



<http://www.geol.umd.edu/>
Tomistoma



<http://www.wildthingsphoto.com/>
Gavia

The intention with BIO332 is that you should have a reasonably good understanding of basic phylogenetic methodology when you have completed the course. It will be a good test of your learning achievements if you can read a recent phylogenetics paper from a journal and actually be able to reproduce some of the analyses conducted by the authors. *Systematic Biology* is one of the most prominent journals in the field of systematics. In this assignment, you will examine a paper by Harshman *et al* (2003) and try to repeat some of their analyses based on their data. The morphological data is provided for you in the file *gavials_morph.nex*. You may execute this file directly in PAUP* to see the accompanying trees. (Respond *yes* to the question about *derooting*.) The file *gavials_mol.nex* contains their aligned nucleotide data. Additionally, at the end of the molecular matrix, there is a series of standard characters representing coded states of what they regarded as phylogenetically informative “indels” (gapped regions) in the alignment.

Your task will be to write scripts to PAUP* and to MrBayes in order to do some of the analyses that Harshman *et al* did on the molecular data. Your scripts and the files resulting from execution of the scripts should be submitted for assessment, together with a short electronic report (word or pdf file) that explains your procedures, results and interpretations.

Your PAUP* script should contain commands that:

- Organize the taxa into birds and crocodiles so that within group distances can be computed for both.
- Define the birds as an outgroup.
- Divide the alignment into intron, CDS (exon coding region), UTR (untranslated region) (see paper, page 390 and table 3) and the separately coded characters for indel regions (see paper page 389). Hint: export the alignment to a format that can be read by BIOEDIT and then manually align a downloaded Genbank file with one of the sequences (with features) to the alignment. Important note!: You may start the CDS at position 391 of the alignment. The start position of the CDS seems to be incorrectly given in Genbank accession AY277491. (Should start at position 344, and not at position 345 in that sequence.)
http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?list_uids=31322819.
- Compute “raw” p-distances among the crocodiles within the three defined regions (intron, CDS, UTR) (see paper page 391).
- Compute most parsimonious tree(s) and bootstrap for branch support.

Find an appropriate single evolutionary model for the complete data set (your version of PAUP* cannot handle mixed ML models for separate partitions) and

- run the model in ML with the search conditions used by the authors for the same analysis (number of replicates, addition sequence, etc.)
- test the molecular clock in Crocodylia (see page 395) (Recall that the degrees of freedom are $n-2$, n being the number of taxa.) Hint: a tree including only crocodiles with a shape like that in fig.2 can be made with the command *showtree/root=midpoint*. You can make this tree rooted or unrooted with additional commands in order to compute log scores (see BIO332 CD).

Find appropriate models for each of the data partitions and make a MrBayes script to

- do a Bayesian tree search with mixed models on the molecular matrix.
- make a table with parameter estimates (see paper table 4).

Do distances and rates differ in the three defined regions (intron, CDS, UTR), and if so, what could be the reason?

Load the molecular matrix and the Bayesian tree to PAUP*. Delete the birds and save the tree with “crocks” only as *gavials_MB.tre*.

Use the file *gavials_morph.nex* and delete all taxa that are not represented in the molecular tree (Hint: open the file in edit mode and see that many of those deletions have been defined already.) Load your *gavials_MB.tre* with the morphological data and set the criterion to parsimony. Reconstruct characters 2, 9, 12, 13, 14, 27, 30, 39, 60, 69, 77 in both an unrooted and rooted version of the MB tree. Which of these morphological characters indicate synapomorphic states for *Gavialis* and *Tomistoma* with this tree topology?

- What is the difference in the numbers of (morphological) steps between the molecular tree and the most parsimonious tree? (Compute the most parsimonious tree with the crocodile taxa in the molecular matrix.)

Discuss your result with particular reference to the paper by Harshman *et al* (2003).

Document your work and provide a table as an index to submitted file names and their contents. Remove trace of your personal identity. Put files in a zipped folder and deliver as an attachment with an e-mail to:

Deadline for submission is