

Lab XI: Molecular Phylogenetics

Objectives: Use mitochondrial DNA sequences to examine the evolutionary relationships among primates.

Mitochondrial DNA

Mitochondria have their own genomes, which consist of small (~15 kb) circular DNA molecules that encode proteins, ribosomal RNAs and transfer RNAs used by mitochondria. Mitochondrial gene sequences are useful for investigating evolutionary relationships among closely related species because they evolve rapidly. Rapid evolution means there are enough sequence differences between species for statistical analysis of these differences. There are also a large number of mitochondrial gene sequences available in public databases. Today you will download sequences for a variety of primate species (including humans), and use these sequences to explore evolutionary relationships among the primates.

Molecular Phylogenetics

The past 25 years has seen the development of sophisticated methods to determine evolutionary relationships (phylogenies) from DNA and amino acid sequences. In many cases, these methods have confirmed traditional classifications based on phenotypes. However, in other cases molecular methods have produced very surprising results. One of the most exciting areas in molecular phylogenetics has been the analysis of human evolutionary history. For example, it had long been debated whether the closest relatives of humans were chimpanzees or gorillas. Today you will see how molecular data can answer this and other questions about primate evolution.

Part I. Build a primate phylogeny with sequences of *Cytochrome Oxidase I*

Step 1: Download lab11.zip

1. Download **lab11.zip** from the class website (<http://seahorse.louisiana.edu/biol224/>) into your **My Documents** folder.
2. Extract lab11.zip, it should create a lab 11 subfolder, with two files (rename.pl and names.txt).

*Step 2: Download mitochondrial genome sequences for primates from the **Organellar Genome Retrieval System (OGRe)** database.*

1. Go to the OGRe website at <http://drake.physics.mcmaster.ca/ogre/>.
2. On the left side of the page, click on **select species from taxonomy**
3. From the list, click on **Chordata**, then **Vertebrate**, then **Mammalia**, then **Eutheria**
4. Click on the small white square next to **Primates**, a check mark should appear.
5. Click on **Display Sequences** at the bottom of the page.
6. Select **COX1** by clicking on it.
7. Click on **download**, and then **OK**. Save the sequences in the **lab11** subfolder in your **My Documents** folder as **primates.fasta**.

Step 3: Edit the sequences file so that it can be analyzed.

The **primates.fasta** file has extra information at the top that we can eliminate.

