Group Project - Phylogenetics

The purpose of this exercise is fourfold:

1) To become familiar with mining biological databases
2) To develop specific, testable evolutionary hypotheses
3) To get hands-on experience in performing phylogenetic reconstructions
4) To develop facility with interpreting phylogenies, quantifying and comparing evolutionary rates, and character mapping

Step 1: In consultation with me or your TA (or both), develop an interesting evolutionary question (e.g. Are HIV and SIV reciprocally monophyletic? Did flight evolve once or twice in mammals? Do different DNA sequence datasets reveal similar evolutionary affinities?). **Be sure to emphasize species-level questions.**

Step 2: Go to Genbank or another sequence repository and check whether sequences for exemplar individuals from species relevant to your question exist; speak to us to ensure that genetic marker you have chosen is appropriate for your specific question (i.e. does it evolve at the appropriate rate?). If so then download all relevant sequences and assemble them in a FASTA format (save this as a text file). Because an increasing the number of taxa in your analysis also increases the computational burden, we ask you to limit the number of sequences to about 30-40 (as you will see, performing a bootstrap analysis is essentially running the same analysis "n" times where n=number of bootstrap replicates). At least one of these sequences must represent an outgroup (a taxon/sequence which is a recent relative that will help root your focal tree, and help reconstruct the evolutionary sequence of changes in key traits). Make sure to have handy the accession numbers for the sequences as these are unique identifiers that allow you to retrieve any salient information.

Step 3: Make a 2-3 hour appointment with me, during which we will:

- **Step 3a:** Perform a sequence alignment using ClustalX.
- **Step 3b:** Perform phylogenetic reconstructions on the aligned sequences using one or more of PAUP*, MEGA BEAST, or MrBayes.
- **Step 3c:** Perform analyses to assess the robustness of your phylogenetic reconstructions.
- **Step 3d:** If relevant perform some additional analyses relevant to your research question (e.g. constraint analysis, character state reconstruction).

Please note that you might have to make multiple appointments as your project might be multi-stage, and some of the analyses might require significant computation time.

GROUP ORAL PRESENTATION ON YOUR FINDINGS

Introduction: introduce your question, why it is interesting, and why a molecular phylogenetic approach is appropriate.

Methods: briefly describe the steps you took to address your question. Make sure to rationalize your choice of molecular marker. What genome is it from? What is the mode of inheritance? Also indicate why you chose the outgroup that you did. Remember that all members of your audience will be familiar with phylogenetics so you do not have to belabour the methods simply indicate which ones you deployed and why.

Results: give trees and support values, and other salient outcomes from your analyses.

Discussion: Answer the following questions:

1. Discuss why your evolutionary question is interesting and why your approach to addressing it is valid (you may cite other papers where this approach was successful).
2. Did the trees from different analyses differ? How and why?
3. What is the relative level of support for your hypothesis in the two trees? What does bootstrapping actually measure?
4. Discuss the implications of your results. What novel insights into speciation and macroevolution to they provide (you may refer to other articles here of course to broaden the relevance of your findings)? Does your hypothesis require further testing (more data, more taxa, appropriateness of marker chosen)? How would you go about doing this? What other lines of evidence would strengthen your conclusions?