Olfactory camouflage and communication in birds

R code for the comparative analysis
on sex differences in preen oil composition

L.A. Grieves, M. Gilles, I.C. Cuthill, T. Székely, E.A. MacDougall-Shackleton, B.A. Caspers

Contents

Required packages

```
library(ape)
library(MCMCglmm)
library(coda)
library(plotMCMC)
```

1 Data

1.1 Data on sex differences

Import data

```
df <- read.csv("data_sex.csv")
```

Sample size = number of occurences (1 occurence = 1 season for a species)

```
nrow(df)
```

```
## [1] 75
```

Number of occurences in each category of each variable

```
ss.sex <- data.frame(table(df$SexDiff)["0"],table(df$SexDiff)["1"])
colnames(ss.sex) <- c("no","yes")
rownames(ss.sex) <- c("sex difference")
ss.sex
```

```
##                no yes
## sex difference 38  37
```

```
ss.inc <- data.frame(table(df$Inc)["0"],table(df$Inc)["1"])
colnames(ss.inc) <- c("biparental","uniparental")
rownames(ss.inc) <- c("incubation type")
ss.inc
```

```
##                 biparental uniparental
## incubation type         38          37
```

```
ss.season <- data.frame(table(df$Season)["0"],table(df$Season)["1"])
colnames(ss.season) <- c("nonbreeding","breeding")
rownames(ss.season) <- c("season")
ss.season
```

```
##        nonbreeding breeding
## season          22       53
```

Number of species

```r
length(unique(df$Species))
```

[1] 49

Number of studies

```r
length(unique(df$Study))
```

[1] 39

1.2 Phylogenetic data

Import phylogenetic data

```r
# birdtree.org, 1000 trees, from source: Stage2 MayrAll Hackett
tree <- read.nexus("data_phylo.nex")
# this tree contains all species included in the review
class(tree)
```

Build consensus tree

```r
# Concensus tree: given a series of trees, this function returns the consensus tree.
# By default, the strict-consensus tree is computed.
# To get the majority-rule consensus tree, use p = 0.5.
# Any value between 0.5 and 1 can be used.
consensustree <- consensus(tree, p = 0.5, check.labels = TRUE)
class(consensustree)

# Check that the tree covers all the species
species <- df$Phylo
setdiff(species, as.character(consensustree$tip.label))
# 0 species listed in our data but not in the tree
setdiff(as.character(consensustree$tip.label),species)
# species listed in the tree but not in our data (df = data with sex differences)

# Trim tree to have only species where sex differences are known
subtree <- drop.tip(consensustree, tip=setdiff(consensustree$tip.label,df$Phylo))

# Check that the tree covers all the species
setdiff(species, as.character(subtree$tip.label))
# 0 species listed in our data but not in the tree
setdiff(as.character(subtree$tip.label),species)
# 0 species listed in the tree but not in our data

# Dealing with polytomies
is.binary.tree(subtree) # FALSE > there are some polytomies
subtree_random <- multi2di(subtree,random=TRUE)
# randomization approach to take care of the polytomies
is.binary.tree(subtree_random) # TRUE > no more polytomies
```

Some name synonyms had to be fixed to match the list of bird tree: Poecile > Parus, Chlorus > Carduelis, Acanthis > Carduelis NB: Cory's shearwater (*Calonectris borealis*) is considered a subspecies of Scopoli's shearwater (*C. diomedea*) by some authors. In birdtree.org Cory's shearwater (*Calonectris borealis*) is absent, so we used the name of Scopoli's shearwater (i.e. Calonectris diomedea) as they are phylogenetically close/equivalent

Plot tree

```
# Plot tree
plot(subtree_random, cex=.7, label.offset =.1, no.margin = TRUE)
```
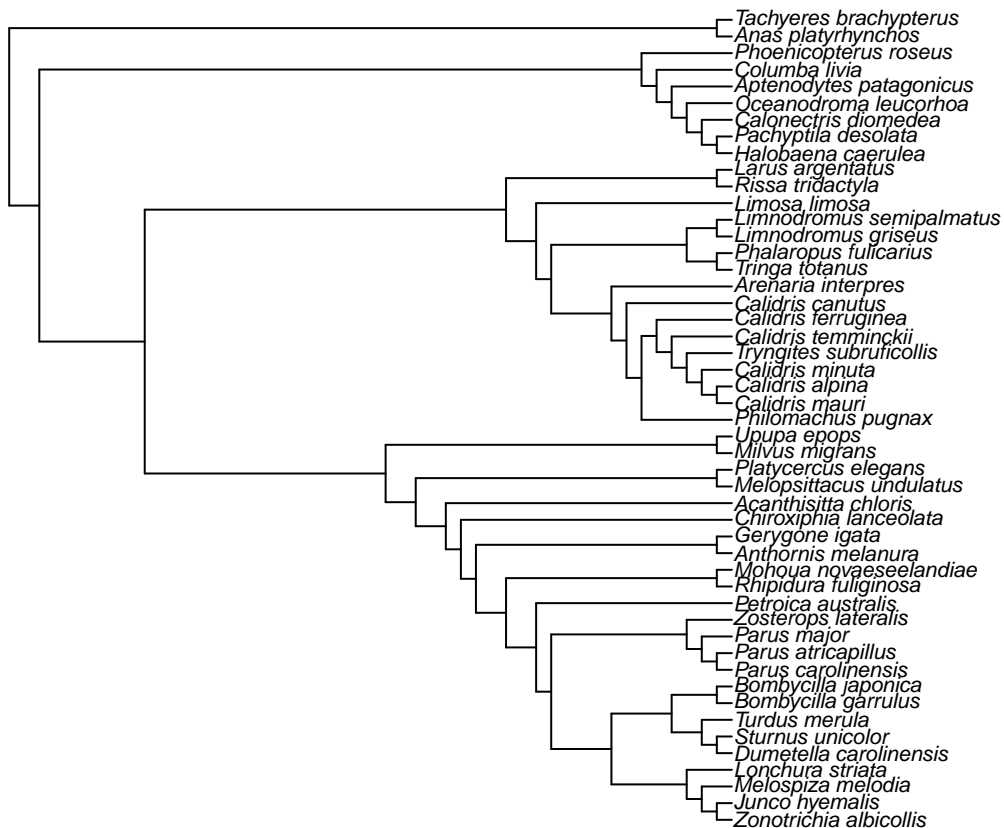