1. Run **Komodo Edit** from the **Start Menu** ⇒ **Programs** ⇒ **Biol 559** ⇒ **Komodo Edit**
2. From the File menu of **Komodo Edit**, select **Open** ⇒ **File**, then select **primates.fasta** to open.
3. You will see that the file has two parts. First there is a list of the sequences in the file, with a name for the sequence by a colon, the species scientific and common names, another colon, and **MITOCHONDRION**. The second part consists of sequences themselves. Each sequence is preceded by a line that starts with > followed by the name of the sequence. This is known as the **fasta** format for sequences.
4. Use Komodo Edit to delete the first part of the file, (shown here in a box) so that the first line starts with: >**NYCCOUMIT(COX1)**
5. Save the file as **CO1.tmp** in the **lab11** folder.

```
#####
# OGRE sequences #
#####

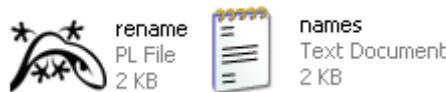
#NYCCOUMIT : _Nycticebus coucang_ (slow loris) : MITOCHONDRION
#PYGNEMHIT : _Pygathrix nemaeus_ (Douc langur) : MITOCHONDRION
#PAPHAMHIT : _Papio hamadryas_ (baboon) : MITOCHONDRION
#PROBADMIT : _Procolobus badius_ (red colobus) : MITOCHONDRION
#GORGORMIT : _Gorilla gorilla_ (gorilla) : MITOCHONDRION
#DAUHADMIT : _Daubentonia madagascariensis_ (aye-aye) : MITOCHONDRION
#PREHELMIT : _Presbytis melalophos_ (mitred leaf monkey) : MITOCHONDRION
#COMPYHMIT : _Pongo pygmaeus abelii_ (Sumatran orangutan) : MITOCHONDRION
#CHLPHYHMIT : _Chlorocebus pygerythrus_ (Cercopithecus pygerythrus) : MITOCHONDRION
#MACSVLHMIT : _Macaca sylvanus_ (Barbary ape) : MITOCHONDRION
#CERAEHMIT : _Cercopithecus aethiops_ (African green monkey) : MITOCHONDRION
#HYLLARHMIT : _Hylobates lar_ (common gibbon) : MITOCHONDRION
#CERALARHMIT : _Cebus albifrons_ (white-fronted capuchin) : MITOCHONDRION
#CERASARHMIT : _Cercopithecus sabaues_ (green monkey) : MITOCHONDRION
#PYGROKMIT : _Pygathrix roxellana_ (golden snub-nosed monkey) : MITOCHONDRION
#TRAOBSHMIT : _Trachypithecus obscurus_ (dusky leaf monkey) : MITOCHONDRION
#HOMASPHMIT : _Homo sapiens_ (human) : MITOCHONDRION
#TARBANHMIT : _Tarsius bancanus_ (western tarsier) : MITOCHONDRION
#PANTRONHMIT : _Pan troglodytes_ (chimpanzee) : MITOCHONDRION
#COLGUEHMIT : _Colobus guereza_ (guereza) : MITOCHONDRION
#NASLARHMIT : _Nasalis larvatus_ (proboscis monkey) : MITOCHONDRION
#EULMONHMIT : _Eulemur mongoz_ (mongoose lemur) : MITOCHONDRION
#SEMENHMIT : _Sennopithecus entellus_ (Hanuman langur) : MITOCHONDRION
#CHLCATHMIT : _Chlorocebus tantalus_ (Cercopithecus tantalus) : MITOCHONDRION
#LEMCATHMIT : _Lemur catta_ (ring-tailed lemur) : MITOCHONDRION
#PONPYHMIT : _Pongo pygmaeus_ (orangutan) : MITOCHONDRION
#PANPANHMIT : _Pan paniscus_ (pygmy chimpanzee) : MITOCHONDRION
#MACMULHMIT : _Macaca mulatta_ (rhesus monkey) : MITOCHONDRION

>NYCCOUMIT (COX1)
atgttcattaaccgctggctctattctaccatcacaaggacattggcactttatacttagtgttcggctgcctgagcgg
ggatagtggaaccgccctaagtttgtaaacccgagcagagctaggtcagccaggctactctgctggcggatgacaaat
ctataacgttatgtaacagcccatgattctgttataactttttatagttataccaataaactcggggcttcggg
aactgattagtagccccaataaattggagccccgacatagctttcccacgaataaataataaagtcttgcactttac
caccctcttccctcttccctacccctctctatgtaaaccccaaccccaaccccaaccccaaccccaaccccaaccccaac
```

Step 4: Rename the sequences

The names of the sequences (like NYCCOUMIT) are based on genus and species names, but are not very useful. Therefore, rename the sequences with the full scientific names of the species they represent. You could do this by hand (e.g. in Komodo Edit), or you do this with a *script* in the Perl language.

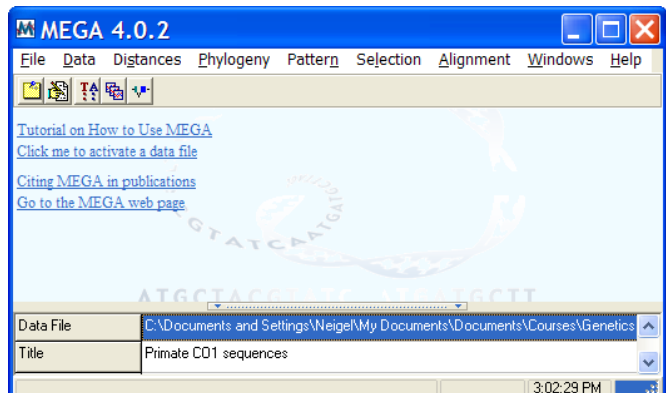
1. Make sure the Perl script (called **rename.pl**), and the file with the scientific names (called **names.txt**), are all in the same folder as **CO1.tmp**. This should be the **lab11** subfolder in your **My Documents** folder. The icons for these files should look like these:



