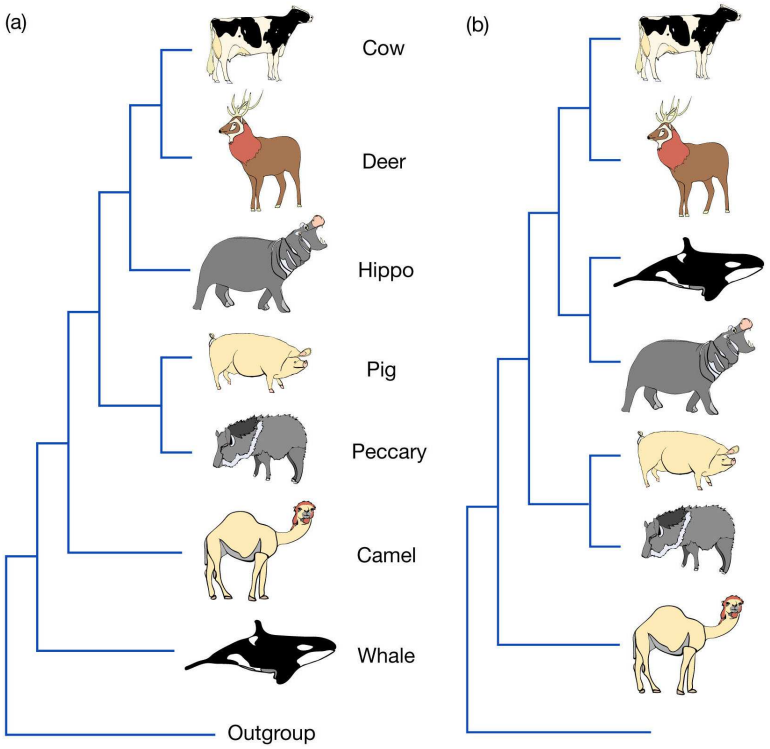
University of Bergen (E. Willassen & K. Meland)
BIO332 Phylogenetics
– take home assignment to be submitted by December 14, 2007

BACKGROUND

Traditional classifications of vertebrate taxa have been challenged by phylogenetic inference from both morphological and molecular data. Researchers have reached different conclusions with respect to the positions of whales and hippos. The figures below, from a textbook in evolutionary biology (Freeman & Herron, 2004: Evolutionary Analysis), shows two trees with differing placements of the two groups derived from (a)morphological and (b) molecular data. In the latter case, phylogeny reconstruction from a number of gene sequences has grouped the whales and hippos as sisters in an assembly that is colloquially known as "Whippos"



Copyright © 2004 Pearson Prentice Hall, Inc. Copyright © 2004 Pearson Prentice Hall, Inc.

The enclosed paper by Geisler (2001) is a contribution to the "Whippo" debate that critically questions the results from molecular analyses. The data used in that paper can be downloaded from the Internet from TREEBASE (http://www.treebase.org/treebase/index.html). Your task here will be to reanalyse these data in order to scrutinize previous conclusions.

THE DATA

Your data comes in two files: `geisler_nuc.nex` and `geisler_morph.nex`. The latter contains the coded morphological data. The key to the codes may be found in Appendix 2 of Geisler's original paper (`Geisler_2001.pdf`). You don't need special morphological knowledge about these characters for this assignment and should not worry about alien terminologies at this point. Notice that Geisler's assumptions about the transformations of multistate characters are given in the `ASSUMPTIONS` block. A rooted tree (Geisler 2001: figure 10A) is also included in the file. The molecular data is given in `geisler_nuc.nex`. The `SETS` block defines the extent of the aligned sequences from several genes. A tree (Geisler 2001: figure 10C) is also provided in the enclosed nexus-file.

YOUR TASK

1) Make a PAUP* script that initially assumes a GTR+I+G model for all character sets in the molecular data and estimates the parameters for each of the genes in the data set, based on the tree (Geisler 2001: fig10C) provided with the molecular data (`geisler_nuc.nex`).
Run your script and present a table in your report showing your results over the different genes. (hint: elegant solutions can be made by applying the `lscore` command with options as in the `MODELTEST` batch file.)

