

Simulate phylogenetic trees and calculate phylometrics.

Supplementary material in: Species and functional
diversity accumulate differently in terrestrial mammals

Brunno F. Oliveira^{1,2}; Antonin Machac^{3,4,5}; Gabriel C. Costa¹; Thomas M. Brooks⁶; Ana D. Davidson⁷; Carlo Rondinini⁸; Catherine H. Graham²;

Global Ecology and Biogeography, 2016 / **1** Departamento de Ecologia, Universidade Federal do Rio Grande do Norte, Natal, Brasil. / **2** Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York 11794, USA. / **3** Center for Macroecology, Evolution, and Climate, Natural History Museum of Denmark, Universitetsparken 15, DK 2100 Copenhagen, Denmark. / **4** Department of Ecology, Charles University, Vinicna 7, 128 44 Praha 2, Czech Republic. / **5** Center for Theoretical Study, Charles University and Academy of Sciences of the Czech Republic, Jilska 1, 110 00 Praha 1, Czech Republic. / **6** International Union for Conservation of Nature, 1196 Gland, Switzerland 7NatureServe, Arlington, VA, USA. / **8** Global Mammal Assessment Program, Department of Biology and Biotechnologies, Sapienza Universita di Roma, Rome, Italy. / **Corresponding author:** (brunno.oliveira@me.com)

Contents

1. Packages versions:	2
2. Load functions:	2
2. List of phylometrics:	2
4. Simulate phylogenetic trees:	3
4.1. Manipulate richness:	3
4.2. Manipulate time (age):	5
4.3. Manipulate diversification rate:	6
5. Table of results:	8

This document follows the principles of reproducible research (Peng, 2011). This document was generated in **R studio** with **kintr** package.

1. Packages versions:

We used R version 3.2.2 (2015-08-14) and the following packages:

```
library(geiger)
library(ape)
library(TreeSim)
library(picante)
library(reshape)
library(e1071)
library(phytools)
library(apTreeshape)
library(plyr)
library(TreeSim)
library(knitr)
```

2. Load functions:

```
### Function range data between zero and one
range01 <- function(x){(x-min(x))/(max(x)-min(x))}

### PSV function by Brunno Oliveira
# Modified from psu function in {picante}
psv.b<-function (samp, tree) {
  Cmatrix <- vcv.phylo(tree, corr = F)
  SR <- rowSums(data.frame(samp))
  nlocations <- 1
  nspecies <- length(tree$tip.label)
  index <- seq(1:nrow(Cmatrix))
  n <- length(index)
  C <- Cmatrix[index, index]
  PSV <- (n * sum(diag(as.matrix(C))) - sum(C))/(n * (n - 1))
  PSV
}

### MRD function by Brunno Oliveira
# Modified from ELIOT MILLER (avaible at http://www.umsl.edu/~emmq7/Menu/Rphylo/MRD.R)
MRD.b <- function(phy)
{
  phylo.bl1 <- compute.br1en(phy, 1)
  all.dist <- dist.nodes(phylo.bl1)
  root.dist <- all.dist[length(phy$tip.label)+1, 1:length(phy$tip.label)]
  tips.to.root <- data.frame(tipnames=phy$tip.label,root.dist)
  mrd <- mean(tips.to.root$root.dist)
  return(mrd)
}
```

2. List of phylometrics:

Phylogenetic diversity (PD) Sum of all branch lengths in a tree. Faith (1992)

Mean phylogenetic distance (MPD) Mean of phylogenetic distances between all species pairs within a tree. Webb (2000)

Mean nearest taxon distance (MNTD) Mean phylogenetic distance between each species and its nearest neighbor in a tree. Webb (2000)

Phylogenetic species variability (PSV) Expected variance among species in a neutrally evolving trait. Helmus et al. (2007)

Mean root distance (MRD) The mean number of nodes that separating species from the root of their tree. Alternatively, when time trees are available MRD can be calculated as the mean distance between each species nodes and the root of its phylogenetic tree (MRD.time). Kerr and Currie (1999); Hawkins et al. (2012)

Species evolutionary distinctiveness (ED) After dividing each branch length by the number of species subtending that branch, the obtained values are summed across all branches from which a species descends. Redding and Mooers (2006); Isaac et al. (2007)

Species-level diversification rate (DivRate) Inverse of the evolutionary distinctness (ED). Species rapidly diversifying will have short edge lengths shared among many species. Jetz et al. (2012), Kembel et al. (2010)

Relative branch length (RBL) Approximates time between speciation events across a tree by calculating the mean ((stem age – crown age) / crown age). Marin and Blair, In prep.

