# Reading a "back to the sea" story from molecular sequences: marine mammals evolution

**Goal: Evolutionary analysis of cetacean evolution using MEGA. Try to identify the closest neighbor terrestrial species related to whales and dolphins.**

## General guidelines for molecular phylogenetic analysis using MEGA

1. Download the three datasets in fasta format. Make a directory and save all the dataset files in fasta format.
2. Align the sequences using the two different programs available in MEGA (Clustal W and Muscle) and save the alignemnt in .meg format.
3. Construct phylogenetic trees using two different methods: a (distance-based) Neighbor-joining approach and a parsimony approach. Save the trees obtained as pdf files.
4. Interpret and compare the results obtained using the different datasets and the different methods. Pay special attention to tree topology and bootstraps values.

**DATASETS**
1. **Casein.fasta** K-casein exon 4 from 13 different mammalian species.
2. **Haemoglobin.fasta** Concatenated protein sequence of haemoglobin -alpha and beta chains from 9 different mammalian species.
3. **DNA_Concatenated.fasta** 10 Concatenated DNA sequences from six different genes from 10 different species (b -casein exon 7,K-casein exon 4,g-fibrinogen exon 2-4,g fibrinogens introns 2-3, protamine P1 exons 1-2, protamine p1 intron1 + 5´-3'non coding region).

These datasets have been compiled by John Gatesy and colleagues (Cladistics, 1999,15: 271-313). The sequences have been retrieved from
http://bioquest.org/bedrock/problem_spaces/whippo/data2.ph
The taxa represented in the datasets are:
Artiodactyl taxa: Bovidae (sheep, cattle, bison, springbok and antelopes),Cervidae (deer),Girafidae (giraffes) ,Tragulidae (chevrotains) ,Hippopotamidae (hippos) , Camelidae (camels and llamas) ,Tayassuidae (peccaries) , Suidae (pigs).
Cetacean taxa: Physeteridae (sperm whales) ,Delphinoidea (beluga whale, dolphins and porpoises), Ziphiidae (beaked whales) Mysticeti (baleen whales) and outgroup taxon
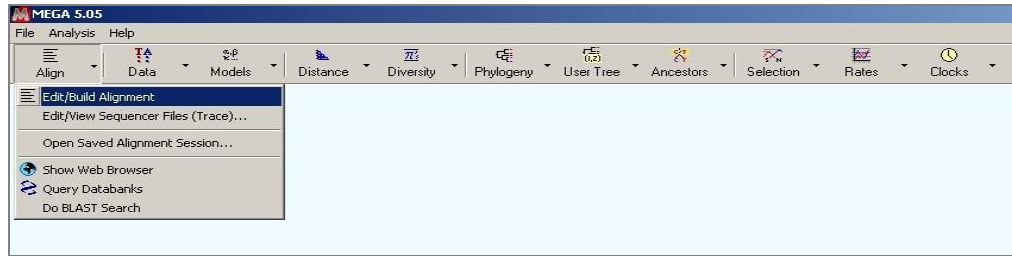Outgroup: (rhinos, horses and guinea pigs).

**STEPS**

## (A) SEQUENCE ALIGNMENT

**Step A.1**:  Start the program MEGA (Molecular Evolutionary Genetics Analysis freely available at http://www.megasoftware.net/)
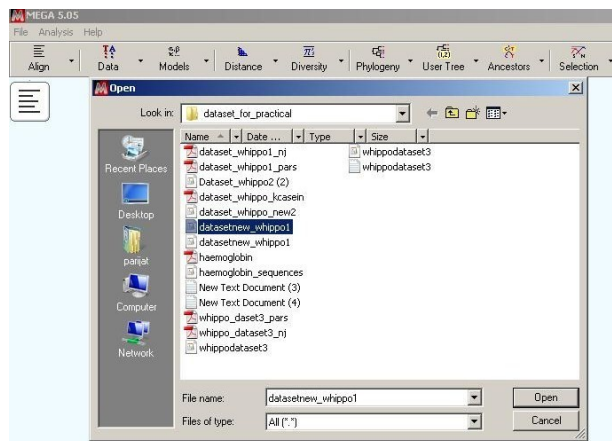
**Step A.2**: From the main MEGA window click on the "Align" tab,  click the  Add /Build alignment  tab from drop down menu, in the left hand corner of the screen.