2. Double-click on **rename.pl** to run it. You can choose the default names by pushing **Enter**. Name the output file, enter **CO1**. If everything works, you should have a new file called **CO1.fas** in your folder.

Step 5: Align the sequences in MEGA 4

1. Double click on the file **CO1.fas**, MEGA should start and load the file. Two windows should open: the **Alignment Explorer** and the main **MEGA** window (which may be under the Alignment Explorer window).
2. In Alignment Explorer, select **Alignment** ⇒ **Align by ClustalW**. This will take a few minutes. The sequences were almost aligned to begin with, so the only changes will be at the ends of sequences.
3. Select **Data** ⇒ **Export Alignment** ⇒ **MEGA format**. Enter **CO1.meg** as the **File name**, and click **Save**. For *Input title of the data*, enter **Primate CO1 sequences**, then click **OK**. For *Protein-coding nucleotide sequence data?* Click **Yes**.
4. Close Alignment Explorer; a small window asks *Open the data file in MEGA?* Click **Yes**.
5. The Sequence Data Explorer window will now open. You don't need this, so close the

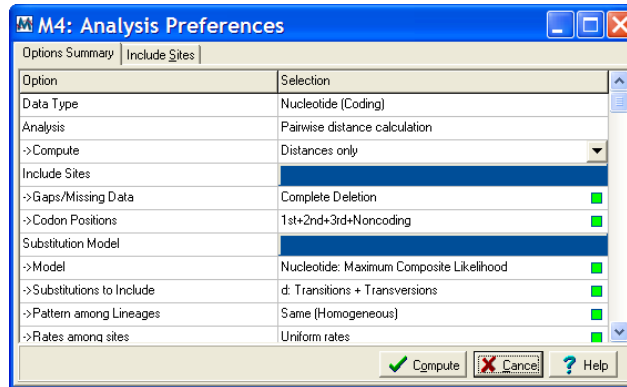


window. You should now see just the main MEGA 4.0.2 window, like the one shown here.

Step 6: Calculate sequence distances

A sequence distance is an estimate of how much change has occurred between two sequences that evolved from the same ancestral sequence. Because multiple changes can occur at the same position in a sequence, sequence distances are generally greater than simply the proportion of nucleotides that are different between two sequences.

1. Select **Distances** ⇒ **Compute Pairwise**. A window will appear showing the options for how the distances will be calculated. The default values are appropriate (as shown below), so just click **Compute**.



2. A **Pairwise Distances** Window showing genetic distances between all the sequences should open.
3. In the **Pairwise Distances** window select **Average** ⇒ **Overall**. Answer the first lab report question.
4. Find the distances between human (*Homo sapiens*) and chimpanzee (*Pan troglodytes*) and between human (*Homo sapiens*) and gorilla (*Gorilla gorilla*). Answer lab report question 2.
5. Close the Pairwise Distances Window.

Step 7: Build a tree

1. From the **Phylogeny** Menu select **Construct Phylogeny** ⇒ **Neighbor Joining**
2. In the Analysis Preferences Window, click on **Compute**. A window showing a phylogenetic tree should appear. Look at this tree, noting the relationships of humans to other primates.