Get phylogenetic matrix

```
# Compute branch lengths of tree
phylo_branch <- compute.brlen(subtree_random, method = "Grafen", power = 1)

# Check tree is ultrametric
is.ultrametric(phylo_branch)

# Calculate the inverse of the matrix of phylogenetic relatedness (to be used in MCMCglmm)
inv.phylo <- inverseA(phylo_branch,nodes="TIPS",scale=TRUE)
```

2 MCMCglmm

2.1 Prior

Because we have no a priori belief on the value of the parameter, we set a weakly informative prior (inverse-gamma)

```r
prior2 <- list(G = list(G1 = list(V = 1, n = 0.002),
                        G2 = list(V = 1, n = 0.002)),
               R = list(V = 1, n = 0.002))
```

2.2 Model

Model parameters: burnin and thin should be big enough to get rid of autocorrelation and ensure convergence, nitt should be big enough to have sufficient sample size (nitt-burnin/thin >=1000).

```r
# Model 1
mcmc1 <- MCMCglmm(SexDiff ~ Season + Inc,
                  random = ~Phylo+Species,
                  family = "categorical",
                  data = df,
                  prior = prior2,
                  ginverse =  list(Phylo = inv.phylo$Ainv),
                  nitt=10000000,burnin=5000,thin=2000,
                  verbose = FALSE)
```

3 Model verifications

Diagnostics

```r
raftery.diag(mcmc1) # tells you how many MCMC iterations you need to run for convergence
```

```
##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
## Probability (s) = 0.95
##
## You need a sample size of at least 3746 with these values of q, r and s
```

3.1 Check autocorrelation

Check that autocorrelation is low

```r
autocorr(mcmc1$Sol[,1])
```

```
## , , 1
##
##                  [,1]
```

```
## Lag 0      1.000000000
## Lag 2000    0.002661889
## Lag 10000   0.010373734
## Lag 20000  -0.010200864
## Lag 1e+05   0.005879281
```
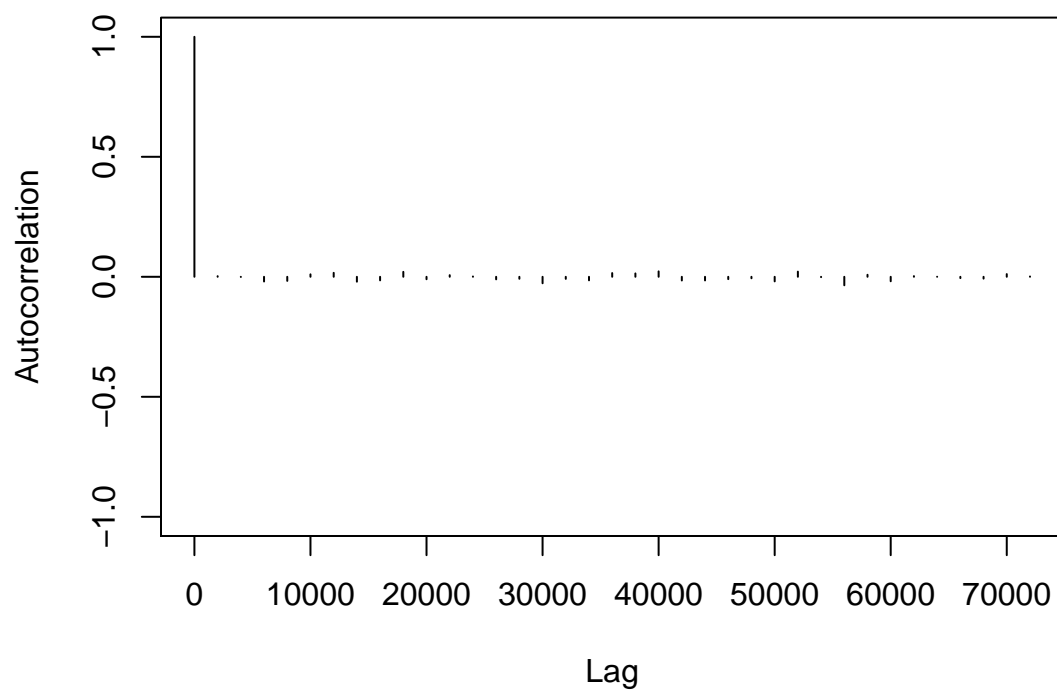
autocorr(mcmc1**$**VCV)

```
## , , Phylo
##
##                 Phylo       Species        units
## Lag 0      1.000000000 -0.021699736  0.0341595230
## Lag 2000    0.018313059  0.002205689 -0.0177896810
## Lag 10000 -0.009429020  0.015071680  0.0234835290
## Lag 20000  0.005574561 -0.003302086  0.0003658392
## Lag 1e+05   0.009518387 -0.003077166  0.0048287579
##
## , , Species
##
##                 Phylo       Species        units
## Lag 0     -0.021699736  1.000000000 -0.038509194
## Lag 2000  -0.008277846  0.008697507 -0.018486979
## Lag 10000  0.018169507 -0.008131379  0.014664103
## Lag 20000  0.012628517 -0.019654551  0.006118190
## Lag 1e+05 -0.006373815  0.014771281  0.002426294
##
## , , units
##
##                 Phylo       Species        units
## Lag 0      0.034159523 -0.038509194  1.000000000
## Lag 2000  -0.005702611 -0.003008078 -0.009079765
## Lag 10000 -0.032844420 -0.003708890  0.010989793
## Lag 20000 -0.006318522  0.010445894  0.013007115
## Lag 1e+05  0.025552133 -0.010334366 -0.010371027
```
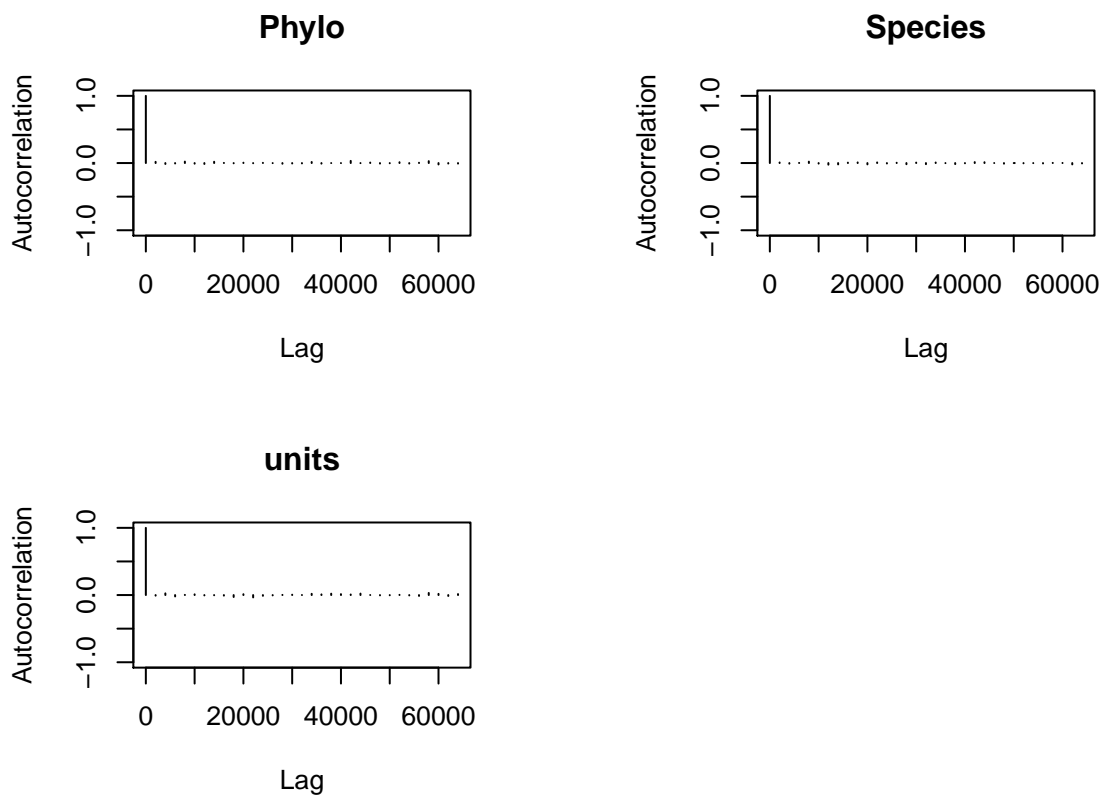
<.05 after Lag 10000 = low autocorrelation

