

Phylogenetic comparative analyses

```
# phylogenetic comparative analysis
library(ggtree)
library(ape)
library(phytools)
library(dplyr)
library(readxl)
library(stringr)
library(ggthemr)
library(reshape2)
library(scales)
library(forcats)
library(readr)

# for comparative analysis
library(caper)
library(yhat)
library(dplyr)
library(GGally)
```

Load the data and do some processing.

```
# phylogeny
tree_final <- read.tree("data/raw/phylogeny/higdon_mod2_28.tre")

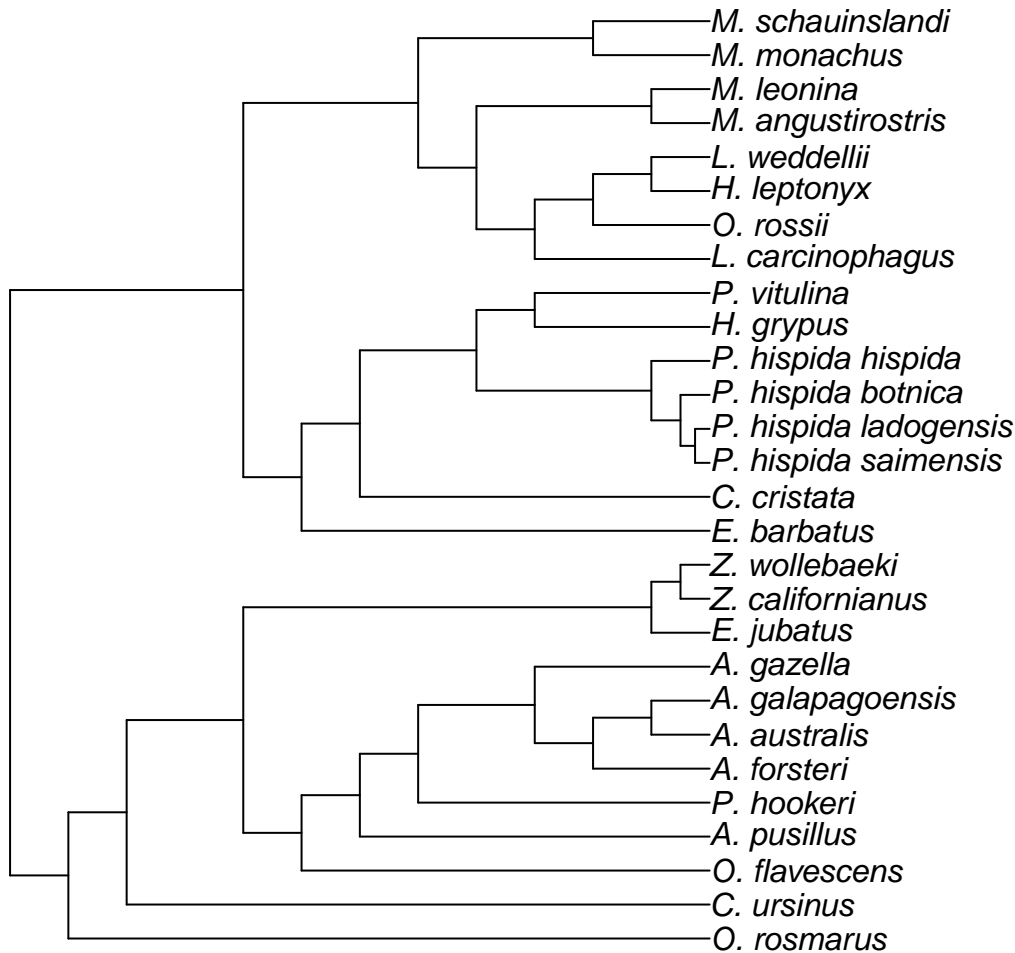
# produce short names for plotting
short <- c("W", "NFS", "SSL", "CSL", "GSL", "SASL", "AFS", "NZSL", "AntFS", "NZFS", "SAFS", "GFS",
  "BS", "HoS", "GS", "HS", "ARS", "SRS", "BRS", "LRS", "MMS", "HMS", "NES", "SES", "CS", "RS", "LS", "
# all_stats$short <- short

# all other seal stats and demographic variables
all_stats <- read_csv("data/processed/all_stats_tree.csv") %>%
  mutate(short = short) %>%
  mutate(SSD = male_weight/female_weight) %>%
  mutate(abc_out = ifelse(bot > 0.5, "bot", "neut")) %>%
  mutate(BreedingType = factor(BreedingType, levels = c("ice", "land", "both"))) %>%
  mutate(logAbundance = log(Abundance),
    logharem_size = log(harem_size),
    logmale_weight = log(male_weight)) %>%
  # order factors according to tree
  mutate(tip_label = fct_inorder(factor(tip_label)),
    species = fct_inorder(factor(species)),
    latin = fct_inorder(factor(latin)),
    common = fct_inorder(factor(common)),
    short = fct_inorder(factor(short)))

# count grey and harbour seal to land breeding
all_stats[all_stats$BreedingType == "both", "BreedingType"] <- "land"
all_stats <- all_stats %>% mutate(BreedingType = as.factor(as.character(BreedingType))) %>% data.frame()
```

Here is the phlogeny.

```
plot(tree_final)
```



Select variables for modeling.

```
stats_mod <- all_stats %>%
  dplyr::select(TPM70_ratio, TPM90_ratio, num_alleles_mean, harem_size, SSD, BreedingType,
    male_weight, breeding_season_length, lactation_length, life_span_years,
    Abundance, Generation_time, Longevity, tip_label, mratio_mean,
    obs_het_mean, mean_allele_range, IUCN_rating, prop_low_afs_mean,
    nloc, nind, bot, logAbundance) %>%
  data.frame()
```

