Lab Exercises – Phylogeny

Exercise 1: Reconstruct a phylogeny using morphological characters

For this exercise, you will measure morphological characters of mammal skulls. Complete the following measurements and fill out the table below. Note that some characters might be difficult to score. This is fine. In fact, we expect it. It is reflective of the challenges that real biologists face. Do your best to score each character.

1. Score the number of incisors on one side of the upper jaw. For example, the human skull has 2 incisors on one side of the upper jaw. Write the number in the table below. If there are no incisors, write 0.

2. Does the skull have a canine tooth? Score 1 for a canine tooth and 0 for no canine tooth.

3. Count the number of post-canine teeth on one side of the upper jaw. If there is no canine, count the post-incisor teeth.

4. Score the total number of incisive foramina. The incisive foramina (singular, foramen) are openings found on the palate just behind the incisors. They allow blood vessels and nerves to pass between the oral and nasal cavities.

5. Does the skull have a sagittal crest? The sagittal crest is a bone ridge running lengthwise along the midline on the top of the skull and is an attachment point for the jaw muscles. It may be present along the whole midline or just on the posterior aspect of the skull. Score 1 for yes or 0 for no.

6. Measure the width of the foramen magnum. This is the largest of the openings at the base of the skull. The spinal cord passes through this foramen. Compare the width of the foramen magnum to the width of the skull at the point of the foramen magnum. Does the foramen magnum occupy more than 1/4 the length of the skull? Score 1 if yes and 0 if no.

7. Is the orbit open or closed? Score 1 for closed and 0 for open.

8. Identify 1 trait that varies between skulls and score it in the last column.
    a. This can be the presence or absence of a character (such as a sagittal crest) OR it can be a measurement.
    b. If you decide to measure a feature of the skull, make sure that you are measuring a ratio, not an absolute measurement (for example, width of foramen magnum divided by width of skull). This is because the skulls are different sizes and taking a single measurement is not informative.

	# incisors	Canine tooth	# post-canine teeth	# Incisive foramina	Sagittal crest	width of foramen magnum	Orbit closed	Your favorite trait	Family
Canis latrans									
Homo sapiens									
Pan troglodytes									
Vulpes vulpes									
Lynx rufus									
Ovis aries									
Pecari tajacu									
Didelphis virginiana									
Castor canadensis									
Oryctolagus cuniculus									

Using the data table above, construct the best phylogenetic tree possible. Note that you do not have to recover the "correct" tree.

1. Which skull did you use as your outgroup? Why?

2. Are there any characters that you scored that you suspect are homoplasies? If so, which ones?

3. For characters that may be homoplasies, are there common selection pressures that could explain them?

Exercise 2: Analyze relatedness using sequence divergence

Modern phylogenetics relies almost exclusively on molecular data to reconstruct phylogenies. Protein or nucleotide sequences are used. We can consider each nucleotide or amino acid to be a character, which vastly increases the amount of information available to perform phylogenetic analysis. As with morphological characters, it is important to carefully select the sequence(s) to be used.

Important considerations include:

- How quickly is the gene evolving? For resolving ancient nodes between distantly related taxa, we would want to use something highly conserved that evolves very slowly, such as 16S/18S ribosomal DNA. For closely related taxa, we may wish to use a region of DNA that does not code for a protein, because it is (presumably) free to vary and will change much more quickly.
- Is the gene under strong selection pressure? This is related to the previous point.
- Do you want to use mitochondrial DNA (mtDNA) or genomic DNA or both? mtDNA is an excellent choice when you want to eliminate the genetic variation due to recombination during meiosis. Since mtDNA does not recombine, all variation is due solely to mutation.
- How many genes do you want to sequence? More is usually better, but practical considerations will limit the actual number.
- What will you use for your outgroup? If the outgroup species is too distant, it will give you an incorrect tree.

Once you have selected a group of taxa and one or more molecular markers, you must also select a method of phylogenetic analysis. This can range from simple comparison of sequence identity or similarity, to incredibly complex statistical analysis, such as maximum likelihood or Bayesian inference. Studying the algorithms that are used to build phylogenetic trees is outside the scope of this course. We will do a couple of very simple analyses and compare them to the morphological tree and to the best current hypothesis of mammalian phylogeny.

The first method is the easiest. We simply measure the number of nucleotides or amino acids that are different between each taxa. For this method, you will be examining 50 amino acids from the cytochrome b protein. This is not the complete sequence; it is one of the more variable regions of the protein. The table below uses the standard one-letter codes for the 20 amino acids.

