Cor-STRATES

Axel Arango

2022-11-03

This is the code to perform the framework correlated speciation and trait rates simulation (Cor-STRATES), proposed in: Cooney, C.R., Thomas, G.H. Heterogeneous relationships between rates of speciation and body size evolution across vertebrate clades. Nat Ecol Evol 5, 101–110 (2021). https://doi.org/10.1038/s41559-020-01321-y

Cor-STRATES aims to diminsh the errors and biases that could negatively impact the results the relationship between specation rates and trait evolutionary rates.

This framework is done in three parts: -Speciation rate calculation -Trait evolutionary rate calculation -Simulations and correlations

#1) Speciation rate calculation

This first part (Speciation rate calculation), can be done in different ways and methods. However, Cooney & Thomas recommend the use of speciation rates obtained from BAMM, but they also mention that the DR measure could also be useful for small trees (~150 tips). They do not test ClaDS, but they conclude that could be as robust as using BAMM.

For Cor-STRATES we will use the *Average branch length* of the speciation rates for BAMM.

Firs, we will load the necessary packages:

```r
library(BAMMtools)
```

```
## Loading required package: ape
```

```r
library(phytools)
```

```
## Loading required package: maps
```

```r
library(geiger)
```

Then the necessary data:

```r
embtree<-read.tree("MCC_corrected_hwipruned.tre") #the phylogenetic tree (already pruned)
embtrait<-read.table("HWI_data_cap3_div.txt",header=F,sep="\t");names(embtrait)<-c("sp","hwi") # the HW
```

We also are importing the event data file from BAMM, previously calculated.

```r
edata<-getEventData(phy = embtree,"BAMM_event_emberizo_capitulo3.txt",burnin = 0.1)
```

```
## Reading event datafile:  BAMM_event_emberizo_capitulo3.txt
##      ...........
## Read a total of 100 samples from posterior
##
## Discarded as burnin: GENERATIONS <  90000
## Analyzing  91  samples from posterior
##
## Setting recursive sequence on tree...
##
## Done with recursive sequence
```

Once you have the event data, Cor-STRATES suggests that we should obtain branch lengths as the value for speciation rates in the case of BAMM.

Obtain them as follows:

```r
meanbranchtree<-getMeanBranchLengthTree(edata,rate = "speciation")


meanbranchrates<-meanbranchtree$phy$edge.length
head(meanbranchrates) #these would be the speciation rates for branch for Emberizoidea
```

```
## [1] 0.9184761 0.4767351 0.2494827 0.1719776 0.1738514 0.1920207
```

#2) Trait evolutionary rate calculation

For obtaining the phenotypic evolutionary rates, Cor-STRATES suggests that, first, we need to calculate the phylogenetic signal of the attribute of interest using *Pagel's lambda*.

To do so, we will emply the *phylosig* function from phytools.

```r
lambda<-phylosig(embtree,embtrait$hwi,method = "lambda",test = T,nsim = 1000)
```

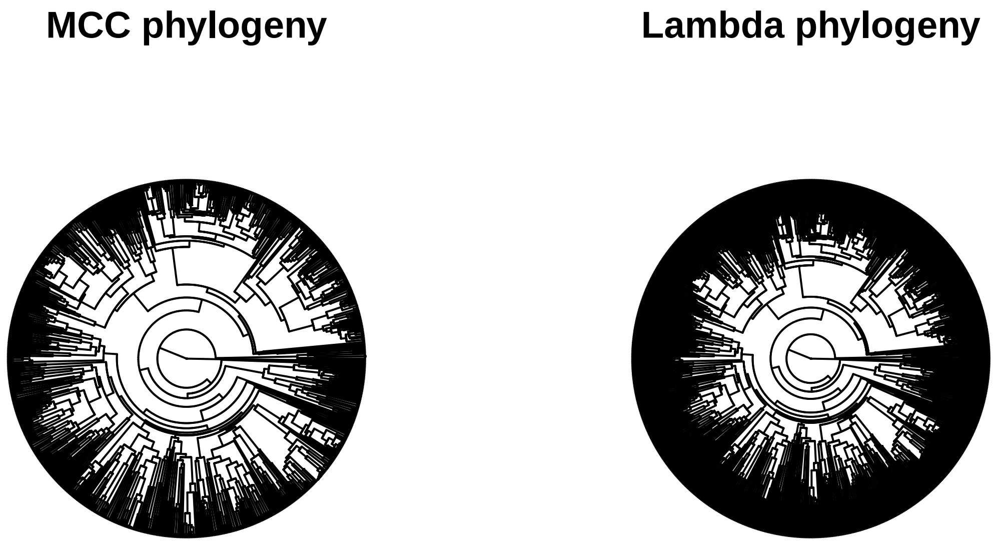
```
## [1] "x has no names; assuming x is in the same order as tree$tip.label"
```

```r
lambda
```

```
##
## Phylogenetic signal lambda : 0.835194
## logL(lambda) : -2184.99
## LR(lambda=0) : 466.142
## P-value (based on LR test) : 2.21497e-103
```