**Step A.3:** Once you carried out the first step , a pop up window with an Alignment Editor will open up , click " Retrieve Sequence from files".
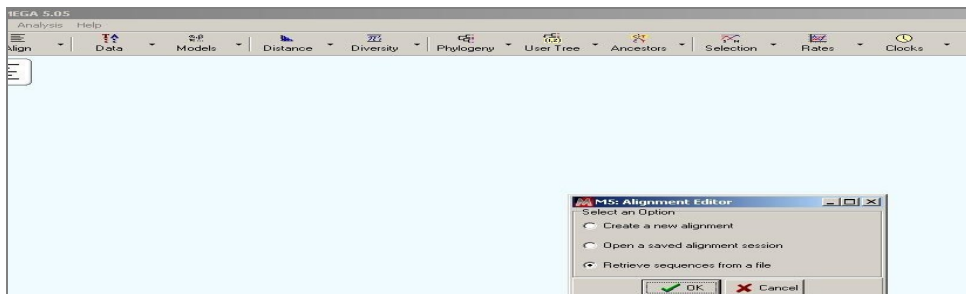


**Step A.4**: Choose the dataset of sequences from the directory/folder.

**Step A.5**:  **Alignment of the dataset.** Two different programs are available to perform the alignment of the data: ClustalW and Muscle. Perform both alignment types on your dataset using default paraeters, and save the aligned data in meg format. Export the alignment in meg format via the tab Data, Export alignment, and chose MEGA format. For each dataset you will have two different alignment files (e.g., kasein_ClustalW.meg, and kasein_Muscle.meg).

For each dataset the comparison of the trees obtained from the two different alignment will help you later to understand the relevance of the alignment process in reconstructing phylogenies.
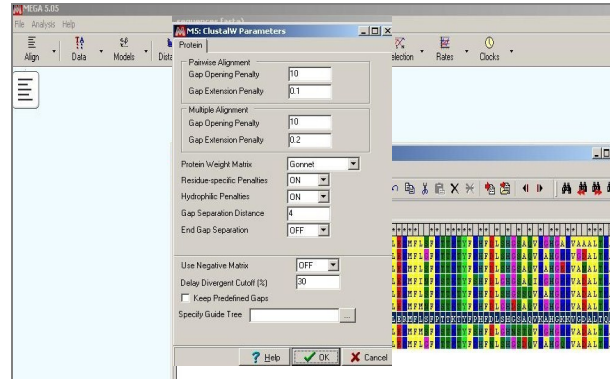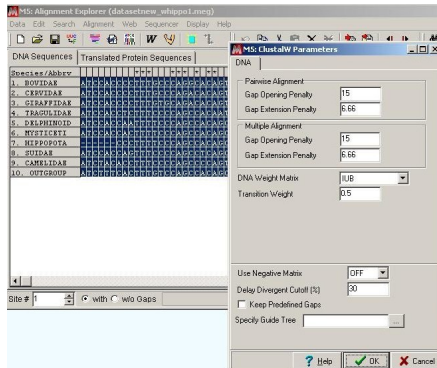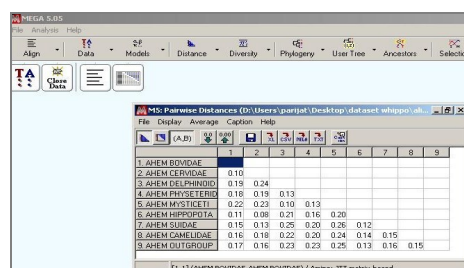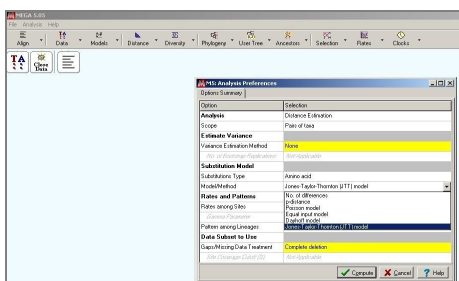


