Microbial Speciation

Jay T. Lennon

20 June, 2018

Setup Work Environment

```
rm(list=ls())
getwd()
```

```
## [1] "/Users/lennonj/GitHub/MicroSpeciation/code"
```

```
setwd("~/GitHub/MicroSpeciation/")

# Load packages
require("png")
```

```
## Loading required package: png
```

```
require("grid")
```

```
## Loading required package: grid
```

Load data

```
data <- read.table("~/GitHub/MicroSpeciation/data/MicroSpeciation.txt", sep = "\t", header = T)
```

Exponential approach

Diversification involves the splitting of a lineage into two and so on, so by definition is an exponential process which has been used by evolutionary biologist to model speciation. We use the exponential function Sf = S0 * e^rt, where Sf = species richness at time final, So is speices richnness at time of origin which equals 1, r is the rate of speciation, and t is the amount of time between origin and final

```
time <- data$Appeared - data$Extinction
rate <- (log(data$Richness/1)/time)
data <- data.frame(data, time, rate)
data <- data[order(rate),]
micro <- data[ which(data$Taxon == 'Microbes'),]
micro.r <- micro[,6]
cat("Microbial speciation rate =", round(micro.r, 3), "per million years")
```

```
## Microbial speciation rate = 0.007 per million years
```

```
# To see if this approach is reasonable, it would be interesting to compare (correlation)
# literature-reported estimates of diversification for plants and animals with estimates from
# the exponential model. Perhaps use data from Scholl and Weins (2016)
```

Comparison with substitution-based diversification estimates

We can define a speciation event based on the time it takes for 16S rRNA gene of a given length to accumulate enough substitutions so that there is 3 % divergence. We can that put the corresponding speciation rate back into the exponential and see what levels of richness are predicted.

```
length.g <- 1600 # length of gene
cutoff <- 0.03 # % divergence for new OTU
subs = length.g * cutoff # number of substitutions needed for new OTU

# Substitution rates (per million years) from Kuo and Ochman (2009):
low.sub <- 0.4
hi.sub <- 1.5

# Corresponding divergence time (million years):
div.t.ls <- subs / low.sub # divergence time w/ low substitution rate
div.t.hs <- subs / hi.sub # divergence time w/ high substitution rate

# Corresponding speciation rates (million years):
spec.ls <- 1 / div.t.ls
spec.hs <- 1 / div.t.hs

cat("Speciation rates range from", round(spec.ls, 3), "to", round(spec.hs, 3), "per million years")
```

```
## Speciation rates range from 0.008 to 0.031 per million years
```

```
# Let's plug these speciation rates back into the exponential model:
mic.yrs <- 4000 # millions of years of microbial evolution
Sf.ls <- 1 * exp(spec.ls*mic.yrs)
Sf.hs <- 1 * exp(spec.hs*mic.yrs)

cat("Speciation rates from literature overshoot global Sobs:", formatC(Sf.ls, format = "e", digits = 2)
```

```
## Speciation rates from literature overshoot global Sobs: 3.00e+14 - 1.94e+54
```

```
# What would substitution rate need to be to get 10^12 species with exponential model?
right.sub <- micro.r * 48 # 0.33

# What if substitution rate was a bit lower (0.3)? How sensitive?
Sf.v.low <- 1 * exp(((right.sub - 0.03) /subs)*mic.yrs)

Sf.v.low <- 1 * exp(((right.sub - 0.03)/subs)*mic.yrs)
formatC(Sf.v.low, format = "e", digits = 2)
```

```
## [1] "8.21e+10"
```

What about a logistic model?

In population biology, a logistic model modifies exponential growth rate based on a carrying capacity. We don't konw what the upper limit for species richness on Earth is. Perhaps it's 10^12 or 10^16. It however, cannot be 10^54 as calculated above with fast substitution rates as this would exceed the presumed equilibrium abundance (N) of the biosphere which is thought to be ~10^30. In calculations below, we will solve for richness with the following equation Sf = (K * So * e^rt) / (K + So (e^rt -1))

```r
K <- 10^14
SF.log <- (K * 1 * exp(spec.hs * mic.yrs)) / (K + 1 * exp(spec.hs * mic.yrs -1))
```

Figure

```r
kern <- density(data$rate)

png(filename="~/GitHub/MicroSpeciation/figures/speciation-kernel.png",
        width = 800, height = 800, res = 96*2)

par(mar = c(4, 5, 1, 1))

plot(kern, main = NA, xaxt = "n", yaxt = "n", cex.lab = 1.5, ylab = "",
     xlab = "", xlim = c(-0.05, 0.16), ylim = c(0,12), lwd = 2)

mtext("Density", side = 2, outer = TRUE, cex = 1.5,
      line = -2, adj = 0.5)

mtext(expression('Speciation Rate (mya'^-1*')'), side = 1, outer = TRUE,
      cex = 1.5, line = -1, adj = 0.5)

     axis(side = 1, lwd.ticks = 2, cex.axis = 1.5, las = 1,
        at = c(-0.05, 0.0, 0.05, 0.10, 0.15), labels = T)

     axis(side = 3, lwd.ticks = 2, cex.axis = 1.5, las = 1,
        at = c(-0.05, 0.0, 0.05, 0.10, 0.15), labels = F)

     axis(side = 2, lwd.ticks = 2, cex.axis = 1.5, las = 1,
        at = c(0, 4, 8, 12), labels = T)

    axis(side = 4, lwd.ticks = 2, cex.axis = 1.5, las = 1,
        at = c(0, 4, 8, 12), labels = F)

box(lwd=2)

arrows(0.006907755, 4, 0.006907755, 7, length = 0.075, lwd = 2, col = "Black")
text(0.02, 3.5, labels = "microbes", cex = 1)

# Close Plot Device
dev.off()
```

```
## pdf
##   2
```

```r
graphics.off()

# Show Plot
img <- readPNG("~/GitHub/MicroSpeciation/figures/speciation-kernel.png")
grid.raster(img)
```