

13. November 2013

Exercise 1: Analyze the primates data data with beast and with RAxML with bootstrapping and compare the results, that is the tree topologies, the branch lengths, and the posterior probabilities vs. bootstrap values. Do the same also for the lizards data. Explore the role of priors in the Bayesian analyses.

Exercise 2: How much do the results of Beast and RAxML depend on the choice of DNA substitution models. Explore this with the primates data and for the lizards data. For the latter, also explore the effect of partitioning the data.

Exercise 3: Let nucleotide frequencies $(\pi_A, \pi_C, \pi_G, \pi_T)$ and mutation rates α and β of the HKY model be given. Under which conditions can you find rates λ and μ of the Felsenstein 84 model, such that the transition matrices of the two models are the same, and which are the appropriate values for λ and μ ? Discuss also the opposite direction. The F84 rate matrix is:

$$\begin{pmatrix} -\lambda(1 - \pi_A) - \frac{\mu\pi_G}{\pi_A + \pi_G} & \lambda\pi_C & \lambda\pi_G + \frac{\mu\pi_G}{\pi_A + \pi_G} & \lambda\pi_T \\ \lambda\pi_A & -\lambda(1 - \pi_C) - \frac{\mu\pi_T}{\pi_C + \pi_T} & \lambda\pi_G & \lambda\pi_T + \frac{\mu\pi_T}{\pi_C + \pi_T} \\ \lambda\pi_A + \frac{\mu\pi_A}{\pi_A + \pi_G} & \lambda\pi_C & -\lambda(1 - \pi_G) - \frac{\mu\pi_A}{\pi_A + \pi_G} & \lambda\pi_T \\ \lambda\pi_A & \lambda\pi_C + \frac{\mu\pi_C}{\pi_C + \pi_T} & \lambda\pi_G & -\lambda(1 - \pi_T) - \frac{\mu\pi_C}{\pi_C + \pi_T} \end{pmatrix}$$

You can insert positive values for the F84 parameters.

Exercise 4 (mainly for biologists): Find publications about at least two different genera, where fossils have been used to time-calibrate phylogenetic trees. Find out in detail which traits of the fossils were used for their taxonomic classification—you may have to trace this back in other publications—and how this was used in the phylogeny analysis.

Exercise 5 (mainly for bioinformaticians): In the lecture we will discuss four different methods to compute a substitution matrix $S(t)$ from a rate matrix R . Investigate for the PAM rate matrix for amino acids, for the pfold rate matrix for RNA stem basepairs and for the following F84 rate matrix how efficient and how accurate and numerically stable these methods are. The PAM rate matrix and the pfold rate matrix will be available from the website of the lecture. (Note that the conventions about matrix notation are not always strictly followed in Bioinformatics. Sometimes transposed matrices are given, i.e. the roles of rows and columns are interchanged. You should always check for this with any matrix that is given to you!)

Exercise 6: Compute the stationary distributions of the pfold rate matrix and the PAM substitution rate matrix. Are these evolutionary dynamics (almost) reversible? (You can use R or a similar program to solve this exercise. In R the function `eigen` may be helpful.)

Exercise 7: Based on the fossil record for lizards it is claimed that the most recent common ancestor (MRCA) of *Sphenodon punctatus* and *Cnemidophorus tigris* is at least 228.0 Mya old, the MRCA of *Cnemidophorus tigris* and *Rhineura floridana* is at least 113 Mya old, the MRCA of *Rhineura floridana* and *Gallotia galloti* at least 64.2 Mya, and the MRCA of *Timon lepidus* and *Dalmatolacerta oxycephala* at least 5.3 Mya.

- (a) Use this information to estimate the phylogeny of lizards with time calibration.
- (b) How does this result depend on the choice of model for the relaxed clock?
- (c) How does the result change if you use only two or three of the given calibration points?
- (d) How well are the given calibration points supported by fossils documented in the following references?

Science (1990) 249:1020-1023,

J Vertebr Paleontol (2006) 26:795-800,

J Vertebr Paleontol (2002) 22:286-298,

J Paleontol (1985) 59:1481-1485