Visualize autocorrelation
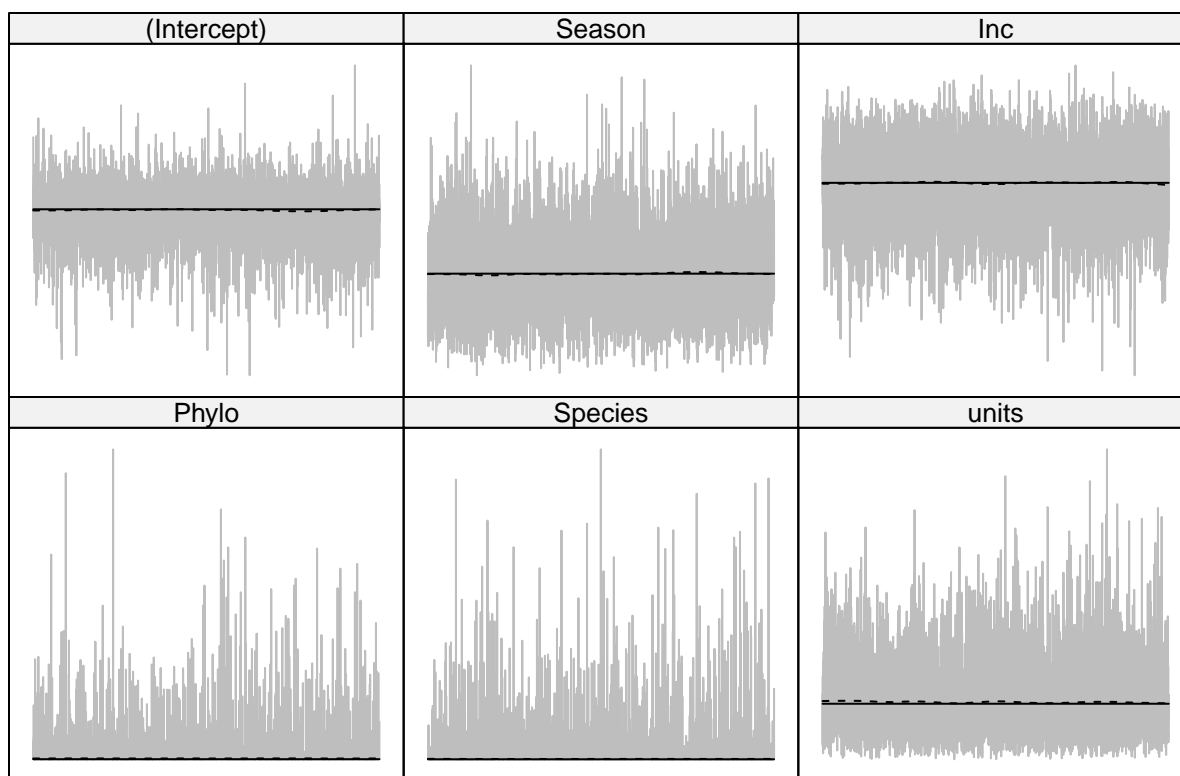
par(mar=**c**(5,5,4,4))
autocorr.plot(mcmc1**$**Sol[,1])

```r
autocorr.plot(mcmc1$VCV)
```
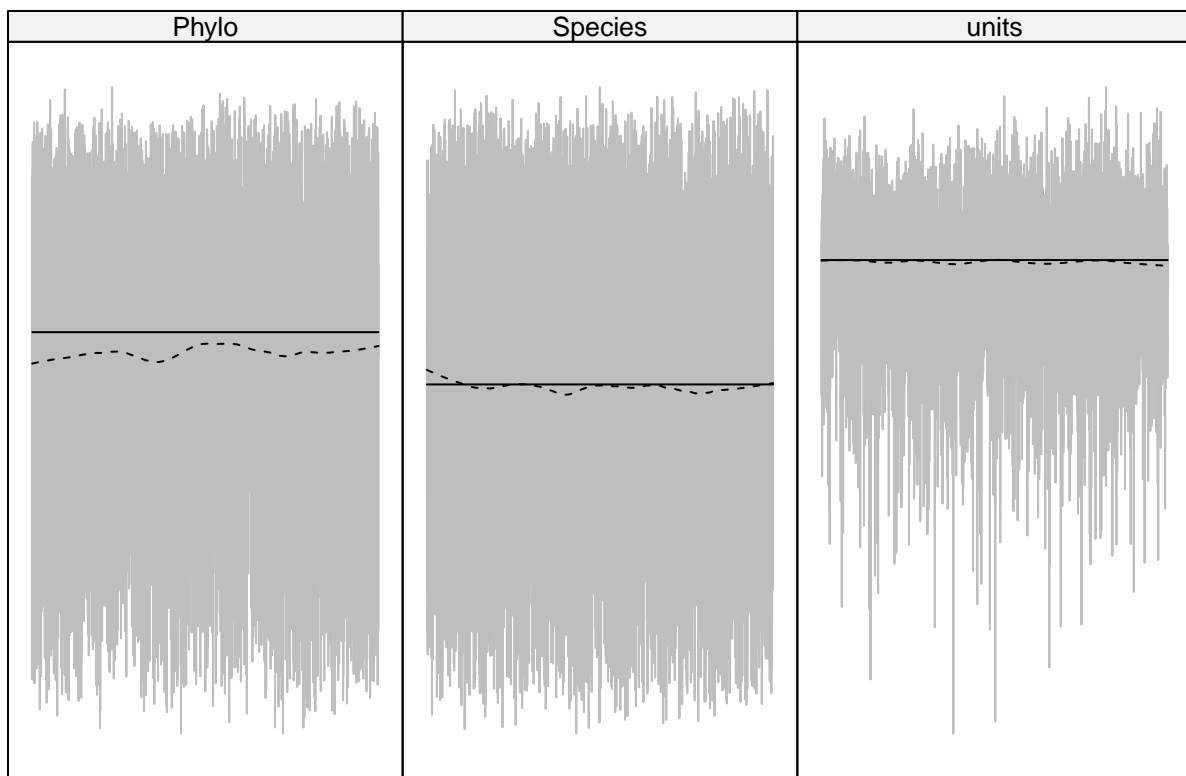
Phylo

