Running MiSSE

Jeremy M. Beaulieu

Background

In a recent publication (Caetano et al., 2018) we pointed out that the trait-independent HiSSE model is basically a model for traits and a separate model for shifts in diversification parameters, much like BAMM (though without priors, discontinuous inheritance of extinction probability, or other mathematical foibles). The hidden states can drive different diversification processes, and the traits just evolve under a regular Markovian trait model. At that point, there is no harm in just dropping the trait altogether and just focusing on diversification driven by unknown factors. That is what this MiSSE function does – it sets up and executes a completely trait-free version of a HiSSE model. Thus, all that is required is a tree. The model allows up to 26 possible hidden states in diversification (denoted by A-Z). Transitions among hidden states are governed by a single global transition rate, q. A "shift" in diversification denotes a lineage tracking some unobserved, hidden state. An interesting byproduct of this assumption is that distantly related clades can actually share the same discrete set of diversification parameters.

Note that we refer to "hidden state" simply as a shorthand. We do not mean that there is a single, discrete character that is solely driving diversification differences. There is some heritable "thing" that affects rates, such as a combination of body size, oxygen concentration, trophic level, and, say, how many total species are competing for resources in an area. In other words, it could be that there is some single discrete trait that drives everything. However, it is more likely that a whole range of factors play a role, and we just slice them up into discrete categories, the same way we slice up mammals into carnivore / omnivore / herbivore or plants into woody / herbaceous when the reality is more continuous. This is true for HiSSE, but this concept is especially important to grasp for MiSSE.

Setting up a MiSSE model

The set up is similar to other functions in `hisse`, except there is need to set up a transition model because obviously we do not have any focal trait(s). For the following example, we will use the phylogeny of Cetaceans of Steeman et al. (2009).

```
suppressWarnings(library(hisse))
```

```
## Loading required package: ape
## Loading required package: deSolve
## Loading required package: GenSA
## Loading required package: subplex
## Loading required package: nloptr
```

```
phy <- read.tree("whales_Steemanetal2009.tre")
```

As with `hisse`, rather than optimizing λ_i and μ_i separately, `MiSSE` optimizes transformations of these variables. Again, we let $\tau_i = \lambda_i + \mu_i$ define "net turnover", and we let $\epsilon_i = \mu_i/\lambda_i$ define the "extinction fraction". This reparameterization alleviates problems associated with over-fitting when λ_i and μ_i are highly correlated, but both matter in explaining the diversity pattern (see discussion of this issue in Beaulieu and O'Meara 2016). The number of free parameters in the model for both net turnover and extinction fraction are specified as index vectors provided to the function call. First, let us fit a single rate model:

```r
turnover <- c(1)
eps <- c(1)
one.rate <- MiSSE(phy, f=1, turnover=turnover, eps=eps)
```

Pretty simple. Now to fit a model that contains two rate classes, we will simply expand out the turnover vector:

```r
turnover <- c(1,2)
eps <- c(1,1)
two.rate <- MiSSE(phy, f=1, turnover=turnover, eps=eps)
```

The model allows up to 26 possible hidden states in diversification (denoted by A-Z), and in this example since we fit two rate classes, we have two hidden states, A and B, impacting turnover rates. We can also let ϵ_i vary across the tree:

```r
turnover <- c(1,2)
eps <- c(1,2)
two.rate.weps <- MiSSE(phy, f=1, turnover=turnover, eps=eps)
```

