Examination Bioinformatics

Linköpings Universitet, IDA, Statistik

Course:	732A51 Bioinformatics
Date:	2025/01/15, 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 exam account id**]_[**own file description**]
	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.
	If you also need to provide some hand–written derivations
	please number each page according to the pattern: Question number . page in
	question number i.e. Q1.1, Q1.2, Q1.3,..., Q2.1, Q2.2, ..., Q3.1, ...

Problem 1 (7p)

You are given two DNA sequences TCAAG and TCTACG. Choose some distance function (for which these two sequences are valid input) and calculate the distance between these two sequences under the chosen function. Perform (manually) a global alignment between the two sequences. Choose all necessary parameters yourself (however they should be nontrivial, i.e., **NOT** 0, and furthermore meaningful). Do not forget to report the dynamic programming matrix and how one obtains the optimal solution from it. Explain what an alignment is from a biological point of view.

Problem 2 (7p)

In the file `LowHeight.fasta` you are provided with the DNA sequences of a certain exon in a certain gene that is suspected to be associated with being short. This fragment has been obtained from a number of individuals: short and tall. Identify which element of this genetic fragment can be associated with being short, propose some statistical test for this and carry it out.

Problem 3 (6p)

A) **(3p)** You are given, in Newick format, the following two rooted trees with tip labels

$$T_1 = (A, (((E, D), C), ((G, F), (B, H))))$$

and

$$T_2 = ((((H, B), (F, G)), (C, (D, E))), A).$$

Are they identical in topology?

B) **(3p)** Draw all possible rooted, without tip labels, trees with 3, 4 and 5 tips.