Step 8: The bootstrap statistic

The phylogeny tree that you just produced is only an *estimate* of the true phylogeny. Like all estimates, it is subject to error, and it would be useful to have some idea of how reliable it is. The bootstrap statistic is one way that the reliability of a phylogeny can be assessed. It works by taking random samples of portions of the original DNA sequences, and constructing a phylogeny for each sample. It then compares the trees from all of these samples, and for each branch in the tree determines the percentage of trees that had branch. A bootstrap value of 100 means that all of the trees had the branch (high reliability), while bootstrap values below 50 mean that less than half of the trees had the branch (low reliability).

1. Select **Phylogeny** ⇒ **Bootstrap Test of Phylogeny** ⇒ Neighbor Joining. The default values are appropriate, so click **Compute**. A new TreeExplorer window will open, with bootstrap values shown on the branches. Branches without numbers had low bootstrap values. Keep this window open for the next step.

Step 9: Rooting the tree

Phylogenetic trees can be either **rooted** or **unrooted**. In a rooted tree, the main “trunk” of the tree is the most ancient part of the tree and taxa that are on the same branch are more closely related than taxa on different branches. The root of a tree is most often determined by using an **outgroup**, which is a taxon or group of taxa that are known to be the most distantly related to the other taxa in the tree.

1. The phylogenetic trees that you have generated are **unrooted trees**, although they appear to be rooted because of the way they are drawn. To see the tree as it truly is, select **View** ⇒ **Tree / Branch Style** ⇒ **Radiation**. Go back to the original by selecting **View** ⇒ **Tree / Branch Style** ⇒ Traditional ⇒ Rectangular.
2. On the left edge of the screen is a button that looks like this:  with the arrowhead in green. This button allows you to place the root of the tree on any branch. Click on this button, and try placing the root on the branch that leads to *Homo sapiens*. The tree now looks very different, it appears that all of the other primates are more related to each other than any are to humans. This demonstrates how important the placement of the root is before a phylogenetic tree can be interpreted.
3. There are two suborders in the order Primates: Strepsirrhini and Haplorrhini. The Strepsirrhini include most of the “Prosimian” primates: lemurs, lorises, pottos, galagos and the Aye-Aye. Use the Strepsirrhini as your outgroup for rooting the tree. In your tree there are four representatives of the Strepsirrhini. Use the appendix to determine which species they are, and answer lab report question 3.
4. Place the root of the phylogeny on the branch that leads to all of the Strepsirrhini. You will probably find that you cannot get a group that includes the non-Strepsirrhini prosimian, *Tarsius bancanus*. That’s OK, let it be in the outgroup as well.

Step 10: Interpreting the Phylogeny

You now have a phylogeny that is rooted and tested with the bootstrap. Use the information in the appendix to answer lab report questions 4-6.

Step 11: Option 1: Another Gene

Option 1: Download another gene (try ATP6 or CYTB) and repeat steps 2, 3, 4, 5, 8 and 9. From TreeExplorer, select **Image** ⇒ **Save as TIFF file**. Answer question 7 on the lab report.

Step 11: Option 2: Another Group of Animals

There are two kinds of panda: the Giant Panda (*Ailuropoda melanoleuca*) and the Red Panda (*Ailurus fulgens*). Both are incredibly cute and share several traits that are clearly adaptations to feeding on bamboo. What group of mammals are pandas related to? Although they eat mostly bamboo, they are clearly in the mammalian order Carnivora. In some ways they are similar to raccoons (family Procyonidae), and in other ways they are similar to bears (family Ursidae). Their classification was long debated, until it was settled by DNA analysis.

1. Go back to the OGRE website at <http://drake.physics.mcmaster.ca/ogre/>.
2. Select **Chordata** ⇒ **Vertebrate** ⇒ **Mammalia** ⇒ **Eutheria**
4. Click on the small white square next to **CARNIVORA**, a check mark should appear.
5. Click on **Display Sequences** at the bottom of the page.
6. Select **COX1** by clicking on it.
7. Click on **download**, and then **OK**. Save the sequences in the **lab11** subfolder in your **My Documents** folder as **carnivores.fasta**.
8. No need to edit this file or align the sequences, just run **rename.pl**, using **names.txt** and **carnivores.fasta** as input files and name the output file **carnivores.fas**.
9. Open **carnivores.fas** in Mega, convert the file to **carnivores.meg**, and build a phylogeny as you did for the primates.
10. Answer question 8 on the lab report.