Species

units

Autocorrelation quickly curves to almost nothing.

View traces to make sure there are no patterns of autocorrelation

```r
allChains <- as.mcmc(cbind(mcmc1$Sol,mcmc1$VCV))
plotTrace(allChains)
```
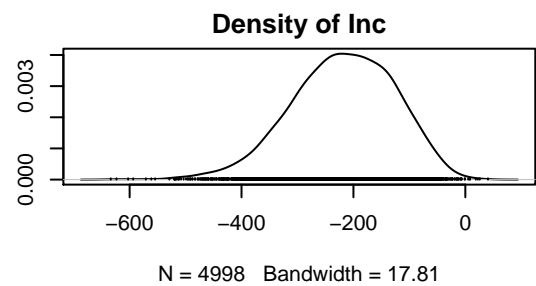
```
logChain<-log10(mcmc1$VCV)
plotTrace(logChain)
```

3.2 Check convergence

Plot trace and density of the inferred intercepts and variances

```r
par(mar=c(4,4,2,2))
plot(mcmc1$Sol) # intercept
```

Trace of (Intercept)

Density of (Intercept)

N = 4998   Bandwidth = 18.22

Trace of Season

Density of Season

N = 4998   Bandwidth = 23.75

Trace of Inc

Density of Inc

N = 4998   Bandwidth = 17.81
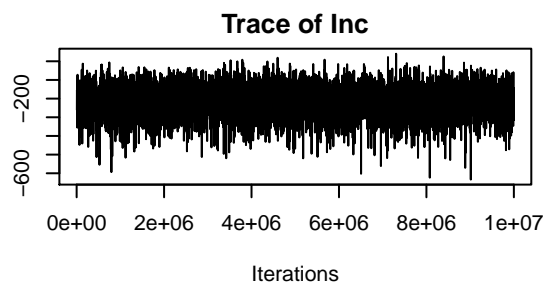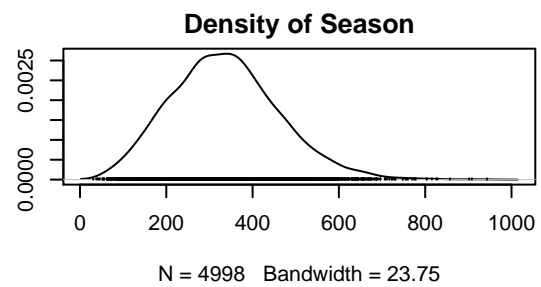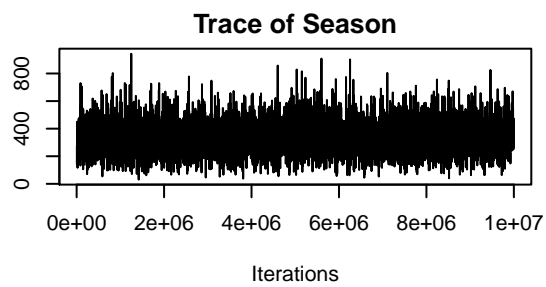
```r
plot(mcmc1$VCV) # variance
```
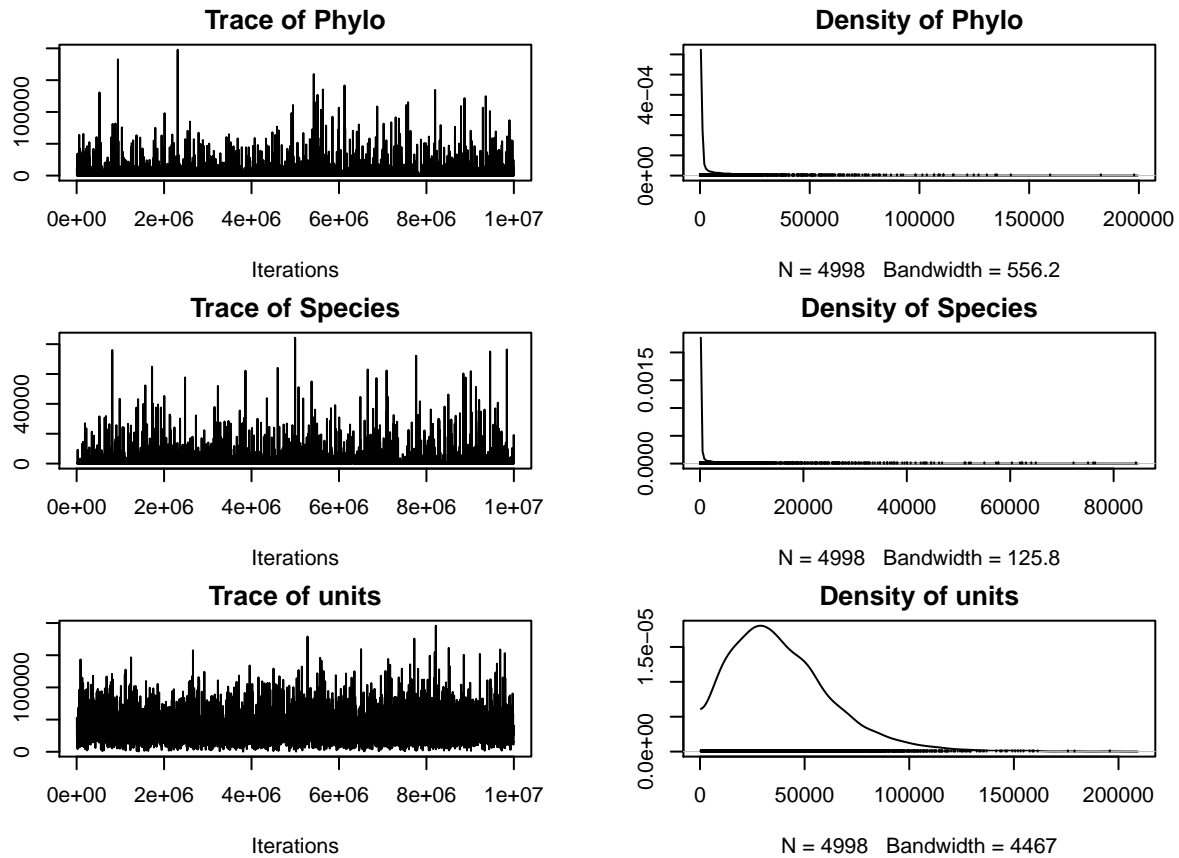
Graphs look like a "hairy caterpillar", suggesting good convergence.

3.3 Gelman-Rubin statistics

Comparison of the same model run multiple times

Run the same model two more times

```r
# Model 2
mcmc2 <- MCMCglmm(SexDiff ~ Season + Inc,
                  random = ~Phylo+Species,
                  family = "categorical",
                  data = df,
                  prior = prior2,
                  ginverse =  list(Phylo = inv.phylo$Ainv),
                  nitt=10000000,burnin=5000,thin=2000,
                  verbose = FALSE)