Diversification rates (DR) Species accumulation through time (log(clade diversity)/clade age). Magallon and Sanderson (2001); Nee (2001); Phillimore et al. (2006)

Gamma statistics (GAM) The distribution of branching events throughout the tree. This phylometric distinguished between trees with relatively long inter-nodal distances towards the tips (tippy trees) and trees with relatively longer inter-nodal distances towards the root of the phylogeny (stemmy trees). Pybus and Harvey (2000)

Colless Measures the degree of tree imbalance. Here, we calculated standardized values of Colless according to the Yule and PDA models. Colless (1982)

4. Simulate phylogenetic trees:

4.1. Manipulate richness:

```
# Phyvars
phyvars <- c("SPD", "PD", "age", "MRD", "MRD.time", "ED", "ED2", "RBL",
            "DR", "MPD", "MNTD", "PSV", "GAM", "IMY", "IMP", "DivRate", "DivRate2")

# We suppose that the greater amount of pruning in a phylogenetic tree greater the greater
# probability of the resulting tree be composed by species with deep evolutionary relationships.
# Conversely, lower levels of tree pruning is likely to comprise more complete clades'
# histories and concentrate nodes towards the tips (high div and young assemblages).
# The objective here is to find a phylometric that is not affected by the level of
# tree pruning (species richness).

# set parameters
N <- 50 # number of simulations
n <- rnorm(10000, mean=100, sd=38) # number of taxa (tips) * Parameters taken from the observed distribut
n = n+(-1*min(n))
n = n +10 # We consider that it is better to have at least 10 species to calculate phymetrics.
```

```

# Simulate a tree in parameters to the actual mammalian tree (Hedges et al., 2015)
trxs <- sim.bd.taxa.age(n=5000, numbsim=1, lambda=0.2, mu=0.14, age=180, mrca=T)
trxs <- trxs[[1]]

# get ED for species
spp.ED <- evol.distinct(trxs, type = c("fair.proportion"), ### Species' evolutionary distinctiveness
                        scale = TRUE, use.branch.lengths = TRUE)

# get age for each species (complete tree)
spp.ages.c<-data.frame(matrix(data=NA,nrow=length(trxs$tip.label),ncol=2))
colnames(spp.ages.c)<-c("Species","age")
max<-max(branching.times(trxs))
for (j in 1:length(trxs$tip.label)){
  Spp<-trxs$tip.label[j]
  spp.ages.c$age[j] <- trxs$edge.length[which.edge(trxs,Spp)]
  spp.ages.c$Species[j]<-Spp
}

# Create table for storing results
model1<-data.frame(matrix(data=NA,nrow=N,ncol=length(phyvars)+2))
colnames(model1)<-c("age","run",phyvars)

# Sample the tree to obtain assemblages with different richness
for(i in 1:N){

  #cat("\r",i,"of", N)

  k<- as.integer(sample(n,1,replace = T)) #n species

  trx<-drop.tip(trxs,sample(1:5000,5000-k)) # sample tips to obtain assemblages of n tips

  # get age for each species (pruned tree)
  spp.ages.p<-data.frame(matrix(data=NA,nrow=length(trx$tip.label),ncol=2))
  colnames(spp.ages.p)<-c("Species","age")
  max<-max(branching.times(trx))
  for (j in 1:length(trx$tip.label)){
    Spp<-trx$tip.label[j]
    spp.ages.p$age[j] <- trx$edge.length[which.edge(trx,Spp)]
    spp.ages.p$Species[j]<-Spp
  }

  subED<-subset(spp.ED,Species%in%trx$tip.label)
  subAge<-subset(spp.ages.c,Species%in%trx$tip.label)

  h<-data.frame(rep(1,length(trx$tip.label)))
  rownames(h)<-trx$tip.label
  h<-t(h)

  model1$run[i] <- i
  model1$SPD[i] <- length(trx$tip.label) ### Species richness
  model1$PD[i] <- sum(trx$edge.length) ### Phylogenetic diversity (PD)
  model1$MPD[i] <- mpd(h,cophenetic(trx)) ### Mean phylogenetic distance (MPD)
  model1$MNTD[i] <- mntd(h,cophenetic(trx)) ### Mean nearest taxon distance (MNTD)
}