With the *lambda* value obtained, we would use it to reescalate the phylogenetic tree.

```r
rescalelambda<-rescale(embtree,model="lambda",lambda$lambda)
par(mfrow=c(1,2))
plot(embtree,show.tip.label =F,type = "fan",main="MCC phylogeny")
plot(rescalelambda,show.tip.label = F,type = "fan",main="Lambda phylogeny")
```



Now we can save the reescaleted phylogeny for future uses:

```
write.tree(rescalelambda,"lambdatree.tre")
```

Using the phylogeny rescaled with lambda values, the models for phenotypic evolution will be fitted, either with *BAMM* or *BayesTraits*

To do this with *BAMM*, we will first calculate the priors. Priors for phenotypic evolution should include the observed values of the trait, so we need to load or fit them as a named vector first

```
phenotrait<-embtrait$hwi#named vector object
names(phenotrait)<-embtrait$sp
phenoprior<-setBAMMpriors(phy = rescalelambda,832,traits =phenotrait,outfile = NULL )
```

Once the priors are calculated, we can proceed to create a control file to perform the phenotypic *BAMM*:

```
generateControlFile(file = "lambdacontrol.txt",type = "trait", params = list(
  treefile = 'lambdatree.tre',
  traitfile= 'HWI_data_cap3_div.txt',
    seed= '-1',
  eventDataOutfile= 'eventrait.txt',
    numberOfGenerations = '100000',
    overwrite = '1',
    betaInitPrior = as.numeric(phenoprior['betaInitPrior']),
     betaShiftPrior  = as.numeric(phenoprior['betaShiftPrior']),
    ObservedMinMaxAsTraitPriors = as.numeric(phenoprior['ObservedMinMaxAsTraitPriors']),
    expectedNumberOfShifts = '1',
  chainSwapFileName='chain_swap_trait.txt',
  mcmcOutfile='mcmc_out_trait.txt'))
```

With this control file we can now perform a phenotypic *BAMM*.

To use BayesTraits for the calculation of the phenotypic evolutionary rate, this has to be done outside of R. See: http://www.evolution.reading.ac.uk/BayesTraitsV4.0.0/BayesTraitsV4.0.0.html.

Now, since we are using *BAMM*, we can load the new phenotypic event data previously calculated using the generated control file.

```
tdata<-getEventData(phy = rescalelambda,"eventrait.txt",burnin = 0.1,type = 'trait')
```

```
## Reading event datafile:  eventrait.txt
##      ...........
## Read a total of 10000 samples from posterior
##
## Discarded as burnin: GENERATIONS <  999000
## Analyzing  9001  samples from posterior
##
## Setting recursive sequence on tree...
##
## Done with recursive sequence
```

Once the phenotypic 'event data' object is loaded, we will proceed to calculate the average branch length to obtain the trait evolution rate, same as we did with the speciation rates:

```
traitbranchtree<-getMeanBranchLengthTree(tdata,rate = "trait")


traitranchrates<-traitbranchtree$phy$edge.length
head(traitranchrates)
```

```
## [1] 6.219594 5.068525 4.262361 3.650371 3.835046 3.391068
```

With both the speciation rate and phenotypic evolutionary rate, we will do a spearman correlation between both rates.

We would use spearman correlation due to the highly non-normal distribution of the rate values:

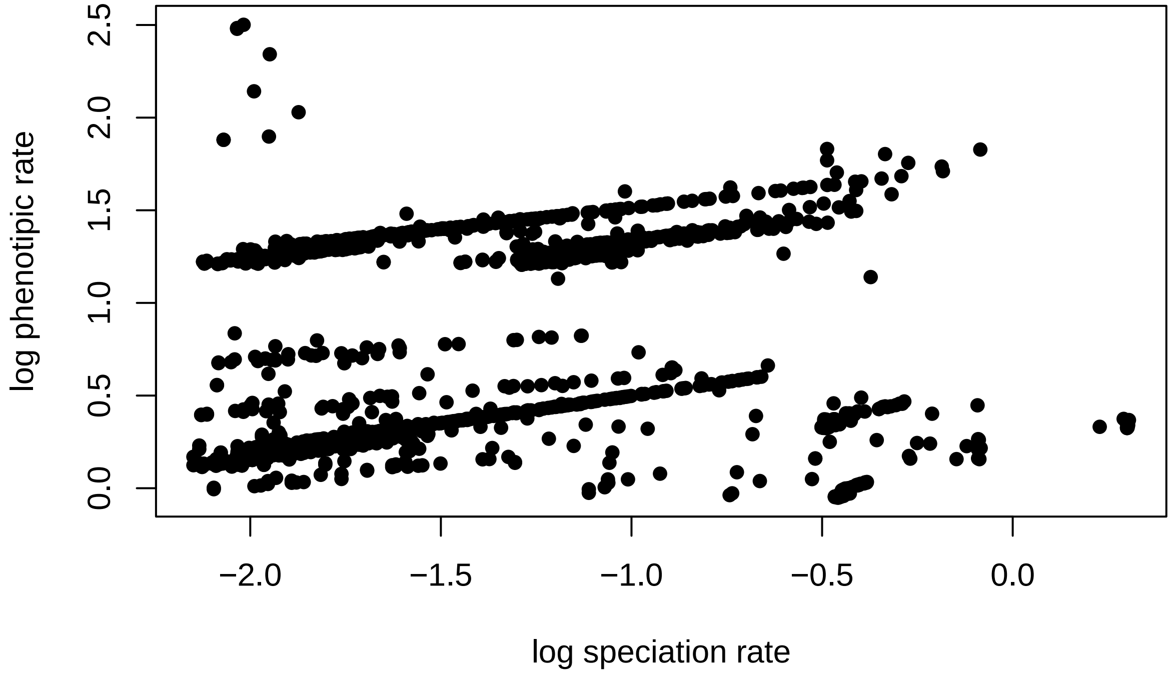
Cooney & Thomas (2021) say: *"We used Spearman's rank correlation (Rho) to measure the association between rates because the distribution of estimated speciation and/or trait rates is often highly non-normal, even after log transformation, which makes applying parametric statistics such as Pearson's r problematic".*

```
obs.cor<-cor.test(log(meanbranchrates),log(traitranchrates),method = "spearman")
obs.cor
```

```
##
##  Spearman's rank correlation rho
##
## data:  log(meanbranchrates) and log(traitranchrates)
## S = 383756650, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2897321
```

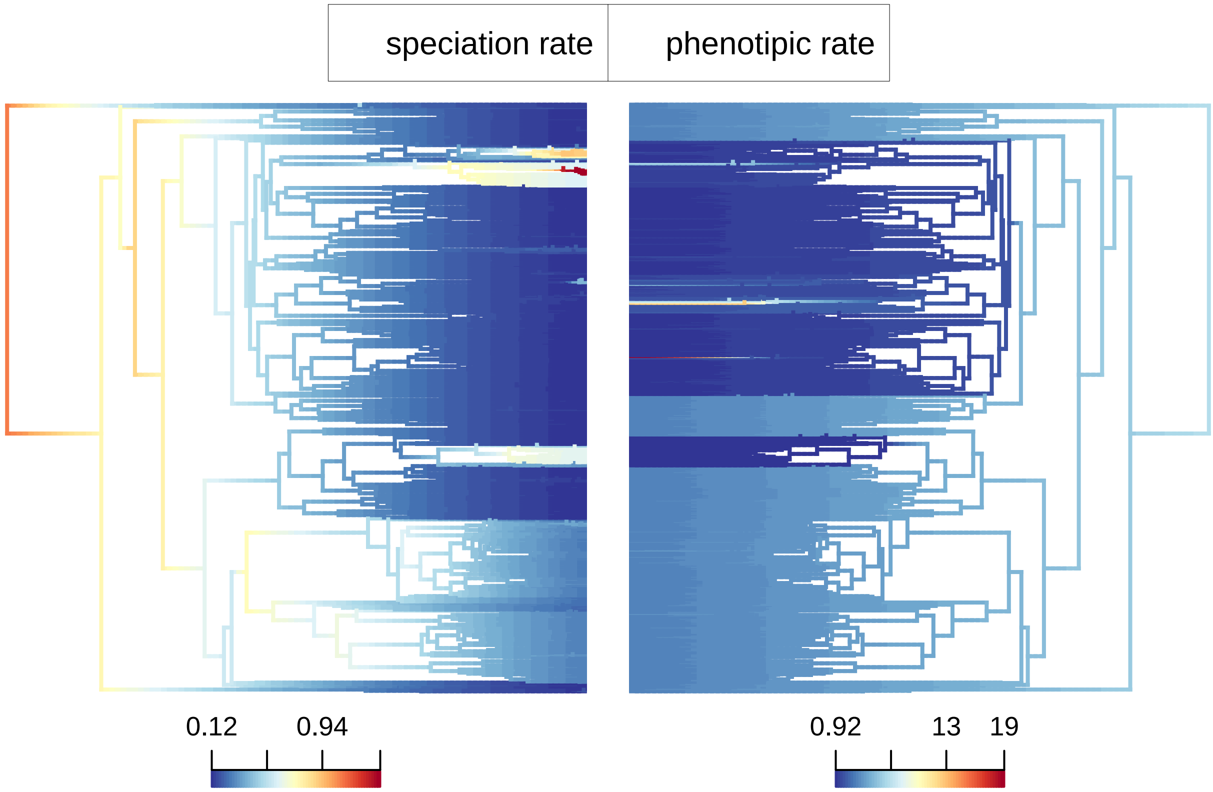
Now we can plot it

```
plot(log(meanbranchrates),log(traitranchrates),xlab="log speciation rate",ylab=("log phenotipic rate"),
```