# Model 3
mcmc3 <- MCMCglmm(SexDiff ~ Season + Inc,
                  random = ~Phylo+Species,
                  family = "categorical",
                  data = df,
                  prior = prior2,
                  ginverse =  list(Phylo = inv.phylo$Ainv),
```
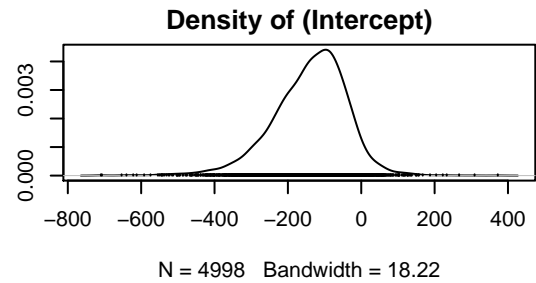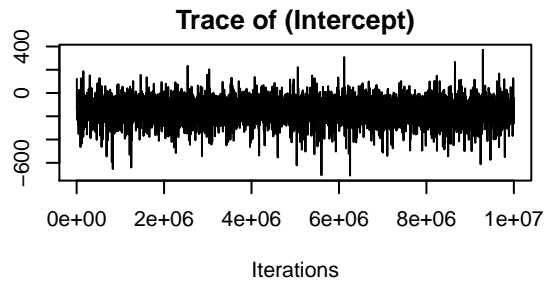
```
                    nitt=10000000,burnin=5000,thin=2000,
                    verbose = FALSE)
```

Compare posterior mode and mean for the three different runs (mean and posterior mode of each model should be similar)

```
##           stat      mcmc.1      mcmc.2      mcmc.3
##           mean -140.60277 -138.34922 -138.84316
## var1 post.mode  -92.96532  -88.13285  -92.78974
```

Posterior modes and means are similar across runs

Gelman-Rubin statistics (the potential scale reduction factor, PSRF) (should be less than 1.1 among chains)

```
# To check for independent chain posterior convergence
model.all <- mcmc.list(mcmc1$Sol,mcmc2$Sol,mcmc3$Sol)
gelman.diag(model.all, confidence = 0.95, transform=FALSE, autoburnin=TRUE)
```

```
## Potential scale reduction factors:
##
##             Point est. Upper C.I.
## (Intercept)          1          1
## Season               1          1
## Inc                  1          1
##
## Multivariate psrf
##
## 1
```

We can see that these are very close to 1, suggesting good convergence.