Phylogenetic mixed modeling: Inverse phylo matrix and define priors.

```
library(MCMCglmm)
inv_phylo <- inverseA(tree_final, nodes="TIPS")$Ainv #,scale=TRUE
prior<-list(G=list(G1=list(V=1,nu=0.002)),R=list(V=1,nu=0.002))
```

R2 for mixed models (marginal)

```
# function to calculate the R2 for gaussian MCMCglmm models
mcmcR2 <- function(mod, type = "marginal", family = "gaussian"){
  if (type != "marginal") stop("At the moment, there is just the marginal R2")
  if (family != "gaussian") stop("At the moment just gaussian errors are supported")
  # based on Shinichis answer on Researchgate
  mVarF <- var(as.vector(apply(mod$Sol,2,mean) %*% t(mod$X)))
  R2 <- mVarF/(mVarF+sum(apply(mod$VCV,2,mean)))

  # alternative with crebile intervals
  mcmc_chain_length <- nrow(mod$VCV)
  vmVarF<-numeric(mcmc_chain_length)

  vmVarF <- vapply(1:mcmc_chain_length, function(x) out <- var(as.vector(mod$Sol[x,] %*% t(mod$X))),
    FUN.VALUE = numeric(length(mcmc_chain_length)))

  R2m<-vmVarF/(vmVarF+ rowSums(mod$VCV)) # include here all random effects plus errors
  outR2m <- R2m
  class(R2m) <- "mcmc"
  R2m<-vmVarF/(vmVarF+mod$VCV[,1]+mod$VCV[,2])

  #mean(R2m)
  #posterior.mode(R2m)
  #data.frame(HPDinterval(R2m), row.names = NULL)

  out <- list("partR2" = data.frame("meanR2" = mean(R2m), "modeR2" = posterior.mode(R2m), data.frame(
    "R2_chain" = outR2m)
  )
}
```

Hypothesis (1) How is heterozygosity-excess as a bottleneck signature correlated to typical diversity indices?

First, standardize input variables to make the estimates comparable.

```
stats_mod_div <- stats_mod %>%
  mutate(num_alleles_mean = scale(num_alleles_mean),
         obs_het_mean = scale(obs_het_mean),
         prop_low_afs_mean = scale(prop_low_afs_mean))

mod_gen <- MCMCglmm(TPM70_ratio ~ num_alleles_mean + obs_het_mean + prop_low_afs_mean, #
  random=~tip_label, nodes = "TIPS", # rcov =~us(trait):units
  family=c("gaussian"),ginverse=list(tip_label=inv_phylo),prior=prior,
  data=stats_mod_div,nitt=110000,burnin=10000,thin=50, verbose = FALSE)

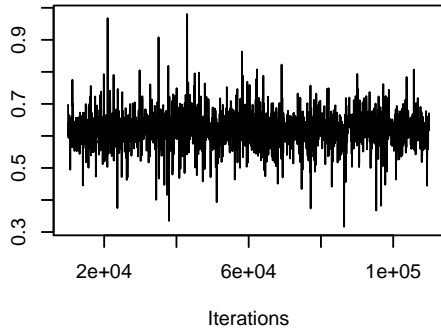
summary(mod_gen)

##
## Iterations = 10001:109951
## Thinning interval = 50
## Sample size = 2000
##
## DIC: -32.28948
```

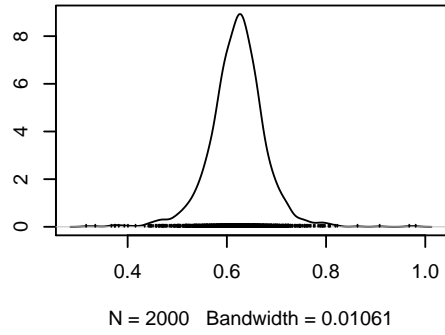
```
##
## G-structure: ~tip_label
##
##           post.mean  l-95% CI u-95% CI eff.samp
## tip_label  0.01161 0.0001672 0.04926      1356
##
## R-structure: ~units
##
##           post.mean  l-95% CI u-95% CI eff.samp
## units      0.01665 0.0004163 0.03029      1610
##
## Location effects: TPM70_ratio ~ num_alleles_mean + obs_het_mean + prop_low_afs_mean
##
##           post.mean  l-95% CI  u-95% CI eff.samp  pMCMC
## (Intercept)    0.621133  0.509843  0.730081     2000 <5e-04 ***
## num_alleles_mean -0.007179 -0.191186  0.181028     2000  0.929
## obs_het_mean     0.138071 -0.021714  0.284422     2000  0.081 .
## prop_low_afs_mean -0.194309 -0.296159 -0.097864     2000 <5e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(mod_gen$Sol)
```

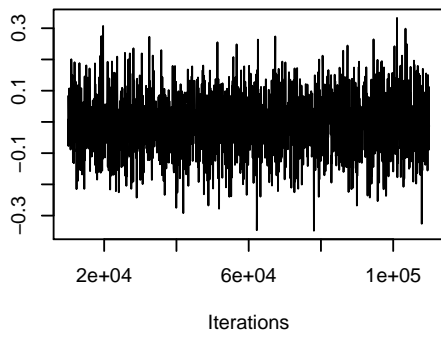
Trace of (Intercept)



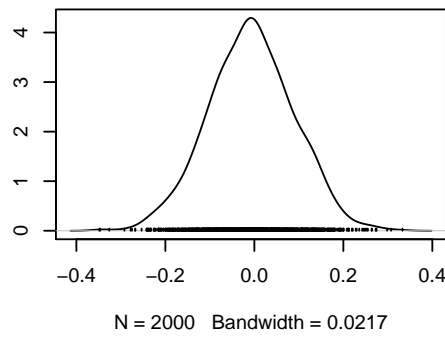
Density of (Intercept)