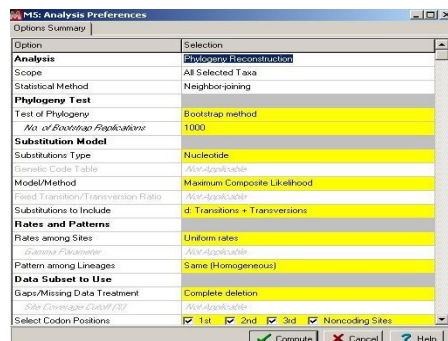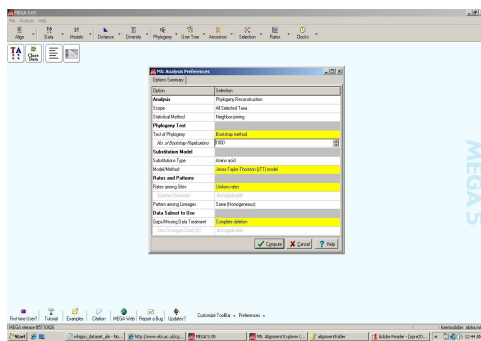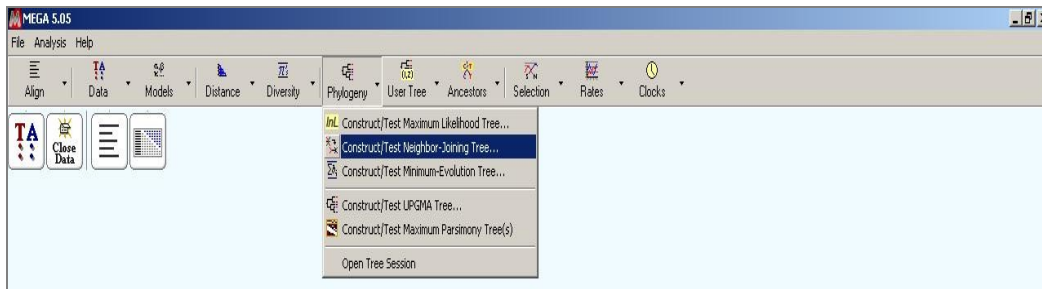## (B) PHYLOGENETIC RECONSTRUCTION: NEIGHBOUR-JOINING APPROACH

**Step B.1** : Open the meg files where you have stored your alignment in step A.5. Compute the pair-wise  distances between all the sequences in your dataset. Click the "**Distance**" tab from the main menu. Choose "Compute pairwise Distances".
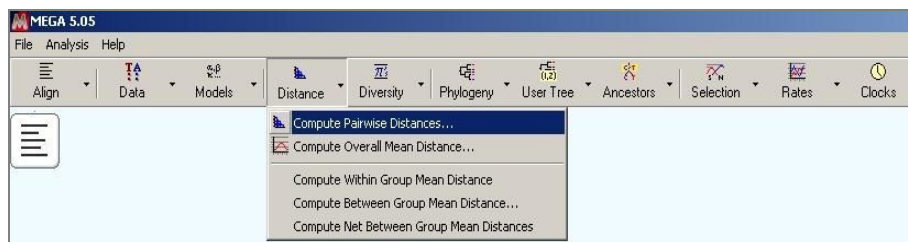


**Step B.2:**  Analysis preference opens up for  Pair wise distance calculations. Click on the yellow colour tab and choose parameters given below. Calculate the pairwise using gamma correction. Once the distance matrix is created, save it.

**Step B.3**: Create a, distance based, Neighbour-Joining phylogenetic tree: Click the tab **Phylogeny** from the main menu and choose Neighbour joining tree construction from drop down box. In the field "Test of phylogeny" set the number of bootstrap replication to 1000.