Gelman plots

```
par(mar=c(3,5,3,2))
gelman.plot(model.all)
```

(Intercept)

Season

Inc
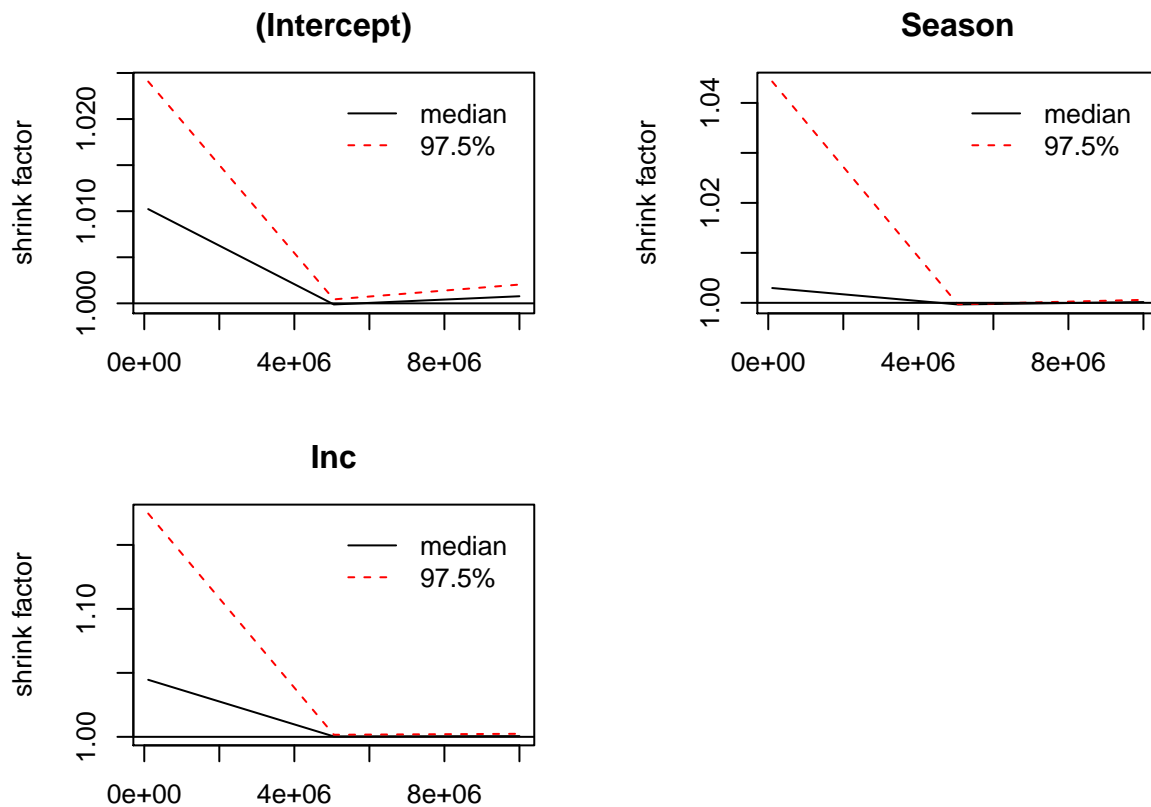
This is also evident when looking at the plot of the values generation per generation as the mixing is very good.

4 Results

Summary of the three models

```
summary(mcmc1)
```

```
##
##  Iterations = 5001:9999001
##  Thinning interval  = 2000
##  Sample size  = 4998
##
##  DIC: 1.601937
##
##  G-structure:  ~Phylo
##
##        post.mean  l-95% CI u-95% CI eff.samp
##  Phylo      5459 0.0002345    28567     4998
##
##                 ~Species
##
##        post.mean  l-95% CI u-95% CI eff.samp
```

14

```
## Species        2371 0.0002795    13773      4998
##
##  R-structure:  ~units
##
##        post.mean l-95% CI u-95% CI eff.samp
## units       39402     1126     85340     4998
##
##  Location effects: SexDiff ~ Season + Inc
##
##              post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept)   -140.60  -352.66     43.29     4998 0.0988 .
## Season         335.82   105.33    586.85     4998 <2e-04 ***
## Inc           -218.80  -389.39    -42.05     5205 0.0028 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

`summary`(mcmc2)

```
##
##  Iterations = 5001:9999001
##  Thinning interval  = 2000
##  Sample size  = 4998
##
##  DIC: 1.609266
##
##  G-structure:  ~Phylo
##
##        post.mean  l-95% CI u-95% CI eff.samp
## Phylo      5768 0.0002523    31016      4998
##
##                ~Species
##
##          post.mean l-95% CI u-95% CI eff.samp
## Species       2123 0.000192    13048      4998
##
##  R-structure:  ~units
##
##        post.mean l-95% CI u-95% CI eff.samp
## units       39866     1766     86303     4998
##
##  Location effects: SexDiff ~ Season + Inc
##
##              post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept)   -138.35  -342.09     47.36     4998 0.1000
## Season         336.65    94.21    568.53     4998 <2e-04 ***
## Inc           -222.35  -403.85    -55.54     4998 0.0024 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