Genes	freqA	freqC	freqG	freqT	R(AC)	R(AG)	R(AT)	R(CG)	R(CT)	p-inv	Gamma shape
CYTB											
12S16S											
BCASX											
BCASI											
KCAS											
GFIB											
PROTEX											
IRBP											
VWF											
ALAC											

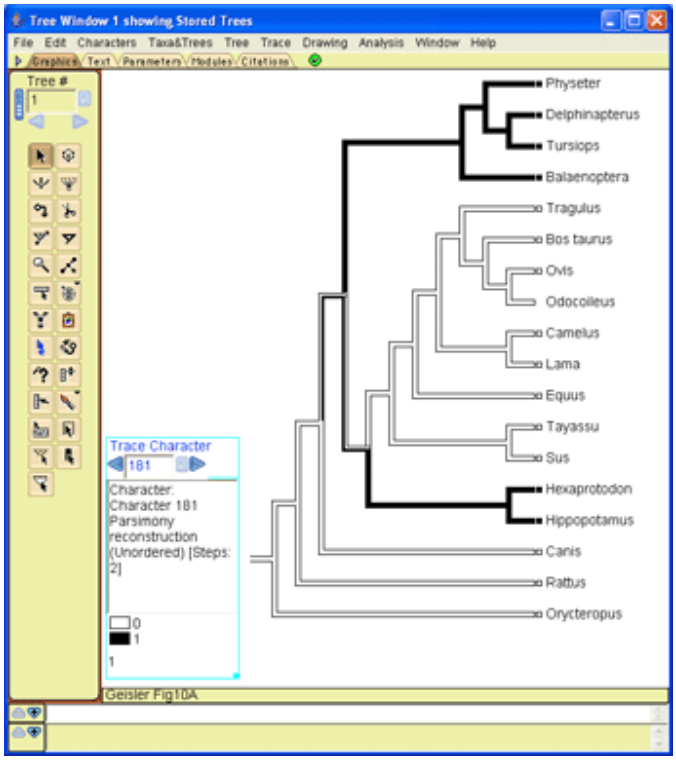
- Just by studying the parameter estimates, could you suggest simplifications of the GTR+I+G models for some of the genes in the data?
- Use your tabulated parameter estimates to calculate (with a spread sheet) the transition / transversion ratio (ti/tv) for each gene.
- Do you find any differences when comparing mitochondrial and nuclear genes?
- Do the ti/tv and gamma shape estimates indicate the possibility of saturation in some genes? If so, why? (Hint: see Page and Holmes Tab 5.1 and Fig.5.13)

2) Use MrModeltest to find suitable nucleotide models for the two partitions of mitochondrial and nuclear genes.

- Run a Bayesian phylogeny inference with mixed models on these two partitions of the molecular data. Do your results agree with the given 'molecular' tree (in `geisler_nuc.nex`), and is the Whippo group well supported?

3) Save the tree in `geisler_nuc.nex` as a separate nexus tree file called `geisler_mol.tre`.
4) The figure below shows morphological character 181 reconstructed on the morphology tree. Look up the character in Appendix 2. What would you say about the evolution of this trait, given

this tree? Load a copy (because you may want to keep the original) of `geisler_morph.nex` to Mesquite. Next, load your `geisler_mol.tre` in a new tree window and trace character 181 by means of parsimony ancestral state reconstruction. Now, what would you say about the evolution of this trait?



5) Run a parsimony search from the morphological data with all characters *unordered*. Does it make a difference in terms of topology and steps when you compare with the parsimony tree in `geisler_morph.nex` (which was derived from a set of *ordered* characters)?

6) Do a Bayesian search on the morphological matrix. (See example in the file `KIM.NEX` that comes with MrBayes. Remember the `ctype` settings for your data and check the MrBayes manual on this point.)

7) To examine the amount of support for the molecular tree in the morphological data, you could make a constraint tree to PAUP* with hippo and whale taxa forced to be monophyletic. Next, you load all the trees from the Bayesian tree files (`.t`files) to PAUP* and delete those that you discarded in the 'burnin'. Finally, you use the `filter` command to assess how many ***n*** of the trees that have clades corresponding to your constraint. The proportion ***n*** divided by the total number of trees is the Bayesian probability that the whippos are monophyletic given the morphological data. (If you can do these calculations you are reeaaally good!)

SUBMIT YOUR RESULTS AND REPORT

Write a report on your methods and your interpretations of the results. Document your submitted material and provide a table as an index to file names and their contents. Remove any trace of your personal identity. Put files in a zipped folder with name corresponding to your candidate/student number. Submit the zip-file as an attachment with an e-mail to:

Eli Neshavn Høie: eli.hoie@bio.uib.no