Table of Contents

Exercise 2

Coalescent multispecies tree (from Stephen Nylinder`s BEAST course, Manaus 2016)

- nuclear_2.nex
- nuclear_1.nex
- concat.nex
- chloroplast.nex

1. Open BEAUTi and load the alignment files Chloroplast.nex, Nuclear_1.nex, and Nuclear_2.nex. They will appear as individual partitions.
2. To begin with, rerun yesterdays exercise using the Concat.nex file where all the genes were concatenated (in this exercise it means linking substitution model, clock model, and tree model for all partititons). Mimic the species tree setup in the next paragraph as close as you can (root height prior, Yule, etc). Summarize the concatenated tree and save it for comparison.
3. Click the top button “Use species tree ancestral reconstruction”. BEAUTi will create a new trait called “species” for you (this trait name is locked to *BEAST). When you click ok you will be immediately transferred to the Traits tab. Here you will assign each taxon in the alignment to a species. (WARNING! A species must not have the same name as a gene in the alignment!). For each pair of taxa (e.g. A1, A2), click “set trait values” and type in “Species_A” for example. Do this for all pairs of taxa, even for the outgroup.
4. If you go back to the Partitions tab again, notice that BEAUTi have unlinked all parameters for all partitions for you. Leave this as it is. Also notice that the Taxa tab has changed name to “Species Sets”. Click on it to define your clades in the species tree. Define a group you call “Ingroup” and set it to be monophyletic. Move all ingroup species there.
5. Go to the “Sites” tab. Assign each gene an HKY+G model with estimated base frequencies.
6. Go to the “Clocks” tab. Change all clocks to Lognormal Relaxed Clocks and click Estimate for each one.
7. Go to “Trees” tab. Stick with the defaults for the species tree prior (Yule) and the population model “Piecewise linear & constant root”. Make sure to change all nuclear genes to be “autosomal nuclear”, and the chloroplast to be “mitochondrial” (haploid).
8. In the “Priors” tab, notice the two new priors at the end. Also notice the improper distributions (1/x). Change these to Exponential distributions with mean 1.0. Also notice there is no speciestree.rootheight! Go back to “Species Sets” and define a group with all taxa in it. Call it “Root”. Go back to “Priors” tab and notice the potential to assign a prior to that group! Set a normal distribution for the Root with mean 10.0 and stdev 0.1. This calibrates the species tree root height to 10 time units. Notice the gene tree root heights are unconstrained per default. Set informative priors on the mean clock rates for each gene (Nuclear_1 = Normal mean 2E-3, stdev 5E-4, Nuclear_2 = Normal mean 1E-3, stdev 2E-4, Chloroplast = Normal mean 9E-4, stdev 2E-4). This is merely to help reach stable state quicker and reduce chain length. Normally this would be a weakly informative prior unless we have other information.
9. Go to the “MCMC” tab and set a chain length (4-5M) and sample every 5.000 generations (~1000 posterior samples). Give it a nice name and generate the file. Place the file in a separate directory and execute it. It will produce a couple of more files as output…
10. Run the analyses, summarize the trees files in TreeAnnotator, and compare the gene tree topology to the estimated species tree. Are they similar? Which one makes more sense? What are the differences in node heights between comparable clades in the trees?

Exercise

1. Open the generated logfile for the species tree analysis in Tracar. Look at how the parameters are behaving. How long does it take until the MCMC samples from the expected species tree root height? What is the relationship between the gene tree root heights and the species tree root height? Which gene tree root height appears closest to the species tree one? Why?
2. Go back to BEAUTi and open the "Traits" tab. Based on the estimated species tree topology from the previous exercise, try to lump at least two pairs of species into the same species (e.g. assign taxa I1, I2, J1, J2 under a new species name). For this to work the current species sets (Ingroup and Root) you have already defined must be deleted. Define a new "Ingroup" and "Root" containing the new species.
3. Rerun the analysis and compare parameter sampling and species tree topology to the previous run. What is different? Node support? Branch lengths? Why?
4. Now, try to assign one specimen to the "wrong" species. Based on the species tree from the first exercise, pick two species far apart in the tree and change one of the species assignment in the "Trait" tab. Generate the xml and rerun the analysis. Compare mcmc behavior, parameter space sampled, and species tree topology to previous runs. What is the difference? How robust is the species tree estimate to mis-assignment of specimens to the correct species?
5. Reproduce the settings of the first analysis in BEAUTi. This time, change the prior on the species.popMean from an exponential distribution with mean 1.0, to an exponential distribution with mean 0.001. This makes a very strong suggestion on the mean population sizes across the species tree. Rerun the analysis and check sampling behavior and parameter space. What are the differences between the two runs? Compare the species tree topologies.

From:
http://www.botanicaamazonica.wiki.br/labotam/ - **Ecologia e Evolução de Plantas Amazônicas**

Permanent link:
http://www.botanicaamazonica.wiki.br/labotam/doku.php?id=cursos:bot99:aulas:praticas:speciestree

Last update: **04/53/2017 10:53**