`summary`(mcmc3)

```
##
##  Iterations = 5001:9999001
```

```
##  Thinning interval  = 2000
##  Sample size  = 4998
##
##  DIC: 1.624105
##
##  G-structure:  ~Phylo
##
##       post.mean l-95% CI u-95% CI eff.samp
## Phylo      5494 0.000274    30882     4998
##
##               ~Species
##
##         post.mean  l-95% CI u-95% CI eff.samp
## Species      2097 0.0002758    12791     4998
##
##  R-structure:  ~units
##
##       post.mean l-95% CI u-95% CI eff.samp
## units     39834    766.3    85089     4998
##
##  Location effects: SexDiff ~ Season + Inc
##
##            post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept)   -138.84  -339.85    43.40     4998 0.1004
## Season         334.78    98.59   582.19     5209 <2e-04 ***
## Inc           -220.65  -398.57   -50.31     4998 0.0024 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The effective sample sizes are all relatively good (higher than the value recommended by the function raftery.diag), both for random and fixed effects.
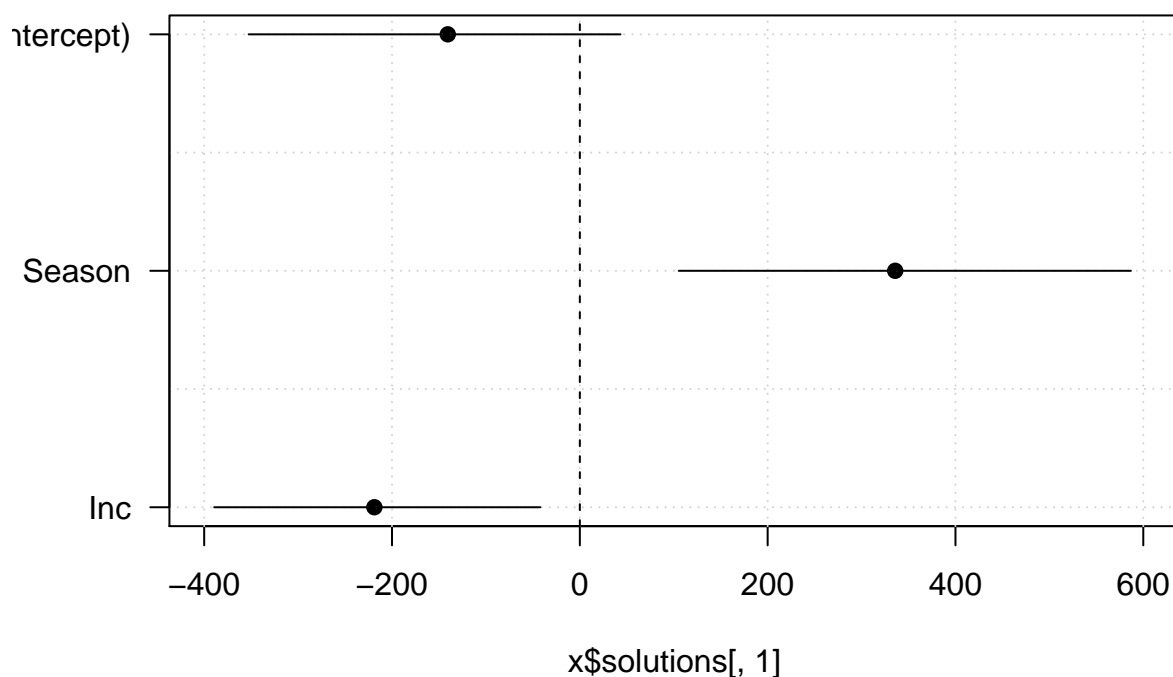
The three runs give the same results but values differ slightly. The second and third models were run only for verification purposes (Gelman-Rubin statistics). We decide to report only the results of the first model (i.e. mcmc1).

4.1   Fixed effects

Plot the parameter estimates and 95% CIs

```
plot.estimates(mcmc1)
```

Posterior means and 95% credible intervals

x$solutions[, 1]

4.2 Random effects

See below.

5 Random effects

We included phylogeny and species as random effects. Let's see if they have an effect on the probability to dtect seasonal differences in preen oil composition.

5.1 Posterior distribution

Plot the posterior distribution as a histogram to check for significance and whether it has been well estimated or not.

```
par(mfrow = c(1,2),mar = c(6, 6, 2, 2)) # plot panel
hist(mcmc(mcmc1$VCV)[,"Phylo"]) # non-significant
hist(mcmc(mcmc1$VCV)[,"Species"]) # non-significant
```

Histogram of mcmc(mcmc1$VCV)[,** **istogram of mcmc(mcmc1$VCV)[, "
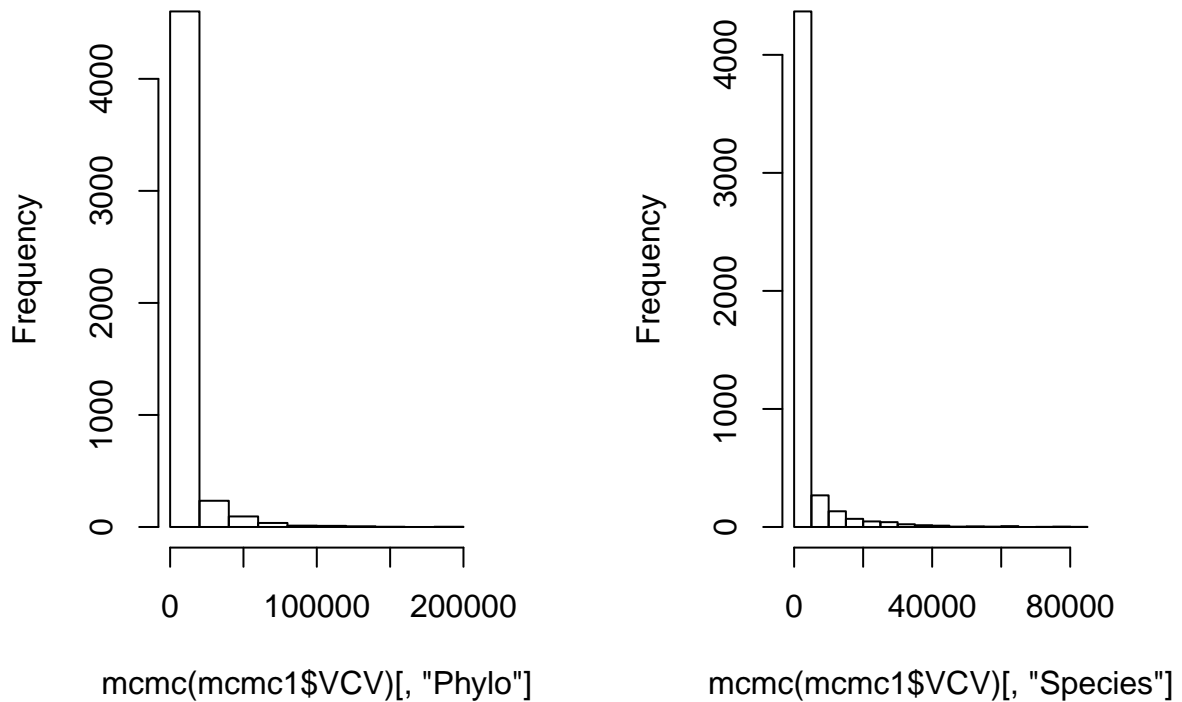
Posterior distribution is pushed up against 0 for both random effects (phylogenetic relatedness and species), which therefore seem to have little effect on seasonal variation in preen oil composition. Indeed, because variance cannot be zero, if the mean value is pushed up against zero, it reflects a non-significant effect. The large spread of the histogram indicate that the distribution is not well estimated (poor precision).

5.2 Proportion of the total variance explained by the random effects

To evaluate variance explained by the random effects, it is useful to look at the relative proportion of the variance explained by each effect and by the residuals (units)

