Population Genetics
Problem Set 10

1. What are the similarities and differences between the multi-regional and the "out of Africa" hypotheses of modern human origins?

Both hypotheses agree that the human genus left Africa ca 2 million years ago and that modern humans first evolved in Africa ca 200,000 years ago. The **multi-regional** hypothesis argues that modern humans also evolved outside Africa with genetic contributions from local pre-existing populations of the genus *Homo* (e.g., Neanderthals contributed to the modern human lineage in Europe; "Java Man" contributed in Australasia, etc.). The **out-of-Africa** hypothesis argues that when modern humans left Africa, they replaced the previous non-African regional populations, which did not contribute to the modern human gene pool. Genetic evidence overwhelmingly favors the out-of-Africa theory.

2. What are the similarities and differences between a cladogram, a phylogram and a chronogram?

All three are types of trees show the relationships that exist among the taxa or OTUs; all three may be either unrooted, or rooted using an outgroup; and all three can be redrawn by rotating branches at the internal nodes without affecting their topology. In a **cladogram**, the branch lengths are not informative, and only the evolutionary relationships are shown with no indication of the degree of difference separating the OTUs. By contrast, a **phylogram** has branch lengths that correspond to the degree of <u>evolutionary change</u> that occurred along each branch. A **chronogram** is a tree where the branch lengths represent the amount of <u>time</u> (eg millions of years) that passed since the taxa diverged.

3. What is the difference between a hard and a soft polytomy on a tree?

A polytomy is an unresolved relationship on a tree, represented by a node connected to more than three branches (i.e., one "ancestor" gives rise to more than two "descendant" branches). A "hard" polytomy is one that represents the true evolutionary history of the lineages (e.g., one ancestral species diverged into three descendant species simultaneously). A "soft" polytomy is an unresolved relationship due to lack of data; additional data may resolve the branching pattern.

4. What are the similarities and differences between monophyly, paraphyly and polyphyly?

All three terms refer to the relationships that individual OTUs of a species or taxon or group have relative to the topology of a tree. A taxon is **monophyletic** if all OTUs of that taxon are more closely related to each other than to the OTUs of other taxa on the tree (example: all chimpanzee mtDNA sequences would form a clade that excluded any gorilla mtDNA sequences on a tree, thus chimpanzees are monophyletic in terms of the mtDNA tree). A taxon is **paraphyletic** if its members are all descended from a common ancestor on the tree, yet some of the descendant lineages are excluded (for example, living reptiles all share descent

from a common ancestral reptile; but the class Reptilia excludes birds, which share the same ancestor and thus form part of the same clade). A **polyphyletic** group is derived from two or more different ancestors, and would thus represent an <u>erroneous</u> category for purposes of taxonomy—examples might include "flying animals", "warm-blooded animals", "shellfish", "weeds", "marine mammals", etc.

5. What characteristics is one searching for in choosing a distance correction method?

Distance correction methods, such as Jukes-Cantor or Kimura-2-Parameter, balance **accuracy** (choosing the right tree, helped by increasing parameters) with **precision** (exclusion of wrong trees; helped by reducing parameters). (The parameters take into account characteristics of the sequence data such as transition-tranversion ratio; unequal base frequencies, etc.) A likelihood test finds the model of evolution that best fits the data while using a minimum number of parameters.

6. Give the best definition for the following terms:

a. Phenetics: Grouping species according to overall similarity of observable phenotypic traits, without regards to their evolutionary relationships

b. Cladistics: the hierarchical classification of species based on evolutionary ancestry

c. True tree: a tree that represents evolutionary events as they actually occurred (as opposed to an inferred tree that is generated by a phylogenetic program, which may or may not be **consistent** with the true tree).

d. Operational taxonomic unit, **OTU**: an individual or taxon or sequence at a terminal node of a tree

e. Internal node of a tree: the point on a tree where two or more branches come together, representing the hypothetical ancestors of the OTUs.

f. Root of a tree: the hypothetical ancestor of all the ingroup taxa of a tree; the point where all branches of the ingroup taxa coalesce. A rooted tree has polarity (directionality of evolutionary change from the root to the OTUs), an **unrooted tree does not have directionality**.

g. Outgroup: A sequence from a divergent species or taxon used to root a tree of sequences from another species or taxon. For example, a set of chimpanzee sequences could be rooted using a human or a gorilla sequence as an outgroup.

h. Tree topology: another name for the shape of a phylogenetic tree.

i. Plesiomorphy: an ancestral or unchanged character state retained in a descendant taxon or O.T.U.

