

# QuaSSE

Axel Arango

2024-01-17

This is the code to perform the Quantitative State-dependent Speciation and Extinction (QuaSSE) models, first proposed in: FitzJohn, R. G. (2010). Quantitative traits and diversification. *Systematic biology*, 59(6), 619-633.

QuaSSE models are used to test the influence of a continuous trait with speciation.

To perform such models we first will need to load the required packages:

```
library(diversitree)
```

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

```
library(caper)
```

```
## Loading required package: MASS
```

```
## Loading required package: mvtnorm
```

```
library(phytools)
```

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

After we have loaded the packages, now we will need the phylogenetic and trait data for our species.

```
phy<-read.tree("hwi_tree.txt")
hwi<-read.table("hwi_data.txt",sep="\t")
head(hwi)
```

```
##                               hwi
## Rhynchophanes_mccownii      34.80553
## Plectrophenax_nivalis       39.47588
## Plectrophenax_hyperboreus  36.65013
## Calcarius_lapponicus        34.76012
## Calcarius_ornatus           26.99860
## Calcarius_pictus            39.47588
```

Now, need a starting point for the models to start looking for the maximum likelihood. This will require the phylogenetic data, a named vector with trait values for our species and the standard deviation of such trait values.

We already have the phylogenetic data, but now we have to transform our trait data into a named vector, and calculate the standard deviation of the trait values.

First the transformation

```
states<-hwi$hwi
names(states)<-row.names(hwi)
head(states)
```

```
##      Rhynchophanes_mccownii      Plectrophenax_nivalis Plectrophenax_hyperboreus
##              34.80553              39.47588              36.65013
##      Calcarius_lapponicus      Calcarius_ornatus      Calcarius_pictus
##              34.76012              26.99860              39.47588
```

And now the trait values standard deviation

```
sx<-sd(states)
sx
```

```
## [1] 6.327958
```

With this, we can calculate the starting point for the likelihood function to find the maximum likelihood.

```
p<-starting.point.quasse(phy,states,sx)
p
```

```
##      lambda      mu diffusion
## 0.2293827 0.0000000 0.9180176
```

Now we create a function to calculate a linear relationship, this would provide us with a minimal set of state-varying functions, useful for the model computation.

```
xr <- range(states) + c(-1,1) * 20 * p["diffusion"]
linear.x <- make.linear.x(xr[1], xr[2])
```

In order to make easier setting up multiple models, we can create a function that considers the estimated proportion of extant species included in the analysis (sampling fraction), takes the phylogenetic and trait data and only varies with the type of relationship for both speciation and extinction.

```
make.hwi <- function(lambda, mu)
  make.quasse(phy, states,sx,sampling.f=0.89, lambda, mu)
```

And, since our first models are going to be the simple ones, we are going to set a function to constrain the drift value (Evolutionary directional tendency of the trait) to 0.

```
nodrift <- function(f)
  constrain(f, drift ~ 0)
```

With these we can create the QuaSSE models. While the function we made can admit varying values of both speciation and extinction, right now we only care about the influence of the trait in speciation, so we are going to leave extinction as constant while creating the likelihood functions.

```

#model where trait does not have influence in neither speciation or extinction
f.c <- make.hwi(constant.x, constant.x)
#model where the influence of the trait in speciation is linear
f.l <- make.hwi(linear.x, constant.x)
#model where the influence of the trait in speciation has a turning point
f.s <- make.hwi(sigmoid.x, constant.x)
#model where the influence of the trait in speciation
#is optimal at intermediate values
f.h <- make.hwi(noroptimal.x, constant.x)

```

Having this likelihood functions, we can set up our first QuaSSE model, which is going to be our null model, where the trait had no relationship with speciation (constant model)

However, since this models can take a long time to find the maximum likelihood, we can import the previously calculated. The upcoming steps were exactly the same while calculating them, we are just skipping the likelihood search.

```

load("Quasse_v3.RData")

#we set up a control object to help us find the maximum likelihood
control <- list(parscale=.1, reltol=0.001)
#find.mle finds the maximum likelihood for our data
#depending on the likelihood function
#see that we have added the drift constrain,
#the constant function and the starting parameters
mle.c <- find.mle(nodrift(f.c), p, lower=0, control=control,
                 verbose=0)