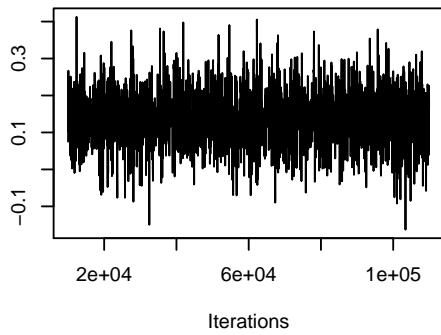
Trace of num_alleles_mean



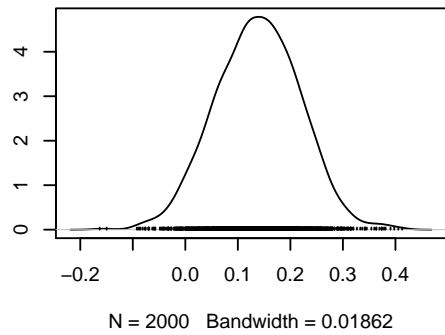
Density of num_alleles_mean



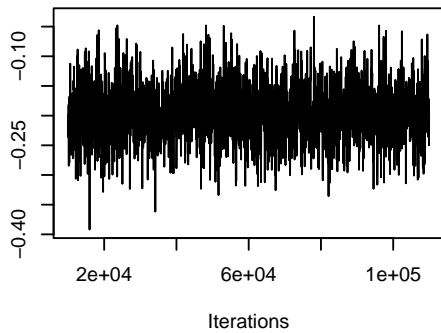
Trace of obs_het_mean



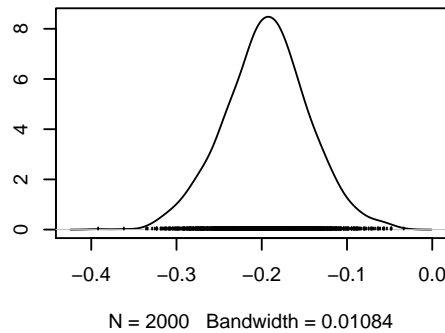
Density of obs_het_mean



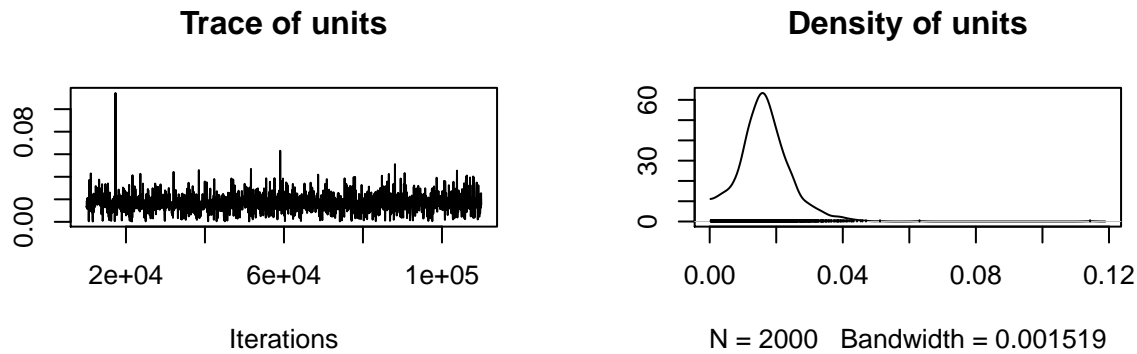
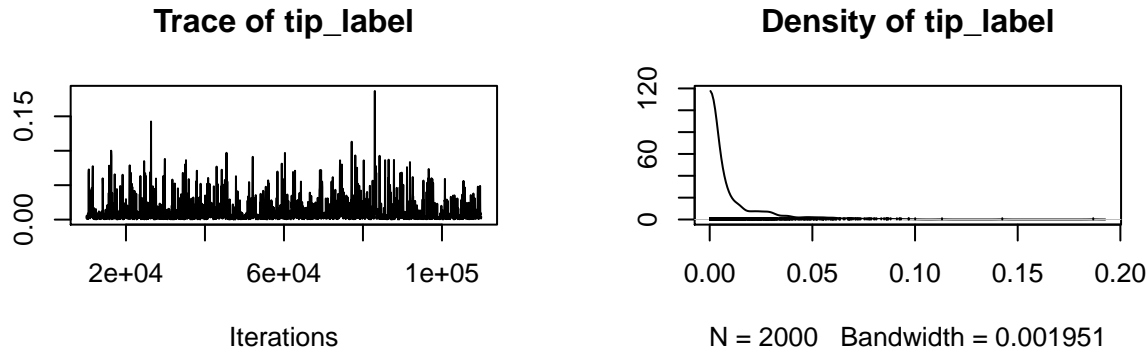
Trace of prop_low_afs_mean