Plotting MiSSE reconstructions

Like with all other functions, we provide plotting functionality in `plot.misse.states()` for hidden state reconstructions of `class misse.states` output by our `MarginReconMiSSE()` function. And, as with other functions, a single `hisse.states` object can be supplied and it will provide a heat map of the diversification rate parameter of choice, or you provide a list of `misse.states` objects the function will "model-average" the results. Users can choose among turnover, net diversification ("net.div"), speciation, extinction, or extinction fraction ("extinction.fraction"). Below is an example `misse.states` output from our data set mentioned above. Let's load this file and check that everything has loaded correctly and is of the proper `misse.states` class:

```r
# two.rate.recon <- MarginReconMiSSE(phy=phy, f=1,  hidden.states=2,
                #pars=two.rate$solution, n.cores=3, aic=two.rate$AIC)
load("misse.vignette.Rsave") # Line above shows the command to create this result.
class(two.rate.recon)
```

```
## [1] "misse.states"
```

```r
two.rate.recon
```

```
##
## Phylogenetic tree with 87 tips and 86 internal nodes.
##
## Tip labels:
##  Balaena_mysticetus, Eubalaena_australis, Eubalaena_glacialis, Eubalaena_japonica, Caperea_marginata
## Node labels:
##  1, 1, 1, 1, 1, 1, ...
##
## Rooted; includes branch lengths.
```

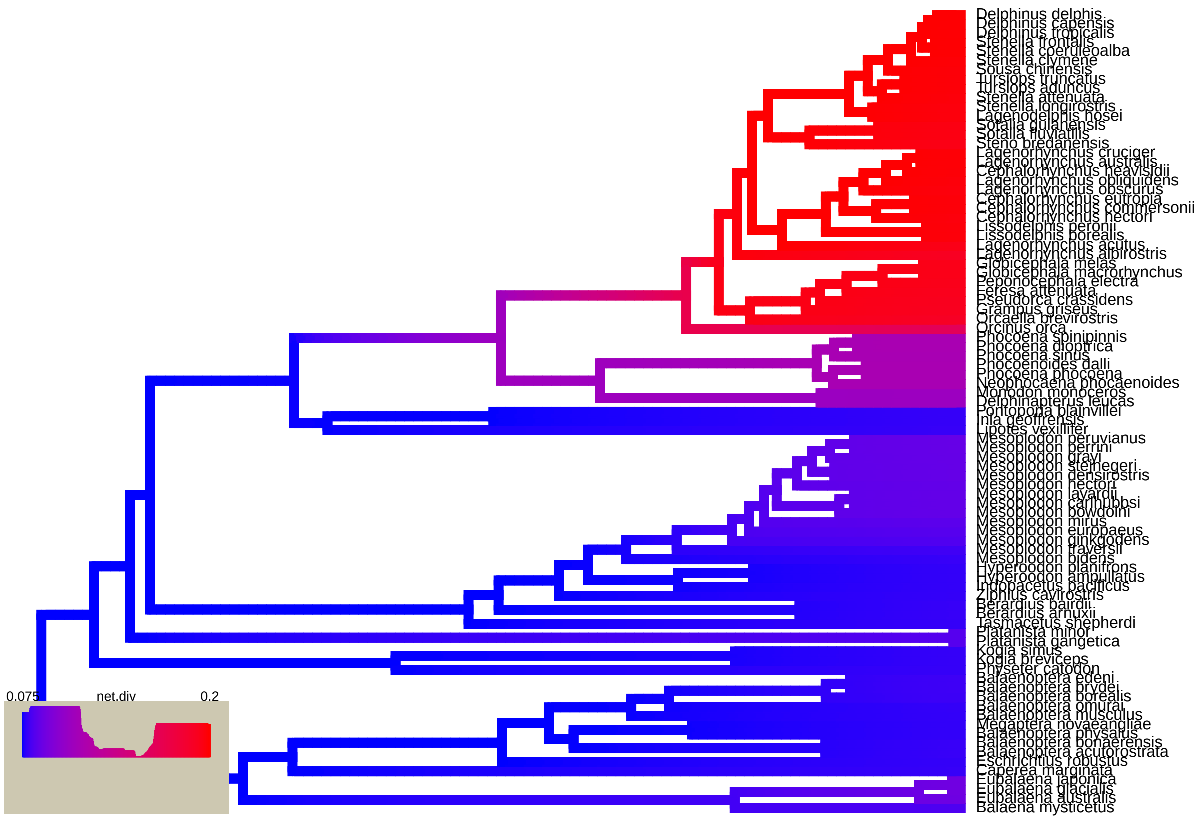
I have conducted reconstructions on the three models shown described above – i.e., one.rate.recon, two.rate.recon, and two.rate.weps.recon. To plot model-averaged rates we simply supply these reconstructions as a list. There are many ways to generate a list, but here is one way, where I'm assuming that the marginal reconstructions from three models are saved to the directory we are working from:

```r
misse.results.list = list()
misse.results.list[[1]] = one.rate.recon
```

```r
misse.results.list[[2]] = two.rate.recon
misse.results.list[[3]] = two.rate.weps.recon
```