	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350
Canis latrans	L	A	F	V	P	F	L	H	T	S	K	Q	R	S	M	M	F	R	P	L	S	Q	C	L	F	W	L	L	V	A	D	L	L	T	L	T	W	I	G	G	Q	P	V	E	H	P	F	I	I	I
Homo sapiens	L	A	M	I	P	I	L	H	M	S	K	Q	Q	S	M	M	F	R	P	L	S	Q	S	L	Y	W	L	L	A	A	D	L	L	I	L	T	W	I	G	G	Q	P	V	S	Y	P	F	T	I	I
Pan troglodytes	L	A	A	I	P	V	L	H	T	S	K	Q	Q	S	M	M	F	R	P	L	S	Q	L	L	Y	W	L	L	A	T	D	L	L	I	L	T	W	I	G	G	Q	P	V	S	Y	P	F	I	T	I
Vulpes vulpes	L	A	L	I	P	H	L	H	T	S	K	Q	R	G	M	M	F	R	P	L	S	Q	C	L	F	W	L	L	T	A	D	L	L	T	L	T	W	V	G	G	Q	P	V	E	H	P	F	I	I	I
Lynx rufus	L	A	I	I	P	I	L	H	T	S	K	Q	R	G	M	M	F	R	P	L	G	Q	C	L	F	W	L	L	V	A	D	L	L	T	L	T	W	I	G	G	Q	P	V	E	Y	P	F	I	T	I
Ovis aries	L	V	I	M	P	L	L	H	T	S	K	Q	R	S	M	M	F	R	P	I	S	Q	C	M	F	W	I	L	V	A	D	L	L	T	L	T	W	I	G	G	Q	P	V	E	H	P	Y	I	I	I
Pecari tajacu	L	A	L	V	P	A	L	H	T	S	K	Q	R	S	M	M	F	R	P	L	S	Q	L	L	F	W	M	L	V	A	D	F	L	T	L	T	W	I	G	S	Q	P	V	E	H	P	F	I	I	I
Didelphis virginiana	L	L	I	I	P	M	L	H	T	S	T	Q	R	S	M	A	F	R	P	I	S	Q	T	L	F	W	M	L	T	A	N	L	I	I	L	T	W	I	G	G	Q	P	V	E	Q	P	Y	I	T	I
Castor canadensis	L	A	F	I	P	F	L	H	M	S	K	Q	R	S	M	M	F	R	P	I	S	Q	V	L	F	W	V	L	V	A	D	L	L	T	L	T	W	I	G	G	Q	P	V	E	H	P	F	I	T	I
O.cuniculus	L	M	A	F	P	L	L	H	T	S	K	Q	R	S	L	M	F	R	P	I	S	Q	C	L	F	W	A	L	A	A	D	L	L	T	L	T	W	I	G	G	Q	P	V	E	P	P	F	I	I	I

1. How many of the characters above are uninformative? A character is uninformative if it is exactly the same for every taxon. (Hint: use a sharpie and black out each column in which every amino acid is exactly the same. It will make the remaining questions easier.)

2. Which residue (i.e., amino acid #) is the least conserved? (this will be the column with the greatest number of different amino acids)

3. Count the number of differences between each pair of sequences. This will take some time, so divide up the work between the whole class and trade answers. Fill in the table below. The first three have been filled in for you.

4. Look at the table you have filled out. Which two species are the most closely related? (these will have the least number of differences. If it is a tie, write both pairs of species).

5. Which species is the most distantly related to the others? (this will be the species that, on average, has the most number of differences).

	Homo sapiens	*Pan troglodytes*	*Vulpes vulpes*	*Lynx rufus*	*Ovis aries*	*Pecari tajacu*	*Didelphis virginiana*	*Castor canadensis*	*Oryctolagus cuniculus*
Canis latrans	12	12							
Homo sapiens		7							
Pan troglodytes									
Vulpes vulpes									
Lynx rufus									
Ovis aries									
Pecari tajacu									
Didelphis virginiana									
Castor canadensis									

