Distance Methods

1. Be able to apply the UPGMA algorithm to find a tree from a distance matrix. (You do not need to memorize the algorithm.)
2. Be able to find the last three edge lengths when combining the final three nodes using the neighbor-joining algorithm.
3. Know what types of trees UPGMA and neighbor-joining produce.
4. Know in principle why distance methods are often less accurate than parsimony and likelihood methods.
5. Know the primary advantage of distance methods.
6. **UPGMA algorithm:**
   (a) Find the i and j with the smallest distance D_{ij}.
   (b) Create a new group (ij) which has $n_{(ij)} = n_i + n_j$ members.
   (c) Connect i and j on the tree to a new node (ij).
   (d) The depth of group (ij) from its leaves is $D_{ij}/2$; set the new edge lengths accordingly.
   (e) Compute the distance between the new group and all other groups except i and j by using
   $$D_{(ij),k} = \left(\frac{n_i}{n_i + n_j}\right) D_{ik} + \left(\frac{n_j}{n_i + n_j}\right) D_{jk}$$
   (f) Delete columns and rows corresponding to i and j and add one for (ij). If there are two or more groups left, go back to the first step.
7. **Neighbor-Joining algorithm:**
   (a) Let n be the number of groups remaining.
   (b) If $n > 3$, for each group, compute $u_i = \sum_{j \neq i} D_{ij}/(n-2)$.
   (c) Choose the i and j for which $D_{ij} - u_i - u_j$ is smallest. (This is a negative number.)
   (d) Join i and j to a new node with lengths $(D_{ij} + u_i - u_j)/2$ to node i and $(D_{ij} + u_j - u_i)/2$ to node j.
   (e) Compute the distance to the new node (ij) and the other groups as
   $$D_{(ij),k} = \frac{D_{ik} + D_{jk} - D_{ij}}{2}$$
   (f) Delete columns and rows corresponding to i and j and add one for (ij). If there are four or more groups left, go back to the first step. Otherwise, connect the three remaining nodes to a single new node using the equation from step (e) to determine the final three edge lengths from the remaining groups to a last internal node.

Maximum Likelihood

1. Know the differences among distance methods, maximum parsimony, and maximum likelihood in terms of the computational effort needed to evaluate trees and to search for trees.
2. Know how to apply AIC to select the best model (among those examined) for maximum likelihood.

Bootstrap

1. Be able to describe how the bootstrap would be applied (conceptually) for parsimony, likelihood, and distance methods.
2. Understand what a bootstrap proportion of, say, 0.90 means.

Bayesian Inference

1. Be able to calculate a Bayesian posterior distribution for a problem similar to the homework.
2. Know in principle how Bayesian inference differs from maximum likelihood.