The second midterm (Nov 26 in-class) covers in-class material days 8-21, labs 5-8, reading weeks 5-11. You may bring a 1 page (front and back), hand-written study guide, but no other notes or resources. You will not need a calculator. I have put vocab in blue.

1. Phylogenetic Trees
   - Study vocab from the end of class 8 + idea of pairwise differences
   - What is a phylogenetic tree and what can we learn from them?
   - What is the input and output of a phylogenetic tree algorithm? Input: dissimilarity map, Output: tree topology (often binary) AND branch lengths
   - Vocab: tree metric (usually induced), ultrametric, rooted, unrooted
   - What are we trying to minimize with phylogenetic tree algorithms?
   - UPGMA algorithm, how to run it and interpret the results
   - Neighbor Joining algorithm, how to run it and interpret the results (how to root tree?)
   - How do UPGMA and NJ compare? Runtime? Advantages/disadvantages depending on the situation?
2. Ancestral State Reconstruction
   - What is ancestral state reconstruction? What can we learn from it?
   - Multiple mutations at the same site are rare. Could be convergent evolution.
   - Fitch's algorithm (small parsimony): what is the input, method, output, and interpretation
   - Sankoff's algorithm (weighted parsimony): same as above + how is it different from Fitch?
   - Runtime for Fitch and Sankoff
   - Perfect phylogeny: what is the input (data from many sites), what is the goal (yes/no answer, ideally + tree and mutation history)
   - Notation (i.e. O_i) and interpretation (containment, disjoint, etc, what do they mean?)
   - Naive algorithm for perfect phylogeny (check all pairs of sites)
   - Gusfield's algorithm: how to run it and interpret the results + why does it work?
   - Why do we use radix sort? What is the runtime of Gusfield's algorithm?
3. Bayesian Phylogenetics
   - Likelihood framework (i.e. probabilistic) for trees
   - Felsenstein's peeling algorithm (basic ideas, not details)
   - Bootstrapping: how it can be used to resample sites and estimate uncertainty of tree inference
   - MCMC: Markov chain Monte Carlo for moving through tree space (basic ideas, not details)

4. Population Genetics

   - What is population genetics? What changes when we consider a single species?
   - What is recombination and what affect does it have on population genetic analysis?
   - Wright-Fisher model of evolution within a population
   - Notation: N for population size, n for sample size, etc
   - Neutrality assumptions: constant N, random mating, no natural selection
   - Measures of sequence diversity: S, π, and the site frequency spectrum (SFS)
   - Finding a common ancestor and how we use that to derive the coalescent
   - Idea that coalescent times $(T_n, T_{n-1}, \cdots T_2)$ are exponentially distributed (not details)
   - Tajima's d: how to compute it and why we expect it to be 0 under neutrality
   - How to interpret π and Tajima's d in terms of deviations from neutrality (e.g. selection)
   - How deep learning models (CNNs in particular) can be used to identify regions under selection

5. Hidden Markov Models

   - What is a Markov chain? What are transition probabilities? Stationary distribution?
   - Difference between a state diagram and a state sequence for a Markov process
   - Probability concepts: conditional probability, probabilities "sum to 1"
   - What is a hidden Markov model (HMM)? Observed sequence $\vec{x}$, hidden state sequence $\vec{z}$
   - Transition, emission, and initial state probabilities (notation, meaning, etc)
   - Viterbi algorithm: input, method (fill in recursive data structure + backtrace to get best path), output and interpretation
   - Forward and Backward algorithms and how we use them to get the posterior decoding and posterior mean
   - Parameter estimation for HMMs when the state sequence is known
   - Baum-Welch algorithm for parameter estimation when the state sequence is unknown (basic idea, not details)
   - HMM example in genetics: time to most recent common ancestor (TMRCA) for $n = 2$
   - Skip: details of log-space (just know why we need to use it), and formulas for A_{kl} and $E_k(b)$ in Baum-Welch