20. December 2011

Exercise 1: 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 2 (mainly for biologists): (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 3 (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.

Exercise 4: 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.)