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.

```r
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

```r
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

```r
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))
```

(3) 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.

```r
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))

```r
stats_mod_botsig <- all_stats %>%
    mutate(SSD = (SSD - mean(SSD) / (2*sd(SSD))))
```

Model 1: $prop_{het-exc}$ ~ **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: p_{bot} ~ **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 + $prop_{het-exc}$ **+** p_{bot}

```
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: $prop_{het-exc}$ ~ 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: p_{bot} ~ 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 ($prop_{het-exc}$, here: TPM80_ratio) under the TPM 80 and the ABC bottleneck model probability (p_{bot}, 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)))
```

p_{bot} ~ $prop_{het-exc}$

```
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)
```