Density of prop_low_afs_mean



```
plot(mod_gen$VCV)
```



```
autocorr(mod_gen$Sol)
```

```
## , , (Intercept)
##
##      (Intercept) num_alleles_mean obs_het_mean prop_low_afs_mean
## Lag 0      1.000000000 -0.007216183  0.0106983038   0.006029174
## Lag 50      0.019811765 -0.006402469  0.0010600891   0.026867955
## Lag 250     0.054086536 -0.008128206 -0.0074720865   0.037468684
## Lag 500     0.026388809  0.003668325 -0.0000135441   0.007716804
## Lag 2500    -0.002929626  0.023854252  0.0035003218   -0.065229803
##
## , , num_alleles_mean
##
##      (Intercept) num_alleles_mean obs_het_mean prop_low_afs_mean
## Lag 0      -0.007216183   1.000000000 -0.866280090   -0.581069231
## Lag 50      0.013058377   0.011592414  0.004163592   -0.021741964
## Lag 250     -0.065132497   0.023012934 -0.007038241   -0.002102486
## Lag 500     -0.017003306  -0.017698531  0.024624161   -0.002078504
## Lag 2500    -0.013765684  -0.008658784  0.003548198    0.014124293
##
## , , obs_het_mean
##
##      (Intercept) num_alleles_mean obs_het_mean prop_low_afs_mean
## Lag 0      0.0106983038  -0.866280090  1.00000000   0.207528634
## Lag 50     -0.0197349358  -0.004769702 -0.02227582   0.040147026
## Lag 250     0.0439929245  -0.017434495  0.01176800   -0.009695746
## Lag 500     0.0083582950   0.005683895 -0.01565171   0.009609889
```

```
## Lag 2500 -0.0001471254      0.020842269 -0.01436467      -0.022253422
##
## , , prop_low_afs_mean
##
##      (Intercept) num_alleles_mean obs_het_mean prop_low_afs_mean
## Lag 0      0.006029174      -0.58106923   0.20752863      1.000000000
## Lag 50      0.018289926      -0.01972430   0.01322466      0.005827153
## Lag 250     0.058359076      -0.00443747  -0.01312518      0.008395246
## Lag 500     0.023275363       0.03493281  -0.02986688      -0.018749257
## Lag 2500    0.035651201       0.01296201  -0.01313968      -0.004565001
```

```
autocorr(mod_gen$VCV)
```

```
## , , tip_label
##
##      tip_label      units
## Lag 0      1.000000000 -0.516534691
## Lag 50      0.148774324 -0.137725656
## Lag 250     0.030833794 -0.014774451
## Lag 500    -0.009459346  0.016052417
## Lag 2500   0.015466380 -0.006465327
##
## , , units
##
##      tip_label      units
## Lag 0     -0.516534691  1.000000000
## Lag 50    -0.110993257  0.10782185
## Lag 250   -0.010278981  0.01216979
## Lag 500    0.009094770 -0.01489854
## Lag 2500  -0.002118572  0.03047730
```

R2

```
R2 <- mcmcR2(mod_gen)
R2$partR2
```

```
##      meanR2  modeR2  lower  upper
## 1 0.448175 0.540029 0.17742 0.6891671
```

Hypothesis (2) Which life-history or demographic traits explain genetic diversity?

Standardize by 2 sd to make estimates comparable with binary BreedingHabitat variable (Gelman (2008), Schielzeth (2014))

```
stats_mod_div <-
  stats_mod %>%
  mutate(Abundance = ((Abundance - mean(Abundance)) / (2*sd(Abundance))),
         Generation_time = (Generation_time - mean(Generation_time) / (2*sd(Generation_time))),
         SSD = (SSD - mean(SSD) / (2*sd(SSD))),
         logAbundance = ((logAbundance - mean(logAbundance)) / (2*sd(logAbundance))))
```

Running the model.

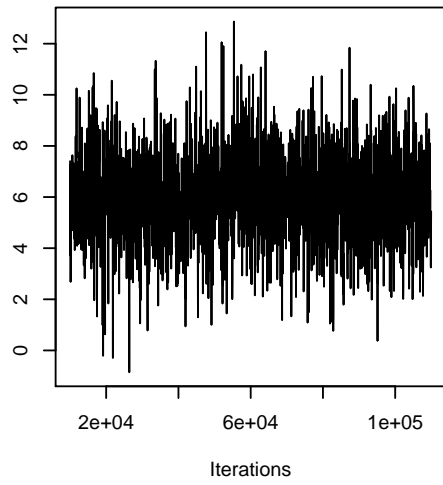
```
mod_div <- MCMCglmm(num_alleles_mean ~ logAbundance + BreedingType + Generation_time + SSD, # , #+ Abun
  random=~tip_label, nodes = "TIPS", # rcov=~us(trait):units
  family=c("gaussian"),ginverse=list(tip_label=inv_phylo),prior=prior,
  data=stats_mod_div,nitt=110000,burnin=10000,thin=50, verbose = FALSE)
```

```
summary(mod_div)
```

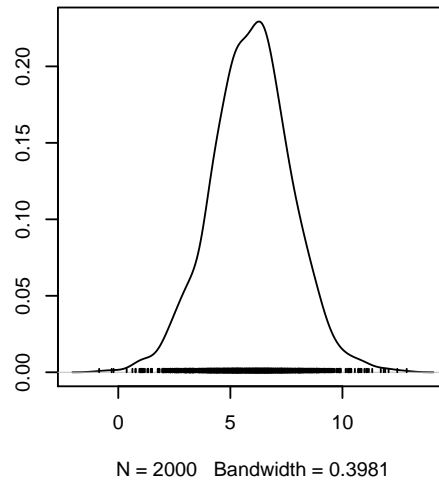
```
##
## Iterations = 10001:109951
## Thinning interval = 50
## Sample size = 2000
##
## DIC: 102.0803
##
## G-structure: ~tip_label
##
##          post.mean  l-95% CI u-95% CI eff.samp
## tip_label    0.4797 0.0002164    2.176    1235
##
## R-structure: ~units
##
##          post.mean l-95% CI u-95% CI eff.samp
## units        1.846   0.6118    3.202    1636
##
## Location effects: num_alleles_mean ~ logAbundance + BreedingType + Generation_time + SSD
##
##          post.mean    l-95% CI    u-95% CI eff.samp pMCMC
## (Intercept)    5.943e+00  2.320e+00  9.360e+00    2000 0.003 **
## logAbundance    2.212e+00  1.033e+00  3.521e+00    2000 0.001 ***
## BreedingTypepeland -1.774e+00 -3.461e+00 -1.638e-01    1418 0.030 *
## Generation_time    9.662e-05 -5.397e-04  6.634e-04    2000 0.728
## SSD              1.190e-01 -3.932e-01  6.053e-01    2000 0.633
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(mod_div$Sol)
```

