

Examination Bioinformatics

Linköpings Universitet, IDA, Statistik

| | |
|----------------|--|
| Course: | 732A51 Bioinformatics |
| Date: | 2020/08/24, 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: | <p>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. 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, Scan/take photos of such derivations preferably into a single pdf file but if this is not possible multiple pdf or .bmp/.jpg/.png files are fine. Please do not use other formats for scanned/photographed solutions. Please submit all your solutions via LISAM or e-mail. If emailing, please email them to BOTH krzysztof.bartoszek@liu.se and KB_LiU_exam@protonmail.ch . During the exam you may ask the examiner questions by emailing them to KB_LiU_exam@protonmail.ch ONLY. Other exam procedures in LISAM.</p> |

Problem 1 (6p)

In the file `DNAseq.fasta` you have a sequence of nucleotides. Translate it into its corresponding RNA sequence and then using the standard genetic code into the appropriate protein sequence.

Problem 2 (7p)

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

Problem 3 (7p)

In an organism's protein interaction network (i.e. the network that describes which proteins interact with which) the number of connections to a target protein follows a binomial distribution, with parameters n and p . That is the probability that a protein interacts with $0 \leq k \leq n$ other proteins equals $\binom{n}{k} p^k (1-p)^{n-k}$.

- a) (3p) Derive the ratio of the expected number of target proteins with four interacting proteins to the expected number with three interacting proteins. Provide a plot for this when n changes from 5 to 100. Take $p = 0.1$.
- b) (3p) Write a short program (or do analytically, but coding is probably easier) to find the maximum value of k for which at least 1% of the proteins interact with at least k proteins for $n = 15$ and $p = 0.3$.