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: Assume that 10000 trees haven been MCMC-sampled from the posterior distribution for a given data set. To summarize these trees we would like to show a single tree that has all branches that appear in more than 5000 of the sampled trees (as splits of the taxa sets). Proof that this is always possible or present a counter-example.

Exercise 3: We flip a coin 1000 times to test whether it is fair. If we apply Bayesian inference with a uniform prior on $[0, 1]$ for the probability p that the coin shows "head", how probable is it that we estimate the probability of $\{p > 0.5\}$ to be higher than 90% or lower than 10%?

Exercise 4: Calculate the equilibrium distributions of the Markov chains with transition matrices U and V and check whether the processes are reversible:

$$U = \begin{pmatrix} 0.3 & 0.7 \\ 0.1 & 0.9 \end{pmatrix} \qquad V = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.5 & 0.4 & 0.1 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$$

Exercise 5: 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 6: Assume an ancestral sequence and a derived sequence are given, and we aim to estimate the evolutionary distance between the sequences. For this, we assume a Jukes-Cantor model with rate $\lambda = 1$. Let n be the length of the sequences, and k be the number of segregating sites. (There are no gaps in the alignment or the alignment is known and positions with gaps are not counted.)

(a) Calculate the function $f(t)$, which is the expected number of segregating sites if t is the true time distance between the sequences (assuming the Jukes Cantor model and sequence length n).

(b) Calculate the log-likelihood function $\ell_k(t)$, which is the log of the probability to observe k segregating sites if t is the right time (assuming Jukes Cantor etc.)

(c) The so-called *moment estimator* for t is the $\widetilde{t}$ such that $f(\widetilde{t}) = k$. How does it depend on the observed k? (It is called moment estimator because the expectation value is also called the first moment of a distribution.)

(d) As you know, the ML estimator $\hat{t}$ is the t that maximizes $\ell_k(t)$. How does $\hat{t}$ depend on k?

(e) Compare $\widetilde{t}$ to $\hat{t}$. Can you find an obvious relationship between the two? And if so, does it also hold for other substitution models that assume independence between the sites?