# Assignment #4: Distance Matrix Methods

## Part I. Distance methods using paper and pencil

The alignment below shows the first 30 bp of the mitochondrial cytochrome b gene for five mammals.

```
Dog          ATG ACC AAC ATT CGA AAA ACC CAC CCA CTA
Cat          ATG ACC AAC ATT CGA AAA TCA CAC CCC CTT
Mouse        ATG ACA AAC ATA CGA AAA ACA CAC CCA TTA
Pig          ATG ACC AAC ATC CGA AAA TCA CAC CCA CTA
Human        ATG ACC CCA ATA CGC AAA ACT AAC CCC CTA
```

### 1. Distance calculation

1. **Find the Jukes-Cantor distance for each pair of species** (I talked briefly about different models of DNA evolution on Thursday, but here is the formula for the JC distance again: `d = –3/4*ln(1–(4/3)p)` , were p is uncorrected (observed) dissimilarity.)

### 2. UPGMA tree

1. Apply the algorithm we used in class to find the UPGMA tree based on JC distances. **Draw this tree and indicate all branch lengths.**

### 3. Neighbor-joining tree

1. Apply the neighbor-joining (NJ) algorithm to find the NJ tree. **Draw this tree and indicate all branch lengths.**

## Part II. Distance methods in PHYLIP.

You may want to check the [PHYLIP documentation](#) and/or [lab3](#) tutorial before doing this part.

Create a multiple sequence alignment using prestin nucleotide sequences we used in class and save it in the PHYLIP format.

1. Calculate 4 different matrices using 4 models available in dnadist and perform a NJ analysis with each of them. Do not submit these trees. Instead, calculate a strict consensus tree. **Root your consensus trees using an appropriate outgroup and include it with the rest of the assignment. What does this tree tell you about the influence of model choice on the results of this analysis?**

2. Create 200 bootsrtap replicates of your data file and build a NJ tree for each of them (you should choose one of the four models of nucleotide substitutions you used in #4). **Build a majority rule consensus tree for these trees, print it out, and submit with the rest of the assignment. What do bootstrap support values tell you about the robustness of your phylogenetic inference?**

# Part III. Extra pointFastME.

1. Use one of the distance matrices you calculated in #4 to calculate NJ trees with/without the tree refinement with NNI and SPR. **Did NNI/SPR post-processing changed the resulting tree phylogeny?**

# Good luck!