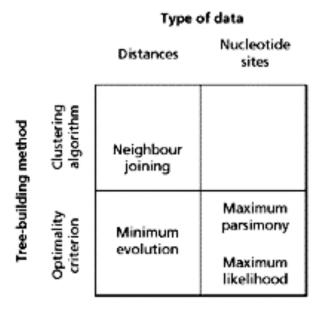j. Apomorphy: a derived sequence or character state, opposite of plesiomorphic

k. **Aut**apomorphy: an apomorphic character state that is unique to **one** taxon or O.T.U. (and thus is not parsimony-informative); **as opposed to a syn**apomorphy, which is an apomorphic character shared by more than one taxon or O.T.U. (and thus can be parsimony-informative for grouping the taxa in a tree).

l. Homology: characters are similar and derived from the ancestral state directly

m. Homoplasy: characters are similar, but their states have evolved independently, due to **convergent** or **parallel evolution**.

n. Lineage sorting: the fixation of ancestral polymorphisms following the phylogenetic divergence of species. Lineage sorting may occur in a way that leads the topology of some **gene trees** to be incongruent with the topology of the **species tree**.

o. Dot plot: a graphical tool for comparing and aligning two sequences

p. Heuristic method: A shortcut used by phylogenetic methods that rely on optimality criteria, for finding the best tree without examining or scoring every possible tree.

q. Long branch attraction: a flaw in the maximum parsimony phylogenetic method, in which two long branches may be wrongly grouped into the same clade.

r. Pseudoreplicates: The novel datasets generated to calculate **bootstrap** support, in which a new dataset of the same length as the original sequences is generated by selecting nucleotide site positions from the original dataset at random, with some of the original sites randomly represented more than once and other sites not represented at all. Each of $\geq$100 pseudoreplicates is used to generate a new phylogeny: the proportion of times that a clade of OTUs appears in these new phylogenies is the bootstrap support for that clade.

7. What are two ways to classify phylogenetic methods; name four commonly used methods and where they fall into these two classifications.

1. Type of data: Discrete vs. distance methods. **discrete** or nucleotide site methods use the actual nucleotide character states; **distance** methods compare sequences to create a distance matrix of differences, with the distances used to generate the tree.
2. Tree building: Optimality criterion vs. clustering algorithm. **Optimality criterion** methods examine every possible tree and give each a score, then select the optimal tree with the best score based on a given criterion. **Clustering algorithm** methods use a defined series of steps in a computer algorithm to generate a single tree.
Four methods are shown at the right

Tree-building method		Type of data	
		Distances	Nucleotide sites
Clustering algorithm		Neighbour joining	
Optimality criterion		Minimum evolution	Maximum parsimony
			Maximum likelihood

8. Which of the following generally apply to the following methods: Neighbor Joining (NJ), minimum evolution (ME), maximum parsimony (MP) or maximum Likelihood (ML)? Write NJ, ME, MP and/or ML if the phrase applies.

May be used to produce a cladogram or phylogram   NJ ME MP ML

Bootstrap support may be calculated using pseudoreplicates   NJ ME MP ML

May be used to analyze either mtDNA or nuclear sequences   NJ ME MP ML

May be used to examine if clades are reciprocally monophyletic   NJ ME MP ML

Can handle outgroup sequences   NJ ME MP ML

Requires an accurate alignment of sequences   NJ ME MP ML

May have internal and external nodes   NJ ME MP ML

Places operational taxonomic units at external nodes   NJ ME MP ML

May be used to draw a tree with rectangular shape   NJ ME MP ML

May be rooted or unrooted   NJ ME MP ML

Generates a distance matrix   NJ ME

Calculates the probability of the data given a hypothesis   ML

Of the 4 methods listed, considered the most computationally intensive ML

May be run using the Jukes-Cantor model   NJ ME ML

May be run using the Kimura-2-Parameter model   NJ ME ML

Ideally tries to examine all possible trees   ME MP ML

May use a heuristic search across tree space   ME MP ML

May generate more than one optimal tree   ME MP ML

May produce a tree with a polytomy   ME MP ML

Uses a clustering algorithm to produce a single resolved tree   NJ

Of the 4 methods listed, the least computationally intensive   NJ

Guaranteed to generate the true evolutionary history for the sequences none