

Molecular phylogenetics mid-term 2016

Instructions – Please read carefully

1. **This exam is to be written entirely on your own, without collaboration or discussion with anyone else. You may not discuss any aspect of it with any other person until you have turned it in. The exam is subject to the university's rules on academic integrity.**
2. There is no limit on how much time you may spend on the exam. The exam is open-book. You may refer to notes, books, articles, and so on. The normal scholarly expectations for attribution of language or ideas apply.
3. Please submit printed/typed responses to the following questions. Use double-spaced format and an easily readable font. Do not forget to put your name on the exam.
4. The purpose of this exam is to evaluate your understanding of fundamental ideas, strategies and methods in phylogenetics that we have covered to date. It is designed to allow you to demonstrate your mastery of the material. Your goal should be to show me that you really understand the concepts we discussed. Citations to the literature strengthen your argument and are encouraged.
5. Please answer each question **clearly and completely**, and be sure to justify your responses. For many of these questions there is no single correct answer; there are often multiple approaches or opinions that are reasonable. Your responses will be judged based on how well you justify your choices.
6. No question should require more than about 2 pages of double-spaced response, and most can be considerably shorter. You will have to use your judgment about what makes for an adequate answer. Do not go on and on to make me think you know what you are talking about: concision is a virtue.
7. The quality of your writing is important, so be sure take the time to read and edit your responses. Think of the exam as a paper, subject to the normal high standards of clarity, insight, and grammar/style.
8. The exam is due in class or by email by 1 pm Wednesday May 24. It can be submitted earlier if you like.

Exam Questions. Please answer all questions.

1. Cladists choose as the best inference of phylogeny the tree that requires the fewest evolutionary character changes. Why? How do they justify this criterion?
2. Please identify the following statement as true or false and justify your answer briefly.

The amino acid character shown below is phylogenetically informative in a cladistic framework. (Be sure you explain the meaning of “phylogenetically informative” as part of your answer).

Wolf	Alanine
Fox	Alanine
Bear	Valine
Cat	Leucine

3. Suppose you are interested in the relationships among mammalian orders. You sequence the gene cytochrome oxidase II from species in the various orders. You calculate for every pair of DNA sequences the percent of sites that differ between the two sequences. Professor Lockhart, your molecular biology professor, proposes to use UPGMA and neighbor-joining on the resulting distance matrix to infer the mammalian phylogeny.

- a. What are some reasons that UPGMA might yield a tree other than the true one as the best inference of phylogeny?

- b. What are some reasons that NJ might yield a tree other than the true one as the best inference of phylogeny?

4. You analyze the phylogenetic relationships among rodents and rabbits using DNA sequences of the epidermal growth factor (EGF) gene. You have EGF sequences from 21 species: one guinea pig species and two from each of the other major rodent families (squirrels, beavers, mice/rats, pocket gophers, mole rats, and jerboas (jumping mice)), as well as two species of rabbit (which are lagomorphs, not rodents). In addition, you have EGF sequences from three species of bats and two primates as outgroups. As is found for many genes, the guinea pig sequence appears to be very fast-evolving. Its uncorrected pairwise differences with each of the other ingroup taxa are 50-60%, while most of the other taxa have ~20-30% pairwise differences from each other.

You conduct an unweighted parsimony analysis using a thorough heuristic search, and you find that the rooted MP tree is (bats, primates, (guinea pigs, (rabbits, other rodents))). This tree is not consistent with the traditional view of rodents as a monophyletic group. The node that unites rabbits and other rodents to the exclusion of guinea pigs has decay index $D=8$ and bootstrap value 98 percent. The branch leading to this node from the next deeper node (the one that unites guinea pigs with rabbits and the other rodents) has length = 0.09 when expressed as the probability of a change per site on this branch.

- a. Do these observations give you reason to be concerned that the nonmonophyly of rodents is an artifact of error in the phylogenetic method rather than the true tree? Please explain why or why not.
- b. What can you do to investigate whether your finding is likely to be an artifact or the true phylogeny? Please suggest three strategies you can use to shed light on this problem. Be specific about how you will interpret the results of your research (i.e., result is X would provide evidence that H1 is true, because....).

5. You are interested in determining whether the Ecdysozoa/ Lophotrochozoa hypothesis for the phylogeny of bilaterally symmetric animals is correct. You address this problem using the sequences of the RNA polymerase II (RNAPol) gene, which you have obtained from 75 animal taxa, including 2 cnidaria, 3 nematodes, 40 vertebrates, 20 insects, 5 mollusks, 2 echinoderms, and 3 annelids. You use both parsimony and maximum likelihood methods.

- a. Please describe how you will choose a model for your analysis. Be specific and justify your approach.
- b. Suppose you conduct both bootstrap and likelihood-ratio analyses. The ML tree contains a clade of coelomates (deuterostomes, arthropods, annelids, and mollusks), to the exclusion of nematodes and Cnidaria. You find that the node that contains coelomates is labeled with the number 66 for bootstrap, and with 3.8 for the likelihood ratio statistic. Explain what each of these results means. Also, explain and justify how much confidence you now have that Ecdysozoa is a true clade.
- c. Suppose you also perform parsimony-based analysis, which also results in a clade consisting of coelomates, to the exclusion of nematodes and Cnidaria. The clade has decay index $D=17$ and a bootstrap value of 94. How do you interpret these results in relation to your findings in part b? Given these results, what do you conclude about this phylogenetic question?
- d. Given all these results, what would you do next?