```

```

model1$PSV[i] <- psv.b(h,trx) ### Phylogenetic species variability (PSV)

model1$MRD[i] <- MRD.b(trx) ### Mean root distance (MRD) The mean number of nodes that separating sp
model1$MRD.time[i] <- mean(max(branching.times(trx))-spp.ages.p$age) ### Mean root distance (MRD.time)

### Species evolutionary distinctiveness (ED)
#1) Prune the tree first and then pass the tree in
model1$ED[i] <- mean(evol.distinct(trx, type = c("fair.proportion"), scale = TRUE, use.branch.lengths
#2) Subset the resulting vector
model1$ED2[i] <- mean(subED[,2])

### Species-level diversification rate (DivRate)
#1) Prune the tree first and then pass the tree in
model1$DivRate[i] <- 1/model1$ED[i]
#2) Subset the resulting vector
model1$DivRate2[i] <- mean(1/subED[,2])

model1$RBL[i] <- mean(trx$edge.length/(as.numeric(branching.times(trx)[1]))) ### Relative branch leng

model1$DR[i] <- log(model1$SPD[i])/max(branching.times(trx)) ### Diversification rates (DR)
model1$GAM[i] <- gamStat(branching.times(trx),return.list=FALSE) ### Gamma statistics (GAM)
model1$IMY[i] <- colless(as.treeshape(trx),norm="yule") ### Colless Yule model (IMY)
model1$IMP[i] <- colless(as.treeshape(trx),norm="pda") ### Colless PDA model (IMP)

model1$age[i] <- max(branching.times(trx)) ### max lineage age
}

```

4.2. Manipulate time (age):

```

# set parameters
ages <- rnorm(1000, 100, 20)
N <- 50 # number of simulation
n <- 100 # number of taxa (tips)

# sample ages
tsamp<-sample(ages,N,replace = T)

# Create table for storing results
model2<-data.frame(matrix(data=NA,nrow=N,ncol=length(phyvars)+2))
colnames(model2)<-c("age", "run", phyvars)

# Simulate N trees under a uniform birth-death process
for(i in 1:N){
  #cat("\r",i,"of", N)
  time<-tsamp[i]
  trx <- sim.bd.taxa.age(n=100, numbsim=1, lambda=log(n/2)/time, mu=0, age=time, mrca=TRUE)
  trx <- trx[[1]]

  # get age for each species
  spp.ages<-data.frame(matrix(data=NA,nrow=length(trx$tip.label),ncol=2))
  colnames(spp.ages)<-c("Species", "age")
}

```

```

max<-max(branching.times(trx))
for (j in 1:length(trx$tip.label)){
  Spp<-trx$tip.label[j]
  spp.ages$age[j] <- trx$edge.length[which.edge(trx,Spp)]
  spp.ages$Species[j]<-Spp
}

# get ED for each species
spp.ED <- evol.distinct(trx, type = c("fair.proportion"), ### Species evolutionary distinctiveness (ED)
                        scale = TRUE, use.branch.lengths = TRUE)

# needed for MPD, MNTD and PSV
h<-data.frame(rep(1,length(trx$tip.label)))
rownames(h)<-trx$tip.label
h<-t(h)

model2$run[i] <- i
model2$SPD[i] <- length(trx$tip.label) ### Species richness
model2$PD[i] <- sum(trx$edge.length) ### Phylogenetic diversity (PD)
model2$MPD[i] <- mpd(h,cophenetic(trx)) ### Mean phylogenetic distance (MPD)
model2$MNTD[i] <- mntd(h,cophenetic(trx)) ### Mean nearest taxon distance (MNTD)
model2$PSV[i] <- psv.b(h,trx) ### Phylogenetic species variability (PSV)

model2$MRD[i] <- MRD.b(trx) ### Mean root distance (MRD) The mean number of nodes that separating sp
model2$MRD.time[i] <- mean(max(branching.times(trx))-spp.ages$age) ### Mean root distance (MRD.time)

model2$ED[i] <- mean(spp.ED[,2]) ### Species evolutionary distinctiveness (ED)
model2$ED2[i] <- mean(spp.ED[,2]) ### Needed to repeat - matching tables of results
model2$DivRate[i] <- mean(1/spp.ED[,2]) ### Species-level diversification rate (DivRate)
model2$DivRate2[i] <- mean(1/spp.ED[,2]) ### Needed to repeat - matching tables of results

model2$RBL[i] <- mean(trx$edge.length/(as.numeric(branching.times(trx)[1]))) ### Relative branch leng

model2$DR[i] <- log(model2$SPD[i])/time ### Diversification rates (DR)
model2$GAM[i] <- gamStat(branching.times(trx),return.list=FALSE) ### Gamma statistics (GAM)
model2$IMY[i] <- colless(as.treeshape(trx),norm="yule") ### Colless Yule model (IMY)
model2$IMP[i] <- colless(as.treeshape(trx),norm="pda") ### Colless PDA model (IMP)
model2$age[i] <- time
}

```

4.3. Manipulate diversification rate:

```

# set parameters
lamb = rexp(10000, rate=2)
lamb = range01(lamb)
lamb = lamb+0.0001
N <- 50 # number of simulation
n <- 100 # number of taxa (tips)

#simulate div values
tsamp=sample(lamb,N,replace = T)