Appendix: Classification of primate species used for this lab

suborder Haplorrhini (tarsiers, monkeys and apes)

infraorder Simiiformes

parvorder Catarrhini

superfamily Hominoidea

family Hominidae (hominids, or “great apes”)

subfamily Homininae

genus *Pan* (chimpanzees)

genus *Homo* (humans)

genus *Gorilla* (gorillas)

Gorilla beringei

subfamily Ponginae

genus *Pongo* (orangutans)

family Hylobatidae (gibbons, or “lesser apes”)

genus *Hylobates*

superfamily Cercopithecoidea

family Cercopithecidae (Old World Monkeys)

subfamily Cercopithecinae

genus *Macaca* (macaques)

genus *Chlorocebus* (vervet monkeys)

genus *Papio* (baboons)

genus *Cercopithecus* (guenons)

subfamily Colobinae

genus *Colobus* (colubus monkeys)

genus *Procolobus* (olive colubus)

genus *Semnopithecus* (gray langurs)

genus *Nasalis* (proboscis monkey)

genus *Pygathrix* (douc langurs)

genus *Trachypithecus* (lutungs)

genus *Presbytis* (surilis)

parvorder Platyrrhini (New World Monkeys)

family Cebidae (marmosets, tamarins, capuchins and squirrel monkeys)

subfamily Cebinae

genus *Cebus* (capuchin)

infraorder Tarsiiformes

family Tarsiidae (tarsiers)

genus *Tarsius*

Appendix I (continued)

suborder Strepsirrhini (non-tarsier prosimians)

infraorder Lemuriformes

family Lemuridae (lemurs)

genus *Eulemur* (true lemurs)

genus *Lemur* (ring-tailed lemur)

infraorder Chiromyiformes

family Daubentoniidae

genus *Daubentonia* (Aye-aye)

family Lorisidae (lorises and pottos)

genus *Nycticebus* (slow lorises)

Name _____ Section _____

Lab XI: Molecular Phylogenetics - Lab Report

1. What is the average sequence distance between the COX1 primate sequences? _____

2. What is the COX1 sequence distance between:
 - a. human (*Homo sapiens*) and chimpanzee (*Pan troglodytes*)? _____
 - b. human (*Homo sapiens*) and gorilla (*Gorilla gorilla*)? _____
 - c. How do these distances compare with the average distances between primates?

3. Which species are in the suborder Strepsirrhini?

4. Are bootstrap values higher for more recent (shallow) or more ancient (deep) relationships in this phylogeny?

5. *Tarsius bancanus* is incorrectly grouped with the Strepsirrhini. Based on all the evidence you have, would you argue that the classification of *Tarsius bancanus* is incorrect? (Why?).

6. Based on the COX1 phylogeny:
 - a. Which species is most closely related to *Homo sapiens*? _____

 - b. How does the sequence divergence between humans and chimpanzees compare with the sequence divergence between the two species of *Macaca*? _____

 - c. There appear to be two major groups of Haplorrhini. What do these groups correspond to in the classification of primates?

6. (continued)

d. How strong is the support for the subfamily Cercopithecidae?

e. Does the placement of *Cebus albifrons* on the tree correspond with its classification?

7. (Step 11 option 1: Another gene).

a. Which gene did you use?

b. On this tree, what is the closest relative to humans?

c. Are there any important differences between this tree and the one based on COX1? (Consider especially cases where the phylogenies did not match traditional classifications).

8. (Step 11 option 2: Another group of animals)

a. Are pandas more related to bears or raccoons? (Or is there a problem with this question?)

b. Why do you think pandas were more difficult to classify than other carnivores before DNA methods became available?