29\_modeling.csv"))

## Modeling preparation

1. create inverse relatedness matrix
2. prior list for all models

inv\_phylo <- inverseA(tree\_final, nodes="TIPS",scale=FALSE)$Ainv #,scale=TRUE  
prior<-list(G=list(G1=list(V=1,nu=0.002)),R=list(V=1,nu=0.002))

1. Here, we use a smaller number of MCMC iterations (nitt), burning and thinning interval for all models. For the paper, we used nitt=1100000,burnin=100000,thin=1000.

nitt\_fast <- 1100  
burnin\_fast <- 100  
thin\_fast <- 10

## Bottleneck signatures, Breeding Habitat and Sexual Size Dimorphism (SSD)

**Corresponding Figure: Fig.3**

### Preparation

Standardize SSD by 2 sd to make estimates comparable with binary predictor breeding habitat, see Gelman (2008))

stats\_mod\_botsig <- all\_stats %>%   
 mutate(SSD = (SSD - mean(SSD) / (2\*sd(SSD))))

### Model 1: ~ breeding habitat + SSD

MCMCglmm(TPM80\_ratio ~ SSD + BreedingType,   
 random=~tip\_label, nodes = "TIPS",   
 family=c("gaussian"),ginverse=list(tip\_label=inv\_phylo),prior=prior,  
 data=stats\_mod\_botsig,nitt=nitt\_fast,burnin=burnin\_fast,thin=thin\_fast)

### Model 2: ~ breeding habitat + SSD

MCMCglmm(bot ~ SSD + BreedingType,   
 random=~tip\_label, nodes = "TIPS",   
 family=c("gaussian"),ginverse=list(tip\_label=inv\_phylo),prior=prior,  
 data=stats\_mod\_botsig,nitt=nitt\_fast,burnin=burnin\_fast,thin=thin\_fast)

## Determinants of genetic diversity

**Corresponding Figure: Fig.4**

### Preparation

Standardize all variables by 2sd for comparability with breeding habitat.

stats\_mod\_gen <- all\_stats %>%   
 mutate(logAbundance = ((logAbundance - mean(logAbundance)) / (2\*sd(logAbundance))),  
 SSD = (SSD - mean(SSD)) / (2\*sd(SSD)),  
 bot = (bot - mean(bot)) / (2\*sd(bot)),  
 TPM80\_ratio = (TPM80\_ratio - mean(TPM80\_ratio)) / (2\*sd(TPM80\_ratio)))

### Model 3: Allelic richness ~ log(Abundance) + Breeding habitat + SSD + +

MCMCglmm(num\_alleles\_mean ~ logAbundance + SSD + BreedingType + bot + TPM80\_ratio,   
 random=~tip\_label, nodes = "TIPS", # rcov =~us(trait):units  
 family=c("gaussian"),ginverse=list(tip\_label=inv\_phylo),prior=prior,  
 data=stats\_mod\_gen,nitt=nitt\_fast,burnin=burnin\_fast,thin=thin\_fast)

## IUCN, bottleneck signatures and genetic diversity

### Preparation

Reformat IUCN status to binary.

stats\_mod\_IUCN <- all\_stats %>%   
 mutate(IUCN\_binary = case\_when(IUCN\_rating == "vulnerable" ~ "concern",  
 IUCN\_rating == "near threatened" ~ "concern",  
 IUCN\_rating == "endangered" ~ "concern",  
 IUCN\_rating == "least concern" ~ "least concern"))

### Model 4a: Allelic richness ~ IUCN

MCMCglmm(num\_alleles\_mean ~ IUCN\_binary,   
 random=~tip\_label, nodes = "TIPS",  
 family=c("gaussian"),ginverse=list(tip\_label=inv\_phylo),prior=prior,  
 data=stats\_mod\_IUCN,nitt=nitt\_fast,burnin=burnin\_fast,thin=thin\_fast)

### Model 4b: ~ IUCN

MCMCglmm(TPM80\_ratio ~ IUCN\_binary, # , #+ Abundance BreedingType + BreedingType + Generation\_time  
 random=~tip\_label, nodes = "TIPS", # rcov =~us(trait):units  
 family=c("gaussian"),ginverse=list(tip\_label=inv\_phylo),prior=prior,  
 data=stats\_mod\_IUCN,nitt=nitt\_fast,burnin=burnin\_fast,thin=thin\_fast)

### Model 4c: ~ IUCN

MCMCglmm(bot ~ IUCN\_binary, # , #+ Abundance BreedingType + BreedingType + Generation\_time  
 random=~tip\_label, nodes = "TIPS", # rcov =~us(trait):units  
 family=c("gaussian"),ginverse=list(tip\_label=inv\_phylo),prior=prior,  
 data=stats\_mod\_IUCN,nitt=nitt\_fast,burnin=burnin\_fast,thin=thin\_fast)

## Supplementary model 1:

**Corresponding Figure: Fig.2**

This model elucidates the correlation between the two bottleneck measures, the proportion of loci in heterozygosity-excess (, here: TPM80\_ratio) under the TPM 80 and the ABC bottleneck model probability (, here: bot) across all species.

### Preparation

Standardize variables (by 1 sd, as no binary predictor is involved, otherwise by 2 sd to make estimates comparable across predictors, see Gelman (2008))

stats\_mod\_gen <- all\_stats %>%   
 mutate(TPM80\_ratio\_stand = as.numeric(scale(TPM80\_ratio)))

### ~

MCMCglmm(bot ~ TPM80\_ratio\_stand,   
 random=~tip\_label, nodes = "TIPS",   
 family=c("gaussian"),ginverse=list(tip\_label=inv\_phylo),prior=prior,  
 data=stats\_mod\_gen,nitt=nitt\_fast,burnin=burnin\_fast,thin=thin\_fast)