And finally, we will supply this list the plotting function, `plot.misse.states()`, and plot net diversification:

```r
plot.misse.states(misse.results.list, rate.param="net.div", show.tip.label=TRUE, type="phylogram",
                  fsize=.5)
```




```
## $rate.tree
## Object of class "contMap" containing:
##
## (1) A phylogenetic tree with 87 tips and 86 internal nodes.
##
## (2) A mapped continuous trait on the range (0.075238, 0.203937).
```

Other considerations

Like with `hisse`, `GeoHiSSE`, and `MuHiSSE`, there are functions available to obtain model averages (i.e., `GetModelAveRates()`) for nodes, for tips, or for both to be used in post-hoc tests.Users are encouraged to read other vignettes and help pages provided for more information. For more conceptual discussions of these functions and ideas, readers are also encouraged to read Caetano et al. (2018).

There are two additional items that are worth mentioning. First, like with `MuHiSSE`, I would recommend users try multiple random starting points when optimizing any given model with `MiSSE`. In Nakov et al. (2018), we found that the default starting values often did not return the highest log likelihood. To alleviate this issue, we performed ≥ 50 maximum likelihood optimizations for each model, each initiated from a distinct starting point. All functions within `hisse` are provided with `starting.vals` option for these purposes.

We also note that `MiSSE` may seem slower than most other functions within `hisse`. This is somewhat intentional. Underneath the hood we have implemented a lot of checks to the integration for calculating probabilities along branches. This will mean that often times weird messages will spit out to the screen. For now, ignore them, the optimization "feels"" them and takes the necessary action. But this also means that users must pay particular attention to the complexity of the models they are fitting and critically think whether or not the parameters make much sense. For example, in whale tree above, I attemped to fit a model with three hidden states, A, B, and C. However, the output indicated that perhaps this model was overly complex for the data at hand:

```
three.rate
```

```
##
## Fit
##           -lnL            AIC           AICc         n.taxa
##      -183.3967       376.7934       377.5341        87.0000
## n.hidden.states
##         3.0000
##
## Model parameters:
##
##  turnover0A       eps0A  turnover0B       eps0B  turnover0C       eps0C
## 0.044841087 0.007690476 0.185779643 0.007690476 0.039766719 0.007690476
##          q0
## 0.010000000
```

Note that the likelihood was a significant improvement from the two rate model, but the transition rate, q, never changed from the initial start value ($q = 0.01$). Also, this particular fit spit out an unusually large number of integration warnings, indicating that this was a rather difficult model to fit to the tree. There are other considerations that users should consider, such as the number of parameters relative to the size of the tree. In the case of the whales, a three rate model consisted of 5 parameters total, which is roughly ~17 taxa per parameter. Our feeling is that, conservatively, there should at least 20 taxa for every parameter in the model. When this assumption is not met, a warning is printed to the screen, but users are still allowed to carry on with their analyses. Again, we hope you take these warnings into consideration when examining the fit of the model and deciding whether or not to cull from the set.

References

Beaulieu, J.M., and B.C. O'Meara. (2016). Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583-601.

Caetano, D.S., B.C. O'Meara, and J.M. Beaulieu. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographic models. Evolution, https://doi.org/10.1111/evo.13602

Nakov, T., Beaulieu, J.M., and Alverson, A.J. 2018. Freshwater diatoms diversify faster than marine in both planktonic and benthic habitats. bioRxiv, doi: https://doi.org/10.1101/406165.

Steeman, M.E., B. M.B. Hebsgaard, E. Fordyce, S.Y.W. Ho, D.L. Rabosky, R. Nielsen, C. Rahbek, H. Glenner, M.V. Sorensen, and E. Willerslev. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. Syst. Biol. 58:573-585.