

Examination Bioinformatics

Linköpings Universitet, IDA, Statistik

Course:	732A51 Bioinformatics
Date:	2022/02/07, 8–12
Teacher:	Krzysztof Bartoszek
Provided aids:	The help material is included in the zip file exam_help_material_732A51.zip .
Grades:	A= [18 – 20] points B= [16 – 18) points C= [14 – 16) points D= [12 – 14) points E= [10 – 12) points F= [0 – 10) points
Instructions:	Provide a detailed report that includes plots, conclusions and interpretations. Give motivated answers to the questions. If an answer is not motivated, the points are reduced. Provide all necessary codes in an appendix. If you are asked to do plots, then make sure that they are informative, have correctly labelled axes, informative axes limits and are correctly described. Points may be deducted for poorly done graphs. Name your digital part solution files as: [your id]_[own file description].[format] If you have problems with creating a pdf you may submit your solutions in text files with unambiguous references to graphics and code that are saved in separate files There are THREE assignments (with sub-questions) to solve. Include all code that was used to obtain your answers in your solution files. Make sure it is clear which code section corresponds to which question. Your code should be complete and readable, possible to run by copying directly into a script. Comment directly in the code whenever something needs to be explained or discussed.

Problem 1 (5p)

Two species of algæ (Chlorophyta), *Acetabularia acetabulum* and *Batophora occidentalis*, have a special non-standard genetic code, the triplets TAA and TAG code for Glutamine.

1. (1p) With the exam you have the file `Lekomtsev2007.pdf` which provides examples of such non-standard genetic codes. In particular, in the table on p. 880 the changes for the two algæ species are provided. Explain how one obtains from that table that TAA and TAG code for Glutamine.
2. (3p) You are provided with a short piece of DNA coming from *Acetabularia acetabulum*:
GCTCACTAAATCATC
Find all possible translations of it. You will find the standard genetic code provided in Lecture 1 of the course.
3. (1p) Provide the protein sequence that this short piece of DNA codes for. Explain why you chose this particular protein sequence.

Problem 2 (7p)

You are given the following set of five DNA sequences:

1. ATCGACGT
2. ACGTTATC
3. ACGTGCAT
4. CGGAACGT
5. CGTTCTAG

1. (2p) Align the sequences by hand, justify your alignment.
2. (1p) Which algorithm would you use to align these sequences? Justify your choice.
3. (2p) Propose and justify a tree joining these sequences. There is no need to provide branch lengths, only the topology.
4. (2p) For what purpose is sequence 5. useful when constructing the tree? Which algorithm would you choose to build the tree joining these sequences. Justify the choice of the algorithm.

Problem 3 (8p)

In the file `phyltree.nex` you are provided with a phylogenetic tree. Use the function `ape::read.nexus()` to read it in. For the following questions you have to provide all the code, calculations or reasoning that you used in order to obtain the answers.

1. (1p) Plot the tree. For this purpose explore the possibilities of the plotting function associated with the tree's class in order to make the plot readable, especially the node labels. You might need to increase the size of the plot. Save the plot of the tree in a separate file and provide part of your solution.
2. (2p) Find the maximum degree of this phylogenetic tree (a tree is a graph).
3. (1p) Find the path length between tip nodes `tip_99` and `tip_100`.
4. (2p) Find the path length between tip nodes `tip_3` and `tip_68`.
5. (2p) Find the path length between tip nodes `tip_71` and `tip_80`.