```

```

# Create table for storing results
model3<-data.frame(matrix(data=NA,nrow=N,ncol=length(phyvars)+2))
colnames(model3)<-c("age","run",phyvars)

# Simulate N trees under a uniform birth-death process

rate<-rep(NA,N)
for(i in 1:N){

  #cat("\r",i,"of", N)
  ts<-tsamp[i]
  trx <- sim.bd.taxa.age(n=100, numbsim=1, lambda=ts, mu=0.14, age=40, mrca=TRUE) # change mu to 0.14
  trx <- trx[[1]]

  # get age for species
  spp.ages<-data.frame(matrix(data=NA,nrow=length(trx$tip.label),ncol=2))
  colnames(spp.ages)<-c("Species","age")
  max<-max(branching.times(trx))
  for (j in 1:length(trx$tip.label)){
    Spp<-trx$tip.label[j]
    spp.ages$age[j] <- trx$edge.length[which.edge(trx,Spp)]
    spp.ages$Species[j]<-Spp
  }

  # get ED for each species
  spp.ED <- evol.distinct(trx, type = c("fair.proportion"), ### Species' evolutionary distinctiveness
    scale = TRUE, use.branch.lengths = TRUE)

  h<-data.frame(rep(1,length(trx$tip.label)))
  rownames(h)<-trx$tip.label
  h<-t(h)

  rate[i] <- ts
  model3$run[i] <- i
  model3$SPD[i] <- length(trx$tip.label) ### Species richness
  model3$PD[i] <- sum(trx$edge.length) ### Phylogenetic diversity (PD)
  model3$MPD[i] <- mpd(h,cophenetic(trx)) ### Mean phylogenetic distance (MPD)
  model3$MNTD[i] <- mntd(h,cophenetic(trx)) ### Mean nearest taxon distance (MNTD)
  model3$PSV[i] <- psv.b(h,trx) ### Phylogenetic species variability (PSV)

  model3$MRD[i] <- MRD.b(trx) ### Mean root distance (MRD) The mean number of nodes that separating sp
  model3$MRD.time[i] <- mean(max(branching.times(trx))-spp.ages$age) ### Mean root distance (MRD.time)

  model3$ED[i] <- mean(spp.ED[,2]) ### Species evolutionary distinctiveness (ED)
  model3$ED2[i] <- mean(spp.ED[,2]) ### Needed to repeat - matching tables of results
  model3$DivRate[i] <- mean(1/spp.ED[,2]) ### Species-level diversification rate (DivRate)
  model3$DivRate2[i] <- mean(1/spp.ED[,2]) ### Needed to repeat - matching tables of results

  model3$RBL[i] <- mean(trx$edge.length/(as.numeric(branching.times(trx)[1]))) ### Relative branch leng

  model3$DR[i] <- log(model3$SPD[i])/max(branching.times(trx)) ### Diversification rates (DR)
  model3$GAM[i] <- gamStat(branching.times(trx),return.list=FALSE) ### Gamma statistics (GAM)

```

```

model3$IMY[i] <- colless(as.treeshape(trx),norm="yule") ### Colless Yule model (IMY)
model3$IMP[i] <- colless(as.treeshape(trx),norm="pda") ### Colless PDA model (IMP)

model3$age[i] <- as.numeric(branching.times(trx)[1]) ### max lineage age
}

```

5. Table of results:

Statistical relationship between examined phylometrics, species richness, time and diversification rate, used in the simulations. Relationships are reported as Spearman's rank correlations.

```
kable(results, format = "markdown")
```

	Richness - rho		Richness - P-value		Time - rho		Time - P-value		Div. Rate - rho		Div. Rate - P-value	
SPD	1.00	<0.001			NA	NA			NA	NA		
PD	0.98	<0.001			0.93	<0.001			-0.96	<0.001		
age	NA	NA			1.00	<0.001			NA	NA		
MRD	0.86	<0.001			0.34	0.016			-0.08	0.588		
MRD.time	0.84	<0.001			0.99	<0.001			0.96	<0.001		
ED	-0.93	<0.001			-0.10	0.479			-0.96	<0.001		
ED2	-0.01	0.926			-0.10	0.479			-0.96	<0.001		
RBL	-0.94	<0.001			-0.10	0.479			-0.96	<0.001		
DR	1.00	<0.001			-1.00	<0.001			NA	NA		
MPD	-0.16	0.275			0.94	<0.001			-0.86	<0.001		
MNTD	-0.84	<0.001			0.85	<0.001			-0.96	<0.001		
PSV	-0.16	0.275			0.94	<0.001			-0.86	<0.001		
GAM	-0.52	<0.001			-0.12	0.414			0.95	<0.001		
IMY	-0.04	0.769			0.35	0.012			-0.11	0.433		
IMP	-0.30	0.033			0.35	0.012			-0.11	0.433		
DivRate	0.93	<0.001			0.21	0.15			0.96	<0.001		
DivRate2	-0.05	0.71			0.21	0.15			0.96	<0.001		

Please cite the original publication.

Contact brunno.oliveira@me.com for any further information.