```
# Proportion of variance explained by random factors
rand <- mcmc1$VCV/apply(mcmc1$VCV,1,sum)
# Get median values (50%) and 95% quantiles
apply(rand,2,function(c) quantile(c,probs = c(0.025,0.5,0.975)))
```

```
##               Phylo        Species      units
## 2.5%   4.900498e-08  3.351996e-08  0.3391211
## 50%    2.402639e-03  2.948618e-04  0.9582726
## 97.5%  5.885787e-01  4.682527e-01  0.9999986
```

```
# Get the mean value
apply(rand,2,mean)
```

```
##      Phylo     Species       units
## 0.09254429  0.05162144  0.85583428
```

The large confidence intervals confirm the poor precision of the estimates.

5.3 Inclusion of random effects and fit of the model

Run the different models (phylo only, species only, no random effect) to check if their inclusion in the model improves its fit. Note that a specific prior is built for each model.

Model phylo only

```
prior1 <- list(R=list(V=1,nu=0.002),G=list(G1=list(V=1,nu=0.002)))

mcmc.phylo <- MCMCglmm(SexDiff ~ Season + Inc,
                  random = ~Phylo,
                  family = "categorical",
                  data = df,
                  prior = prior1,
                  ginverse =  list(Phylo = inv.phylo$Ainv),
                  nitt=10000000,burnin=5000,thin=2000,
                  verbose = FALSE)

summary(mcmc.phylo)
```

```
##
##  Iterations = 5001:9999001
##  Thinning interval  = 2000
##  Sample size  = 4998
##
##  DIC: 1.631763
##
##  G-structure:  ~Phylo
##
##        post.mean  l-95% CI u-95% CI eff.samp
## Phylo      6271 0.0001926    33416     4743
##
##  R-structure:  ~units
##
##        post.mean l-95% CI u-95% CI eff.samp
## units     41259      181    87201     4998
##
##  Location effects: SexDiff ~ Season + Inc
##
##              post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept)   -142.68  -359.16    31.17     4998 0.0928 .
## Season         337.47    99.31   580.67     4998 <2e-04 ***
## Inc           -218.47  -389.88   -48.95     5215 0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model species only

```
mcmc.species <- MCMCglmm(SexDiff ~ Season + Inc,
                  random = ~Species,
                  family = "categorical",
```

```
                data = df,
                prior = prior1,
                ginverse =  list(Phylo = inv.phylo$Ainv),
                nitt=10000000,burnin=5000,thin=2000,
                verbose = FALSE)
summary(mcmc.species)
```

```
##
##  Iterations = 5001:9999001
##  Thinning interval  = 2000
##  Sample size  = 4998
##
##  DIC: 1.635616
##
##  G-structure:  ~Species
##
##          post.mean  l-95% CI u-95% CI eff.samp
## Species       2335 0.0002644     14892      4904
##
##  R-structure:  ~units
##
##        post.mean l-95% CI u-95% CI eff.samp
## units      40813    873.7     88373     5328
##
##  Location effects: SexDiff ~ Season + Inc
##
##              post.mean l-95% CI u-95% CI eff.samp   pMCMC
## (Intercept)    -143.09  -323.88     12.13     4998 0.0452 *
## Season          335.81    87.39    567.57     4998 <2e-04 ***
## Inc            -226.16  -397.54    -50.45     4998 0.0004 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model no random effect

```
prior0 <- list(R=list(V=1,nu=0.002))

mcmc.norandom <- MCMCglmm(SexDiff ~ Season + Inc,
                family = "categorical",
                data = df,
                prior = prior0,
                ginverse =  list(Phylo = inv.phylo$Ainv),
                nitt=10000000,burnin=5000,thin=2000,
                verbose = FALSE)

summary(mcmc.norandom)
```

```
##
##  Iterations = 5001:9999001
##  Thinning interval  = 2000
##  Sample size  = 4998
##
```

```
##  DIC: 1.67544
##
##  R-structure:  ~units
##
##       post.mean l-95% CI u-95% CI eff.samp
## units      42180    72.01    89066     4881
##
##  Location effects: SexDiff ~ Season + Inc
##
##             post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept) -145.236 -333.353    5.462     4998  0.034 *
## Season       336.729   97.852  583.886     4998 <2e-04 ***
## Inc         -222.473 -395.694  -56.975     5226 <2e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare the different models using the Deviance Information Criterion (DIC; lower values are best).

```
# Compare fit of the models (DIC)
data.frame(models=c("mcmc.norandom","mcmc.phylo","mcmc.species","mcmc1"),
           random.effects=c("NA","phylo","species","phylo+species"),
           DIC=c(mcmc.norandom$DIC,mcmc.phylo$DIC,mcmc.species$DIC,mcmc1$DIC))
```

```
##           models random.effects      DIC
## 1 mcmc.norandom             NA 1.675440
## 2    mcmc.phylo          phylo 1.631763
## 3  mcmc.species        species 1.635616
## 4         mcmc1  phylo+species 1.601937
```

Inclusion of both phylogeny and species increases the fit of the model (lower DIC).