Exercise 3: Reconstruct a phylogeny using molecular characters

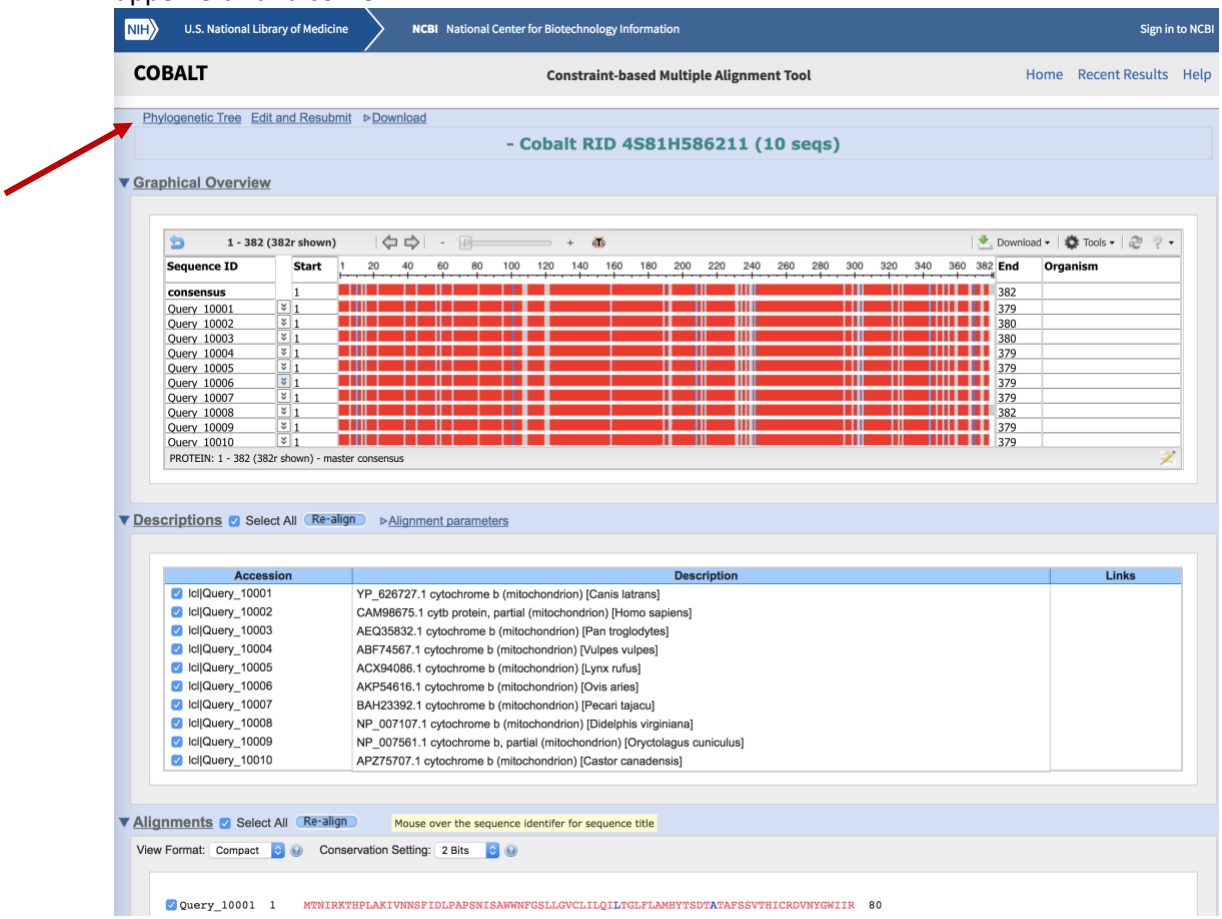
1. Find the files on Canvas that contain the complete protein sequence for cytochrome b: MammalCYTB.fasta and MammalCYTB.docx

2. Open a browser and load the COBALT tool on the NCBI website. https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi

3. Load the sequence data. There are two options to do this:

    a. **Either** download the .fasta file to your computer and then click <Choose File> and select the file from your computer.
    b. **Or** open the .docx file. Select all and copy, then paste in the box that says "**Enter at least 2 protein accessions, gis, or FASTA sequences"**

4. Click <ALIGN>.
5. After a few seconds, you should see a screen that looks like this. Click <Phylogenetic Tree> in the upper left-hand corner.



6. Sketch the tree that is produced. You do not need to include the accession numbers or gene name on the tips of the branches. Only the scientific name of the organism is needed (e.g., *Homo sapiens*).

7. How does this tree compare to the tree you drew using skull morphological characters?

8. Why do you think these two trees are different?

9. Which tree do you think is closer to the "true" phylogeny of these mammals?

10. What is the function of cytochrome b?

11. Now look back at the table of amino acids in Exercise 2. Find a position that is a **homoplasy** between two species**. (**Hint: you are looking for an amino acid position that is shared between two species that are not closely related, but not by more closely related species).