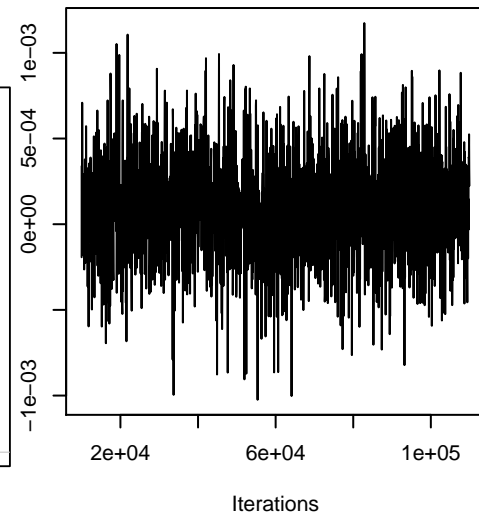

Trace of (Intercept)



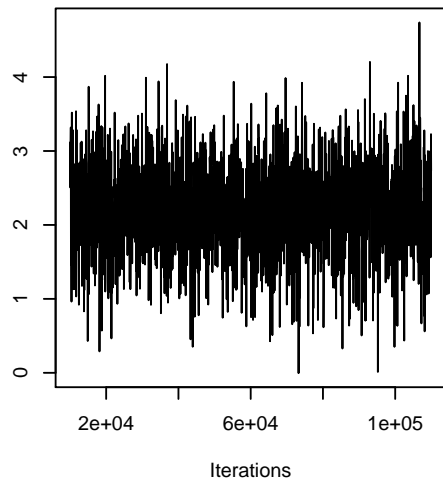
Density of (Intercept)