And compare both rates in the phyologeny:

```
par(mfrow=c(1,2),mar=c(3,0,3,0),xpd=T)
plotedata<-plot.bammdata(edata,lwd=2,direction = "rightwards")
addBAMMlegend(plotedata,location = "bottom")
legend("bottomright",inset=c(0,1),legend = "speciation rate",box.lwd = 0)
plottdata<-plot.bammdata(tdata,lwd=2,direction="leftwards")
addBAMMlegend(plottdata,location = "bottom",)
legend("bottomleft",inset=c(0,1),legend="phenotipic rate",box.lwd=0)
```



#Simulación de datasets nulos

Once we got the observed correlation between the specation and the phenotypic evolutionary rate, we will create null datasets simulating dispersal ability values. After this, we will adjust models of phenotypic evolution, to finally test the significance between the observed relationship and the null relationships.

Cor-STRATES uses datasets simulated under Brownian Motion (BM). To recreate them, we will first adjust a BM model to the trait of interest, using the reescaled phylogeny.

```
trait<-embtrait$hwi
names(trait)<-embtrait$sp
HWIBM<-fitContinuous(rescalelambda,trait,model = "BM")
```

Then, we will use the sigma squared value from the BM adjusted model, to simulate the attribute value for a 100 datasets.

```
sigsqz<-HWIBM$opt$sigsq
simus<-sim.char(rescalelambda,sigsqz,nsim=100,model="BM")
```

Finayll, we will write every simulation on independendent files, with the goal to use them to adjust models of phenotypic evolution using BAMM.

```
setwd("BM_sim/")
for(i in 1:100){
  simx<-simus[,,i]
  simx2<-as.data.frame(simx)
  names(simx2)<-NULL
  write.table(simx2,paste("BMSIM",i,".txt",sep = ""),sep="\t",col.names = F,row.names = T,quote = F)

}
```

Using these datasets of simulated attributes, we calculate the phenotypic evolutionary rates, using the same BAMM protocols that the ones from the observed phenotypic evolutionary rate.

We should have (in this particular case) 100 Event Data files for phenotypic evolution, which are going to be used to calculate the null correlations.

```
nullcors<-data.frame()

#nullrates<-data.frame()
for (i in 1:100){
  sdata<-getEventData(phy = rescalelambda,paste0("Sims/ES",i,".txt"),burnin = 0.1,type = 'trait',verbos
  print(i)
  simbranchtree<-getMeanBranchLengthTree(sdata,rate = "trait") #we aquire its phenotypic evolutionary r
  simbranchrates<-simbranchtree$phy$edge.length #we extract such rates
  ncor<-cor.test(log(meanbranchrates),log(simbranchrates),method = "spearman") #correlate the observed
  nrho<-ncor$estimate #save the estimated value for Rho
  print(nrho)
  names(nrho)<-"Rho"
  nullcors<-rbind(nullcors,nrho)


}
sim.cor<-nullcors
names(sim.cor)<-"Rho"
```

Finally we calculate the p value of the observed correlation in comparison with the null correlations acquired using simulated data.

