

Question 1: Tree terminology, topology and rooting



1. Look at the picture of the above species tree. This is a tree of a selection of animal species. How many existing/"present day" species (=leaf nodes) are in this tree? Explain why you think this *is* or *is not* a rooted picture of a tree? If it is indeed rooted what "organism" would this root represent? How many internal nodes do you encounter when you go from human to pufferfish in this tree?
2. We have another species tree in the following file: [alternative_species_tree.ph](#). (download via e.g. right click save as) This tree displays the phylogenetic relations between a set of globin genes from the same animals as present in the previous question. This tree is given in the newick format, which is one of the official tree formats and normally characterized by the extension .ph. Download this tree file (follow the link and select "Save As" or "Save Page As" from the "File" menu). We are going to look at this file using the tree viewing web-page iTOL at <http://itol.embl.de> (which we hope will hold up during the course). On the home page you can browse some examples of the kind of images you can create using iTOL. Upload your tree via the "DATA UPLOAD" tab as file via "**Tree file:**" → browse, or paste your tree as text in the in the "**Tree text:**" input field. How many species (=leaf nodes) are there in this tree? This tree (i.e. the data structure in the newick file that is used to draw the picture you are now viewing in iTOL) was generated by a maximum likelihood method which is a state of the art phylogenetic method. Given this origin, is this tree a rooted tree or not? How many internal nodes do you encounter when you go from human to pufferfish in *this* tree?
3. Use iTOL to change the tree from question 1.2 (alternative_species_tree.ph) into the tree pictured for question 1.1. You can do this by clicking on the branch where you would like to root on, under "editing" → "tree structure" → "reroot the tree here". In addition to rerooting you might want to flip the order around internal nodes by selecting rotate branch when you click on the branch you want to rotate along.

Can you obtain the same tree as pictured above?

4. Draw on a piece of paper the explicitly **unrooted** tree of the alternative_species tree. Pinpoint in this tree where you can place the root to make it identical to the species tree at the top of this page. You can check your drawing in iTOL by selecting the unrooted mode for displaying the tree (in the right-hand side panel at the “display mode” option click the unrooted button).

Question 2: Ubiquitin trees

1. Construct phylogenies for ubiquitin using the protein sequences given [ubiq_prot2.fas](#). Download these sequences to your computer for example using right click save as, alt-click or by copying them into notepad and saving as a text file. These sequences have not yet been aligned. You thus first need to align the sequences. To align your protein sequences, we will use clustal omega at <https://www.ebi.ac.uk/Tools/msa/clustalo/>. Upload the sequences to clustal omega using the “browse” button or paste the sequences as unformatted text in the sequences field. Then press the submit button. Go to the “phylogenetic tree” tab and download the treefile by clicking the “Download Phylogenetic Tree Data”. (also download this file by using right click save as, alt-click or by copying them into notepad and saving as a text file). View the resulting tree by opening the file in iTOL as above. Based also on the alignment, explain why the tree looks the way it does?

2. Now we are going to construct phylogenies for ubiquitin using the DNA sequences of the genes given in [ubiq_dna2.fas](#) . Again, you need to align them using clustal omega, which also makes a tree for you. View the tree in iTOL and reflect on its quality in light of the species phylogeny: and be specific!, i.e. which species and relations agree and which do not ? *If you are not sure about animal evolution/phylogeny, you can look at <http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi> for information on the phylogenetic relationships within the animals. Specifically you can do this there by searching for relevant species under “Enter name or id” pressing the add button (sometimes you need to use latin species names, such as chicken = Gallus gallus, or frog = Xenopus tropicalis.) To see more resolution in the tree, which is likely necessary check the “include unranked taxa” checkbox).*
3. Make another tree for these species using the protein sequences of methionine adenosyltransferase (an essential metabolic enzyme) given in [MAT1.fas](#). To what extent does this tree based on the MAT protein correctly reflect the species phylogeny?

Question 3: SARS

SARS (Severe Acute Respiratory Syndrome) is a respiratory disease in humans which is caused by SARS coronavirus (SARS-CoV). The outbreak of SARS that nearly became pandemic caused quite a scare in 2003 (see wikipedia). We have collected matrix proteins from 15 corona viruses (note that only one of them is SARS) in <http://bioinformatics.bio.uu.nl/snel/ME/CoronaM.fas>

1. Make a phylogenetic analysis of these viruses based on the matrix protein. What does the tree suggest with regards to the origin of SARS?
2. How does this compare to the [relevant literature](#) at the time?