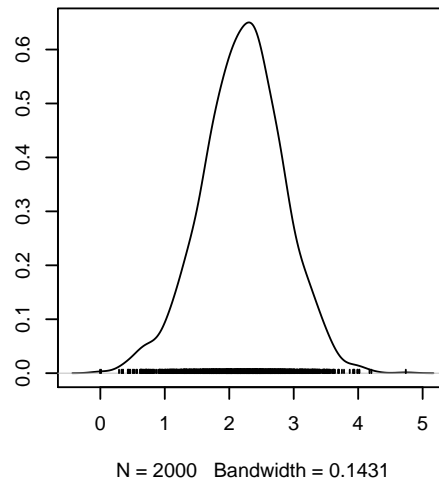
Trace of Generation_time



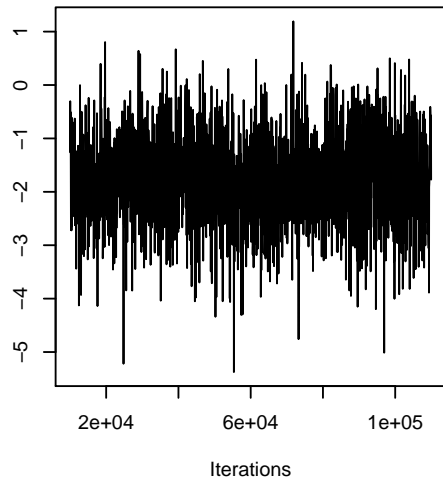
Trace of logAbundance



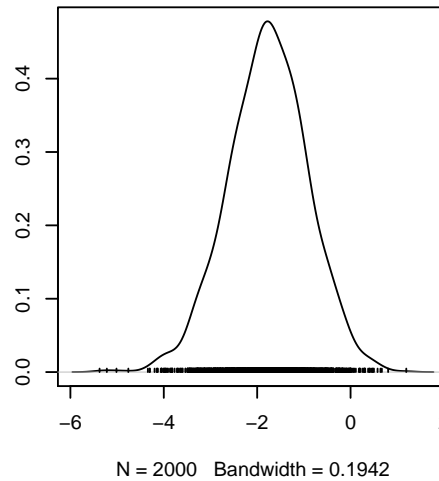
Density of logAbundance



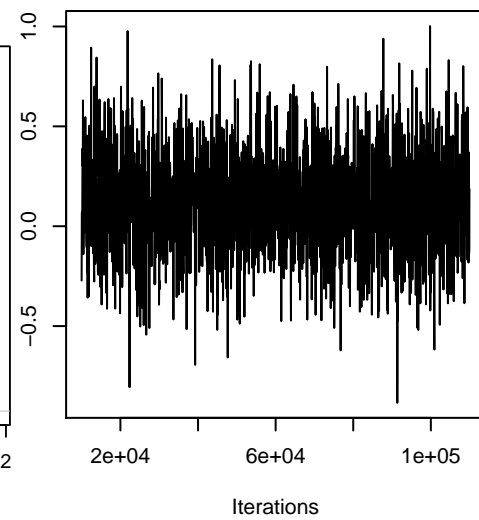
Trace of BreedingTypeland



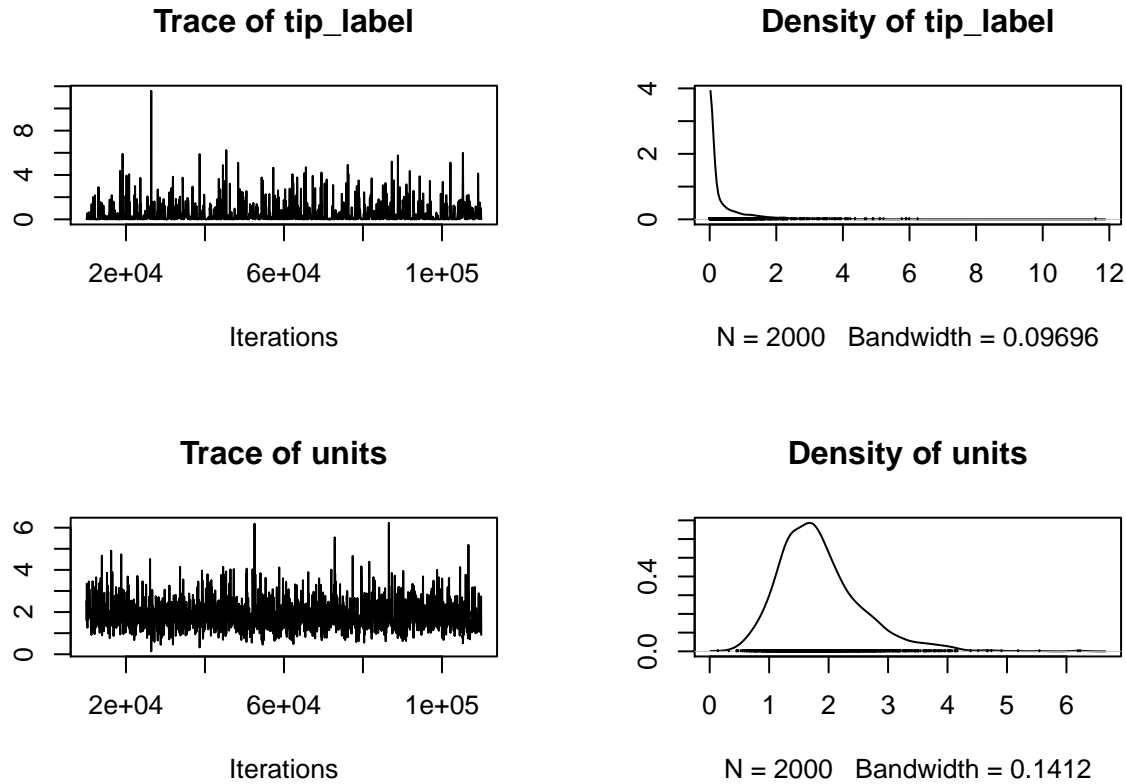
Density of BreedingTypeland



Trace of SSD



```
plot(mod_div$VCV)
```



```
autocorr(mod_div$Sol)
```

```
## , , (Intercept)
##
##      (Intercept) logAbundance BreedingTypeland Generation_time
## Lag 0      1.000000000 -0.089158222    -0.5026076953   -0.9454294806
## Lag 50      0.006566086 -0.026957322     0.0424859716   -0.0045718142
## Lag 250     -0.017005587 -0.009107329     0.0090833638    0.0072204484
## Lag 500      0.003704807  0.006267896    -0.0006367604   -0.0004585829
## Lag 2500     0.001054940 -0.031027487    -0.0009974573   -0.0135600095
##
##      SSD
## Lag 0      -0.230423594
## Lag 50      -0.018348267
## Lag 250     -0.003626477
## Lag 500      0.017852744
## Lag 2500     0.011918177
##
## , , logAbundance
##
##      (Intercept) logAbundance BreedingTypeland Generation_time
## Lag 0      -0.089158222  1.000000000     0.310946787    0.105744262
## Lag 50      0.056440729 -0.030612285    -0.051114617   -0.064436489
## Lag 250     -0.023584694  0.006035796     0.016996770    0.046109076
## Lag 500     -0.030925323 -0.004554234     0.028550151    0.032200098
## Lag 2500     0.002263715  0.025552139     0.006464162    0.003160304
##
##      SSD
## Lag 0      -0.412973445
```

```

## Lag 50    0.039174655
## Lag 250  -0.019832652
## Lag 500  -0.013996654
## Lag 2500 0.004724514
##
## , , BreedingTypeland
##
##      (Intercept) logAbundance BreedingTypeland Generation_time
## Lag 0    -0.50260770  0.310946787      1.00000000      0.406721032
## Lag 50     0.02524953 -0.005282061     -0.02984347     -0.028560717
## Lag 250    -0.02637593  0.023664383      0.01744997      0.032002057
## Lag 500    -0.01277296  0.013750272      0.01275751      0.009903714
## Lag 2500   0.03089908  0.026295638     -0.01765329     -0.018133930
##      SSD
## Lag 0    -0.46130989
## Lag 50     0.03730128
## Lag 250    0.00191946
## Lag 500   -0.01649061
## Lag 2500  -0.02259842
##
## , , Generation_time
##
##      (Intercept) logAbundance BreedingTypeland Generation_time
## Lag 0    -0.9454294806  0.105744262      0.4067210321      1.000000000
## Lag 50    -0.0062780209  0.023203181     -0.0330979199      0.007579702
## Lag 250    0.0034527323  0.003589161     -0.0124874676      0.007125897
## Lag 500   -0.0002388019 -0.012285741      0.0002762982     -0.004154276
## Lag 2500  -0.0147869763  0.019331421      0.0076845522      0.027645650
##      SSD
## Lag 0     0.167195404
## Lag 50     0.003869608
## Lag 250    0.011200541
## Lag 500   -0.017085810
## Lag 2500  -0.013599481
##
## , , SSD
##
##      (Intercept) logAbundance BreedingTypeland Generation_time
## Lag 0    -0.23042359  -0.41297345     -0.461309894      0.16719540
## Lag 50    -0.03435781  0.03179945     -0.005941092      0.02947681
## Lag 250    0.03690658  -0.02215181     -0.02236239      -0.03977316
## Lag 500   -0.01210796  0.01442442     -0.006470692      0.01477277
## Lag 2500  -0.02805839  0.00597352      0.005185650      0.02154808
##      SSD
## Lag 0     1.000000000
## Lag 50     0.007741551
## Lag 250    0.029638182
## Lag 500   -0.005087331
## Lag 2500   0.011607454

```

```
autocorr(mod_div$VCV)
```

```

## , , tip_label
##
##      tip_label      units

```

```
## Lag 0      1.000000000 -0.303547067
## Lag 50     0.201625203 -0.102747313
## Lag 250    -0.001444543  0.002086793
## Lag 500     0.002522338 -0.036162115
## Lag 2500   -0.015098425  0.036687508
##
## , , units
##
##          tip_label      units
## Lag 0     -0.303547067  1.000000000
## Lag 50     -0.124356730  0.046945420
## Lag 250    -0.011629029 -0.005601897
## Lag 500     0.005217020 -0.040226478
## Lag 2500    0.003911496 -0.002434900
```

```
R2 <- mcmcR2(mod_div)
R2$partR2
```

```
##      meanR2      modeR2      lower      upper
## 1 0.5093314 0.5397348 0.2766534 0.6926358
```

Hypothesis (3) Which life-history or demographic traits explain heterozygosity-excess (bottleneck signature)?