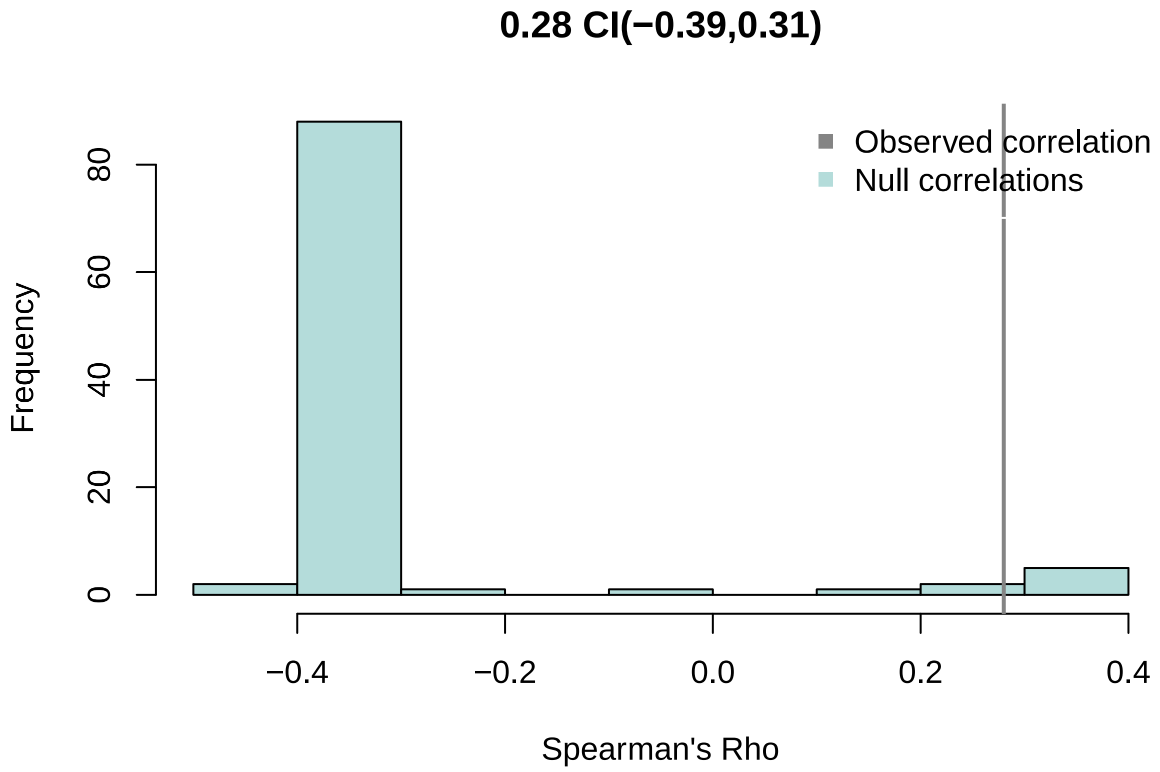
```
sim.corx<-sim.cor$Rho
sims<-100
upper <- (length(sim.corx[sim.corx >= obs.cor$estimate])+1)/(sims+1)
  lower <- (length(sim.corx[sim.corx <= obs.cor$estimate])+1)/(sims+1)
  pval <- 2*min(c(upper,lower))
  if (pval == 0) { pval <- 2*(1/sims) }
  cis <- quantile(sim.corx, probs = c(0.025, 0.975))
  ses <- (obs.cor$estimate - mean(sim.corx)) / sd(sim.corx)
  out <- c(obs.cor$estimate, mean(sim.corx), cis[1], cis[2], pval, ses)
  names(out) <- c("obs.cor", "sim.cor.mean", "sim.cor.lci", "sim.cor.uci", "pval", "ses")



out
```

```
##      obs.cor sim.cor.mean  sim.cor.lci  sim.cor.uci         pval          ses
##    0.2897321   -0.3029550   -0.3917119    0.3162362    0.1386139    3.3416480
```

Unfortunately, it seems that the phenotypic evolution rate, despite being correlated with speciation, is not different from what would be expected by Brownian motion.

We can plot the distribution of the null correlations and the value of the observed correlation to corroborate.

```
hist(sim.cor$Rho,main="0.28 CI(-0.39,0.31)",xlab="Spearman's Rho",col="#B4DCDA")
abline(v=0.28,col="#848484",lwd=2)
legend(x="topright",legend = c("Observed correlation", "Null correlations"),pch = 15,cex=1, col= c("#848
```


0.28 CI(−0.39,0.31)
■ Observed correlation
■ Null correlations
Frequency
Spearman's Rho