Mini project 1:

Building a tree from scratch: What are the closest relatives of whales?

How did whales evolve? What taxonomic groups are closely related to whales? What is the closest (non-whale) relative of whales that you can find? Follow the guidelines below and use everything you know about phylogenetic reconstruction (and bioinformatics, and molecular biology) to answer these questions.

Mini projects: General introduction

An important aim of the mini projects is to place the competences you learn on this course in the context of a full bioinformatics/systems biology work flow, and at the same time to get you to think about how to solve practical bioinformatics-type problems. The idea is that you should use everything you've learned so far (also on other courses!) so you will only get minimal instructions. If you get stuck, then check previous exercises and reading material (also from other courses!)

The mini project consists of 5 sub-problems (described below) that you should solve. Submit your response in the form of a report in a single PDF. Keep your answers as brief as possible while making sure that you have responded to all questions.

Peer assessment

An important aspect of the mini projects is that we will use peer assessment: Each student will evaluate the reports submitted from at least 3 other students. Both the evaluator and the report's author(s) are anonymous (you don't know whose report you are evaluating, and the other person will not know who evaluated their report). You will evaluate a number of aspects for each section of the report, typically by selecting a score on a drop-down menu. In some cases you will have the option of adding comments. The evaluation is done using a form on DTU Learn, which will become available next week. The evaluation form contains explanations about the meaning of different scores.

On this course, peer assessment is used mostly as a pedagogical tool: Seeing the work of other students can help you gain a better understanding of how best to write a report, and will put your own work in perspective. The quantitative and qualitative feedback you receive from fellow students can further help you understand how best to improve your work. Please keep this use of the qualitative feedback in mind when evaluating the work of others: be polite and constructive in your criticism, and focus on helping rather than on judging.

Good luck and have fun! Anders

1: Construction of data set

Construct a data set of nucleotide sequences suitable for analysing the evolutionary history of whales. The first step is to select a gene that is present in whales, and for which you will be able to find homologues in a suitable number of other species. You can use any method to build the data set including, for instance, keyword-based and sequence-based searching of databases. The aim is to construct a data set that contains both whale and non-whale species, and that has good taxonomic coverage. We are especially interested in analysing the nearest neighbours of whales in the tree, and for this purpose we would like good coverage of mammals.

The data set you construct should fulfil the following criteria:

- At least one whale-sequence (any whale species is acceptable; you can have several whales if you want)

- At least 10 non-whale, mammal species

- A number of non-mammal species. Try to cover a wide range of taxonomic groups (e.g., fish, amphibia, insects, birds, ...)

- The total size of the data set should be 20-40 sequences

- The data set should contain an outgroup (one or more species) on which you can root the tree (see later steps for more on tree reconstruction)

Briefly describe how you constructed the data set: what gene did you select? What database(s) did you use? How did you search? Include a list of the species covered in your data set. Also specify what your outgroup is, and include arguments for this choice (how do you know this is an outgroup?)

Also include an appendix with all your sequences in fasta-format.

2: Alignment of sequences

Align your data set using three different multiple alignment methods. Prepare each of the three alignments for use in PAUP (convert to NEXUS format).

Briefly describe which alignment methods you used and how you converted to NEXUS format (e.g., provide URLs for servers, or state the version numbers for stand-alone applications). Include the 3 alignments in Clustal format in an appendix.

3: Phylogenetic tree reconstruction

From each of the tree alignments: use parsimony to construct a phylogenetic tree rooted on your chosen outgroup. In each case use tree-rearrangements of the following two types:

- NNI
- SPR

This means **you have to construct a total of 6 different trees** (3 alignments x 2 tree-rearrangement types).

List the commands you used to construct the 6 trees. Briefly explain the purpose of tree-rearrangement in heuristic methods for tree-reconstruction. Also briefly explain NNI and SPR.

Include figures (e.g., screen-dumps, or images exported from FigTree) showing all 6 trees. Make sure trees are labeled with alignment method and tree-rearrangement method

.

4: Examine differences and similarities between trees

Check whether the 6 reconstructed trees differ. Quantify HOW different the trees are using the treedist command (find each pairwise tree-distance).

Include the results from the treedist command. Also list the commands you used. Briefly summarize the main similarities and differences between the 6 trees (do the trees agree on the most important/major aspects of the tree? What are the differences?). Briefly explain why different alignment methods (often) yield different results.

5: Conclusion: Whale evolution

Briefly discuss the questions you set out to analyze: What taxonomic groups are closely related to whales? What is the closest (non-whale) relative of whales in your trees? Were all ancestors of whales sea dwelling creatures? Any other thoughts on what your trees indicate about whale evolution?