

Section 1: Impact of k -type fossils on FBD likelihood calculation

In the main text, we describe in the information provided by k -type fossils is simply how numerous they are in the fossil set. That is, it is irrelevant where they are placed on a branch. In this section we provide the R code used to obtain the log-likelihood of the tree and fossil set shown in Figure 2d.

The tree file is found in the distribution of files deposited on Dryad.

```
suppressPackageStartupMessages(library(hisse))
suppressPackageStartupMessages(library(geiger))
load("../Single_Rate/Contour/Surfaces/simTreeContour100.Rsave")
```

First step is to obtain the log-likelihood of the fossil set based on the exact placement of both k -type and m -type fossils. In the following block of code, we will obtain the log-likelihood after conducting an ML search using the equations from Stadler (2010):

```
dat.tab <- hisse::OrganizeDataMiSSE(phy=pp$phy, f=1, hidden.states=1)
fossil.taxa <- which(dat.tab$branch.type == 1)
#This is for starting values:
fossil.ages <- dat.tab$TipwardAge[which(dat.tab$branch.type == 1)]
#Drop all k.samples to get split times:
k.sample.tip.no <- grep("Ksamp*", x=phy$tip.label)
phy.no.k <- drop.tip(pp$phy, k.sample.tip.no)
split.times <- paleotree::dateNodes(phy.no.k,
  rootAge=max(node.depth.edglength(phy.no.k))[-c(1:Ntip(phy.no.k))])
n <- Ntip(phy.no.k)-length(fossil.taxa)
m <- length(fossil.taxa)
x_times <- split.times
y_times <- fossil.ages
k <- dim(pp$k.samples)[1]

starting.point.code <- hisse::starting.point.generator.fossils(n.tax=n, k=1,
  samp.freq.tree=1, q.div=5, fossil.taxa=fossil.taxa,
  fossil.ages=fossil.ages, no.k.samples=k, split.times=split.times,
  get.likelihood=TRUE)

ml.full <- -starting.point.code[4]
mle.pars <- c(starting.point.code[1], starting.point.code[2],
  starting.point.code[3])
mle.pars.trans <- c(mle.pars[1]+mle.pars[2], mle.pars[2]/mle.pars[1],
  mle.pars[3])

print(ml.full)

## [1] -1224.937
```

Next, we will then obtain the log-likelihood using only m -type fossils:

```
lik.m.only <- -hisse::starting.point.tree.fossils(x=log(mle.pars.trans), rho=1,
  n=n, m=length(fossil.taxa), k=0, x_times=split.times,
  y_times=fossil.ages)
```

We can then simply add back the contribution of k -type fossils directly to the log-likelihood:

```
k.portion <- k * log(mle.pars.trans[3])  
tot.lik <- lik.m.only + k.portion  
print(tot.lik)
```

```
## [1] -1224.937
```

Finally, we compare that the two log-likelihoods are identical:

```
comparison <- identical(round(tot.lik, 4), round(ml.full, 4))  
print(comparison)
```

```
## [1] TRUE
```

References

Stadler T. 2010. Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, 267:396-404.