03-727 Project Description

Deliverables

1. Title, team members by email on November 7th
2. Project plan: 1 page, due November 14th
3. In-class presentation: December 12th (5:30pm – 8:30pm)
4. Final paper: 3 – 5 pages (not including appendices), due December 14th

Project content:

1. *Ask a phylogenetic question:* Give a brief statement of the motivation for studying a given data set. Examples include determining the evolutionary relationships of a group of species; correlating the expansion of a given gene family with the evolution of a biological character of interest (susceptibility to a disease, novel metabolism, a new developmental program, invasion of an ecological niche, etc.); and investigating which methods work best for a specific type of data.

2. *Collect appropriate sequences:* Given your question, decide which sequences and species to include in your data set. You will have to find the right compromise between thorough sampling and limiting computational and conceptual complexity. Avoid bias as much as possible by sampling from a broad set of taxa and keep in mind the importance of outgroups. For some projects, characterizing the phylogenetic distribution of the gene family may be an important part of your analysis.

3. *Multiple sequence alignment:* Try several different programs to align your data and experiment with different parameters, substitution matrices, etc. Consider refining your alignment manually.

4. *Phylogeny estimation:* Consider which algorithmic strategy is appropriate for your data set, given the age and conservation of the sequences. Prior to phylogeny reconstruction, you will want to use a program like ModelGenerator to select a sequence evolution mode for your analysisl. As above, try several different programs and check to see if the results agree.

5. *Assess the reliability of your tree:* Bootstrapping can be used to assess the reliability of individual branches in the tree. In a probabilistic framework, various approaches for testing the reliability of the entire tree are also available. You can also assess the reliability of your tree by comparing its “predictions” with known biological relationships.

6. *Interpret results:* Discuss your results in light of your original question. What conclusions can be drawn from your analysis? You may wish to compare your results with previously reported results in the literature. However, the main goal of this project is phylogenetic analysis, not a

literature survey. You will also want to discuss the limitations of your results. Given your assessment of reliability, how robust are your conclusions? If you had more time, what additional tests would you perform?

Project plan: 1 page, due November 14th

Your project plan should describe your proposed dataset and a preliminary statement of your question. You should include a plan for acquiring sequences, whether you will use DNA or amino acid sequences; which taxa you plan to sample; and where such data can be obtained. If possible, include a preliminary plan of which methods you plan to use. You may wish to cite one or two relevant papers.

Your initial plan is due on November 14th. If I am concerned about the feasibility of your plan, I may make suggestions and ask you to submit a revised plan that addresses these concerns.

Final paper: 3 – 5 pages (not including appendices), due December 14th

Your paper should include the following sections: Introduction, Methods, Results/Discussion, an appendix, and references. Given the nature of this project, the methods section will be a substantial part of your paper. The key goal is to demonstrate what you have learned about using phylogenetics to investigate a biological question, the relative merits of various phylogenetic methods, and which methods to use in a given context. Your final paper should include an appendix with supplementary figures, raw data, alignments and trees. If your paper describes a team effort, also include a brief summary of who did what.