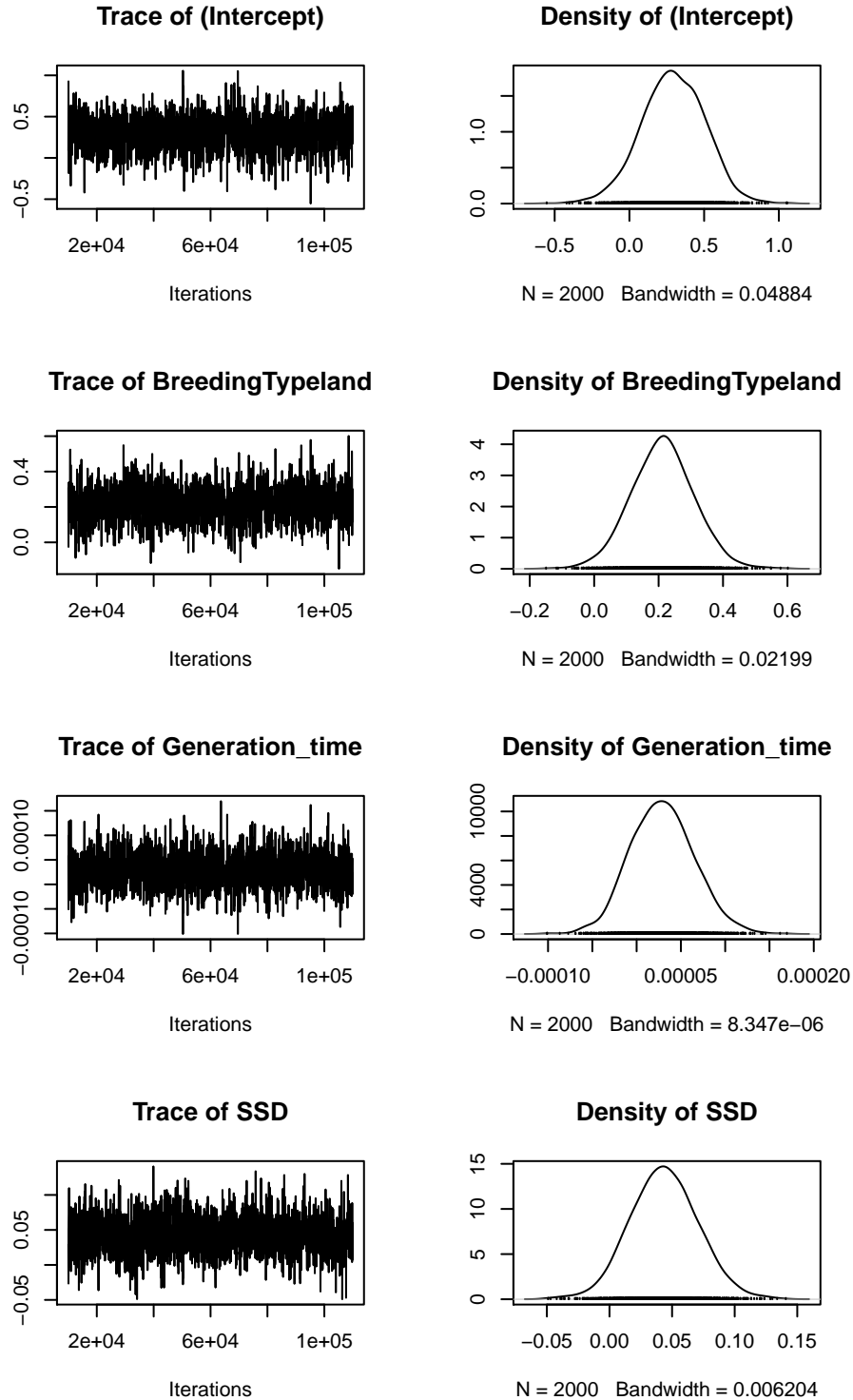
```
mod_bot <- MCMCglmm(TPM70_ratio ~ BreedingType + Generation_time + SSD, #
  random=~tip_label, nodes = "TIPS", # rcov=~us(trait):units
  family=c("gaussian"),ginverse=list(tip_label=inv_phylo),prior=prior,
  data=stats_mod_div,nitt=110000,burnin=10000,thin=50, verbose = FALSE)
```

```
summary(mod_bot)
```

