

Phylogeny using Genomics (Bioinformatics)
Evolution
Texas A&M International University

Purpose: Exercise in cladistics using genomics.

Assignment:

Compare alternate cladograms using DNA sequences.
Remember that this is just an exercise.

Recommended Procedure:

1. Locate a paper that has revised a phylogeny because a taxon is considered paraphyletic or polyphyletic. There are many revisions of phylogenies recently published because of paraphyletic or polyphyletic taxa. A short list of revisions done or being considered includes kingdom Monera to two domains (Bacteria and Archaea), angiosperms revised (dicots paraphyletic), vertebrates revised (bony fish paraphyletic), arthropods considered polyphyletic, nematodes and arthropods closely related, and birds included with dinosaurs and crocodylians (reptiles paraphyletic). Compare the possible paraphyletic (or polyphyletic) cladogram to the revised cladogram in the paper using a short sequence of DNA that you will locate in the gene bank. If you choose a paper that has used DNA sequences from various species for a character(s) in their cladogram, you can use some of the species and a short homologous sequence of the DNA from the paper.

2. Simplify the cladograms to 6-7 taxa (species). One taxa should be an outgroup that is closely related to the taxa of interest but with the taxa of interest being one monophyletic group. The other taxa selected should show the difference in the proposed paraphyletic (or polyphyletic) cladogram vs. the revised cladogram.

3. Select a gene/loci for comparison. Start at National Center for Biotechnology Information

<http://www.ncbi.nih.gov/>

There is useful information about the site and a number of tools to use.

Choose this gene carefully. It should be one that is found in all the species in the exercise and shows variation between the species. Some genes can be too conservative as a character for some cladograms. You should have an homologous sequence for all species being compared but stills varies between species. You will search for differences within the DNA sequence.

There are a number of possible ways to locate the gene for each taxa.

a) Search **TaxBrowser**:

Identify one or more species to work from the paper you have chosen. Search for the species in:

<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>

It is best to use one species as your target species and make sure the other species (or closely related species) occur in the searches you perform.

Then go to the **Entrez Records** and click on **Nucleotide direct links**. Select a potential

gene/locus from the list of nucleotides. Once a potential gene/locus has been found, then click on the **accession number or version** (e.g., CG705225 is the **Accession number** for the first nucleotide listed in the human genome). Other information that can be used is the **gi** (e.g., 37693027). Information will be found in the **Sequence Viewer** including **Locus, Accession Number, Version, Source Organism, gi, Origin** (DNA Sequence for the positive strand - coding strand, 1 will be the 5' end).

Find a DNA sequence from **Origin**. Each base pair only has the base from the positive strand listed.

Select a portion of the sequence between 100-150 bp (base pairs). Select **BLAST** from the NCBI home page (open in a new window):

<http://www.ncbi.nlm.nih.gov/BLAST/>

Click on **Nucleotide-nucleotide BLAST (blastn)**

Copy the DNA sequence from the loci into the **Search** and then click **BLAST**.

Click on **Format** in next window.

There will be a wait for the **Results of BLAST**.

After the results are shown, click on **Taxonomy Reports**. There should be 5 to 6 more species for the cladogram present. Make sure you can fit them into the cladograms you are testing and you have a good choice for the outgroup. Some of the species that could be used in the cladogram may not be in the **Taxonomy Reports** (perhaps the DNA sequencing has not been done yet for that species).

b) Use **Accession Number** from the paper for the 6-7 species selected to build the cladogram. Only a gene/locus from the paper will be used in this case. The other method will show other potential genes to compare with the paper's cladograms.

Consult with the instructor to determine if the choices are suitable.

Report next week on the choice of species and locus

Copy the information for each species.

Include in the information for each species:

Locus/gene studied (**remember it should be the same gene for each species**), species (6-7 spp.), + higher taxa, **Accession Number and version** and **gi** for each species. The **Accession Number** or **gi** will be used for the next step in the exercise.

4. To align the base sequences between the species use **BLAST 2**.

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>

Enter the **Accession Number** or **gi** for two species in **sequence 1** and **2** and click **Align**

Select a sequence that aligns in all species. Copy the sequence for each species to a spread sheet for comparison.

Compare sequences to establish cladogram. Follow the concepts described in class for doing cladistics such as parsimony. Use each base pair as character. The information from each base pair can be:

a) Ancestral homologous - base pair matches in all species (assumed to be ancestral).

b) Change in only one taxa (species) - information not useful to determine phylogeny but has been used as a molecular clock. The number of differences between taxa can indicate the evolution since divergence, and assuming a constant rate of mutation and genetic drift the number of changes can indicate time since divergence.

c) Synapomorphy - change in one clade (branch) of the cladogram.

d) Homoplasy - Same change occurring in more than one clade of the cladogram. This would require two or more changes taking place in the same base pair. With parsimony, the cladogram with the most synapomorphies and the fewest homoplasies would be supported.

Note: Not all homoplasies will be detected but with parsimony homoplasies should be rare.

Report results showing the alternate cladograms (draw the cladograms with species labeled). Which cladogram is supported by your comparison. Draw conclusions: What taxa are a good monophyletic taxa according to your analysis? What taxon (or taxa) is paraphyletic (or polyphyletic) according to your analysis? Justify the conclusions with evidence from the genomics data. Does your results support the revision? Is the old phylogeny paraphyletic or monophyletic according to your results?