**Step B.4** : Save the tree in PDF format by clicking on save button on the left hand corner.
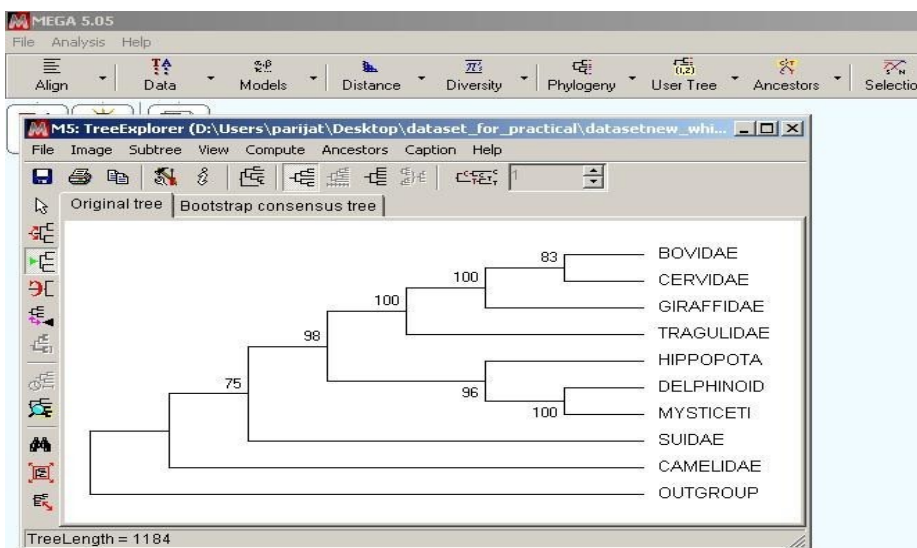
## (C) PHYLOGENETIC RECONSTRUCTION: PARSIMONY APPROACH

**Step C.1**:  Construct a phylogenetic tree using a **parsimony method. Repeat step B.3**  Choose parsimony method this time for construction. Open the  Sequence data explorer just by clicking the tab " TA" in the work space  and calculate " Parsim-info-sites", by clicking the **highlight** section, Save the tree as you have done before .
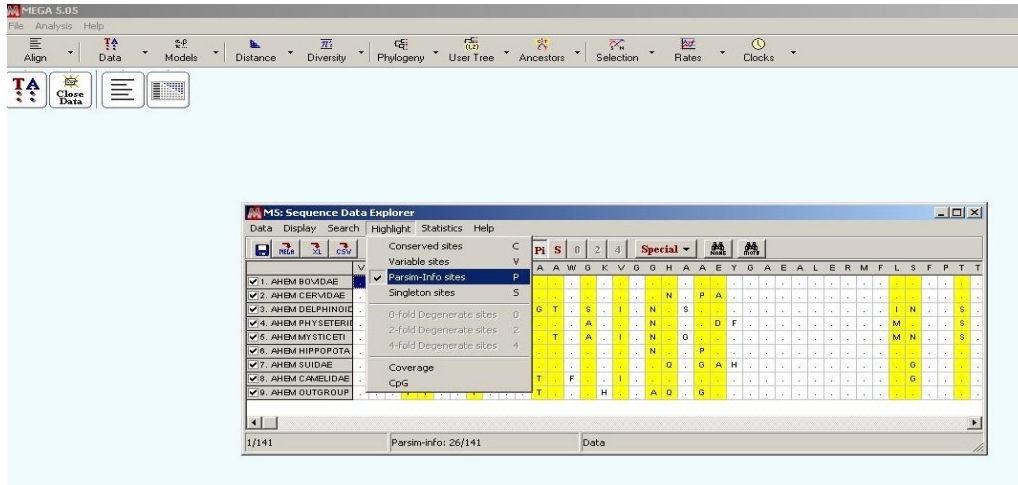


## (D)    COMPARATIVE ANALYSIS OF THE RESULTS OBTAINED WITH DIFFERENT DATASETS AND METHODS

Do all the datasets give the same tree topology?

How do the bootstrap values vary between the different datasets? Can you see a relationship between the amount of information available in the dataset and the bootstrap values?

Which is the terrestrial closest relative of cetaceans in the dataset?