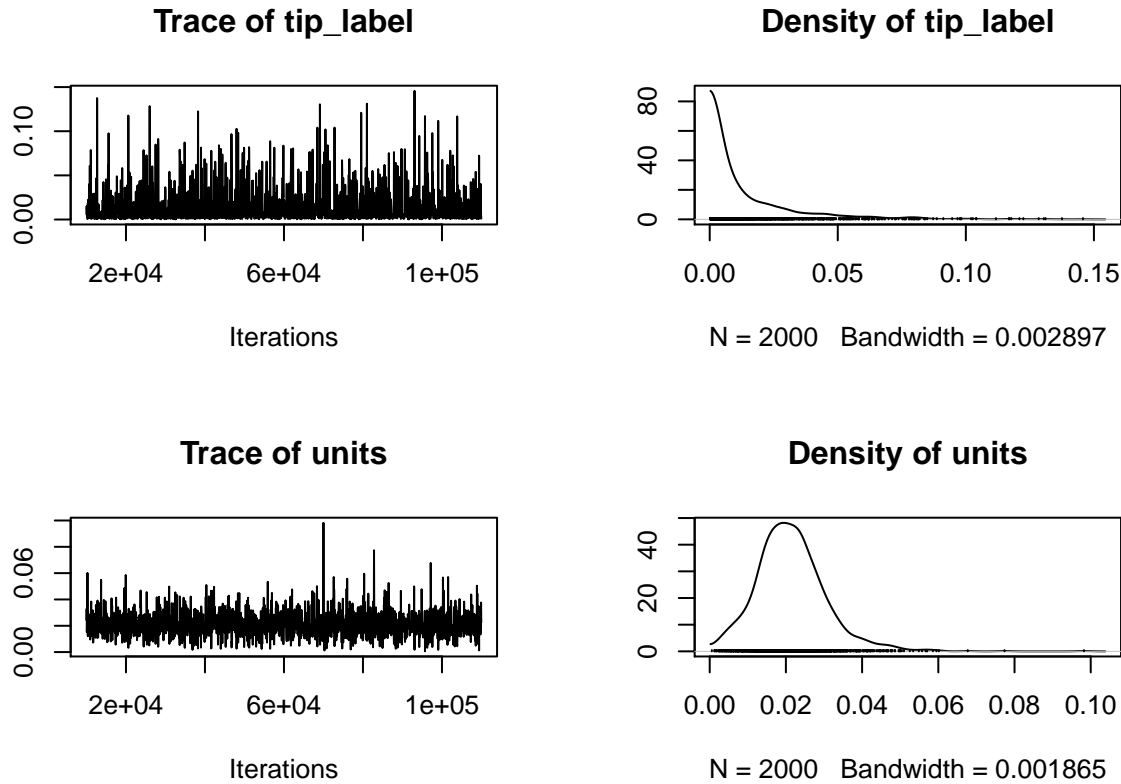
```
##
## Iterations = 10001:109951
## Thinning interval = 50
## Sample size = 2000
##
## DIC: -24.01734
##
## G-structure: ~tip_label
##
##          post.mean  l-95% CI u-95% CI eff.samp
## tip_label  0.01467 0.0001949 0.05578    1567
##
## R-structure: ~units
##
##          post.mean l-95% CI u-95% CI eff.samp
## units  0.02139 0.003805 0.03941    1845
##
## Location effects: TPM70_ratio ~ BreedingType + Generation_time + SSD
##
##          post.mean  l-95% CI  u-95% CI eff.samp pMCMC
## (Intercept)  2.915e-01 -1.187e-01  6.970e-01    2000 0.162
## BreedingType  2.124e-01  1.898e-02  4.004e-01    2000 0.032 *
```

```
## Generation_time  2.907e-05 -3.642e-05  1.038e-04    2000 0.433
## SSD             4.366e-02 -7.012e-03  9.912e-02    2000 0.097 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(mod_bot$Sol)
```



```
plot(mod_bot$VCV)
```



```
autocorr(mod_bot$Sol)
```

```
## , , (Intercept)
##
##      (Intercept) BreedingTypeland Generation_time      SSD
## Lag 0      1.000000000      -0.482322153      -0.921542128      -0.272436658
## Lag 50      0.007697499      -0.007207004      -0.004155664      -0.021295687
## Lag 250     0.024545147      -0.017603832      -0.037210393      0.005340046
## Lag 500     0.041865358      -0.020803152      -0.039785714      -0.007512185
## Lag 2500    -0.012962346      0.002152234      0.024253903      -0.017184784
##
## , , BreedingTypeland
##
##      (Intercept) BreedingTypeland Generation_time      SSD
## Lag 0      -0.48232215      1.000000000      0.35749659      -0.38567453
## Lag 50      0.02235272      -0.019612251      -0.01920396      0.02444578
## Lag 250     -0.02785504      0.008871643      0.02261031      0.03954827
## Lag 500     -0.01878100      0.027685318      0.02488179      -0.00759166
## Lag 2500     0.02794233      -0.023576014      -0.02030574      -0.01118340
##
## , , Generation_time
##
##      (Intercept) BreedingTypeland Generation_time      SSD
## Lag 0      -0.92154213      0.35749659      1.000000000      0.21152935
## Lag 50      -0.01446548      0.00681156      0.005053822      0.02053067
## Lag 250     -0.01868965      0.02017989      0.034709105      -0.03188938
## Lag 500     -0.03880242      0.01804476      0.037334062      0.01088474
```

```
## Lag 2500 0.01590575      -0.01234888      -0.025608820 0.02912999
##
## , , SSD
##
##      (Intercept) BreedingTypeland Generation_time      SSD
## Lag 0      -0.272436658      -0.385674533      0.211529348 1.000000000
## Lag 50      -0.006134405      -0.008739523      0.018012704 0.002220987
## Lag 250     -0.019191405      0.014942030      0.033617680 -0.016179615
## Lag 500     -0.027747841      -0.018982544      0.019197947 0.020414588
## Lag 2500    -0.031873223      0.040538727      0.005638393 0.005137925
```

```
autocorr(mod_bot$VCV)
```

```
## , , tip_label
##
##      tip_label      units
## Lag 0      1.0000000000 -0.462640212
## Lag 50      0.1212641717 -0.081247958
## Lag 250     -0.0009616816 0.029549508
## Lag 500     -0.0165188033 0.002688724
## Lag 2500    -0.0120662041 0.039905862
##
## , , units
##
##      tip_label      units
## Lag 0      -0.46264021 1.000000000
## Lag 50      -0.04723052 0.040063368
## Lag 250      0.01363505 -0.013776547
## Lag 500      0.01771435 -0.004124196
## Lag 2500     0.02537694 -0.026325137
```

```
R2 <- mcmcR2(mod_bot)
R2$partR2
```

```
##      meanR2      modeR2      lower      upper
## 1 0.3863231 0.4591928 0.1378802 0.6378461
```