```

With our Null model out, we can extract the constant parameters for the parameter estimation of the other models.

```

p.c <- mle.c$par
p.c

##      lambda      mu      diffusion
## 2.445650e-01 4.761118e-09 9.182057e-01

```

In case of the linear model, the slope parameter

```

p.l <- c(p.c[1], l.m=0, p.c[2:3])
p.l

##      lambda      l.m      mu      diffusion
## 2.445650e-01 0.000000e+00 4.761118e-09 9.182057e-01

```

For the sigmoidal and the hump model we add the parameters for the minimum and maximum values of lambda, along with the inflexion point for the sigmoidal model and the optimal values for the hump model.

```

p.l <- c(p.c[1], l.m=0, p.c[2:3])
p.s <- p.h <- c(p.c[1], p.c[1], mean(xr), 1, p.c[2:3])
names(p.s) <- argnames(nodrift(f.s))
names(p.h) <- argnames(nodrift(f.h))
print(c(p.s,p.h))

```

```
##          l.y0          l.y1          l.xmid          l.r          m.c          diffusion
## 2.445650e-01 2.445650e-01 2.073794e+01 1.000000e+00 4.761118e-09 9.182057e-01
##          l.y0          l.y1          l.xmid          l.s2          m.c          diffusion
## 2.445650e-01 2.445650e-01 2.073794e+01 1.000000e+00 4.761118e-09 9.182057e-01
```

Having set up these parameters, we can now run our models where we test the influence and the type of relationship that the trait has on speciation rates

```
mle.l <- find.mle(nodrift(f.l), p.l, control=control, verbose=0)
mle.s <- find.mle(nodrift(f.s), p.s, control=control, verbose=0)
mle.h <- find.mle(nodrift(f.h), p.h, control=control, verbose=0)
```

Having found the maximum likelihood for all our likelihood functions, we can do a likelihood ratio test to see if they are significantly different from the null model:

```
anova(mle.c, linear=mle.l, sigmoidal=mle.s, hump=mle.h)
```

```
##          Df    lnLik    AIC  ChiSq Pr(>|Chi|)
## minimal    3 -4049.7 8105.5
## linear     4 -4049.7 8107.4  0.098    0.7542
## sigmoidal   6 -4048.9 8109.8  1.689    0.6395
## hump        6 -4031.4 8074.8 36.723  5.266e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let's see the parameter estimates for the best model

```
mle.h$par
```

```
##          l.y0          l.y1          l.xmid          l.s2          m.c          diffusion
## 1.781648e-01 9.138044e-01 2.131558e+01 5.507799e-02 2.354549e-08 9.198203e-01
```

Since the only model that showed to be significantly different from the null model was the unimodal model, we can add the drift parameter and compare them. Furthermore, this parameter will help us determine the directional trend in which the trait value is evolving to.

```
mle.d.h <- find.mle(f.h, coef(mle.h, TRUE), control=control, verbose=0)
```

Let's find out the differences when including the drift in the model

```
anova(mle.c, linear=mle.l, sigmoidal=mle.s, hump=mle.h,
      drift.hump=mle.d.h)
```

```
##          Df    lnLik    AIC  ChiSq Pr(>|Chi|)
## minimal    3 -4049.7 8105.5
## linear     4 -4049.7 8107.4  0.098    0.7542
## sigmoidal   6 -4048.9 8109.8  1.689    0.6395
## hump        6 -4031.4 8074.8 36.723  5.266e-08 ***
## drift.hump   7 -4017.5 8049.0 64.446  3.367e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

And how the estimation of the drift parameter affects the rest of the parameter for the unimodal QuaSSE model

```
mle.d.h$par
```

```
##          l.y0          l.y1          l.xmid          l.s2          m.c
## 1.704500e-01 8.526504e-01 2.489266e+01 1.509508e-01 5.117482e-08
##          drift          diffusion
## -4.990577e-01 5.813238e-01
```