

Exercise 1: Simulate sequence datasets with different trees with 5 and with 10 taxa and with different mutation models (e.g. JC and HKY). Explore how the accuracy of the ML tree found by DNAML depends on the substitution model used in DNAML and the sequence length. Apply also RAxML to the data.

Exercise 2: (Mainly for bioinformaticians) Find a way to simulate data for a gene that has two exons and one intron. In the coding region, the third codon position evolves faster than the first two, and the intron evolves even faster. You can use seq-gen and postprocess the output. For simplicity, neglect that the mutation probabilities of the third codon position depend on the current states of the first two. Simulate several phylogenetic datasets and analyze them with RAxML with and without partitioning. How is the effect of partitioning on the accuracy of the result and how does this depend on the sequence lengths, the number of taxa, and the branch lengths of the tree?

Exercise 3: Let X_1, X_2, \dots be a Markov chain on a finite state space \mathcal{S} with transition probabilities $P_{x \rightarrow y}$ for all $x, y \in \mathcal{S}$. Let $(\pi_x)_{x \in \mathcal{S}}$ be a distribution that fulfills the detailed balance condition

$$\forall x, y \in \mathcal{S} : \pi_x \cdot P_{x \rightarrow y} = \pi_y \cdot P_{y \rightarrow x}.$$

- Show that $(\pi_x)_{x \in \mathcal{S}}$ is a stationary distribution of the Markov chain.
- Do stationary distributions of a Markov chain always fulfill the detailed balance condition? If yes, give a proof; if no, give a counter example.

Exercise 4: 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 5: Find the central branch length ℓ that maximizes the likelihood of the tree shown below (for fixed lengths of the other branches). For the substitution process assume a Jukes-Cantor model with $\lambda = 1$, such that the rate of a change from nucleotide x to any *other* nucleotide y is $\lambda/4$.

