UNIVERSITY
OF SKÖVDE

School of Bioscience

WRITTEN EXAMINATION

Course: Introduction to bioinformatics

Examination: Supervised written examination

Course code: BI123G

Credits for written examination: 6.5

Date: 24 August 2026

Examination time: 4 hours

Examination responsible: Zelmina Lubovac

Aid at the exam/appendices: None

Other

Instructions

- ☐ Take a new sheet of paper for each teacher.
- ☒ Take a new sheet of paper when starting a new question.
- ☒ Write only on one side of the paper.
- ☒ Write your name and personal ID No. on all pages you hand in.
- ☒ Use page numbering.
- ☒ Don´t use a red pen.
- ☒ Mark answered questions with a cross on the cover sheet.

Grade points: 0-18 = F; 19-21 = E; 22-25 = D; 26-28 = C; 29-32 = B; 33-36 = A

Examination results should be made public within 18 working days

Good luck!

Total number of pages: 6

UNIVERSITY
OF SKÖVDE

Question 1 (6p)

State if the claim about primary and secondary databases is true or false. You do not need to give any motivations in your answer, just writing (for each claim) "true" or "false" is sufficient.

1a) Secondary databases contain information derived from the analysis and interpretation of data stored in primary databases. **(1p)**

1b) GenBank is a secondary database that mainly stores protein families and conserved sequence patterns. **(1p)**

1c) Primary databases may store experimental data, such as nucleotide sequences or gene expression values. **(1p)**

1d) PROSITE is a primary database because researchers submit raw DNA sequencing results directly to it. **(1p)**

1e) Gene Ontology provides standardized terms for describing features such as molecular functions, biological processes, and cellular components. **(1p)**

1f) ENA is primary database that stores only derived protein family information. **(1p)**

UNIVERSITY
OF SKÖVDE

Question 2 (6p)

Part 1 – BLAST general (4 points)

Which BLAST version is most appropriate for identifying a protein of unknown function against known protein sequences? State if the claim is true or false. You do not need to give any motivations in your answer, just writing (for each claim) "true" or "false" is sufficient.

2a) tblastn **(1p)**

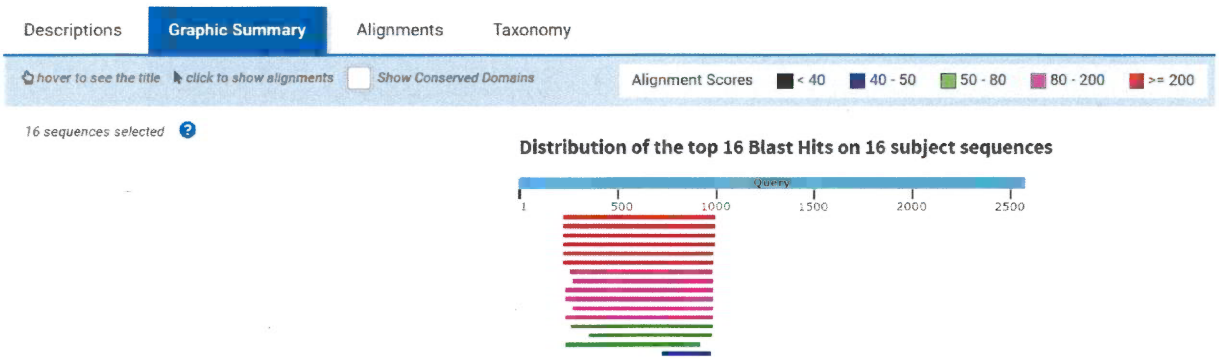
2b) blastp **(1p)**

2c) blastx **(1p)**

2d) blastn **(1p)**

Part 2 – BLAST specific interpretation (2 points)

Analyse the "Graphic Summary" below for the search results for human PDYN gene (NM_024411.5) using blastn.



2e) Write a brief interpretation of the information that could be gained from the summary. Your interpretation should include the **colors of the bars**, their **lengths** and their **location** relative to the Query **(2p)**.

UNIVERSITY
OF SKÖVDE

Question 3 (6p)

Part 1 (4p)

What is the primary objective of protein global alignment compared with protein local alignment? State if the claim is true or false. You do not need to give any motivations in your answer, just writing (for each claim) "true" or "false" is sufficient.

3a) To identify short conserved motifs in non-coding DNA **(1p)**

3b) To ignore gap penalties entirely. **(1p)**

3c) To find the highest-scoring subsegment between the protein sequences. **(1p)**

3d) To align protein sequences from end to end across their entire lengths. **(1p)**

Part 2 (2p)

3e) State the main steps used by ClustalW to produce a multiple sequence alignment. **(2p)**

UNIVERSITY
OF SKÖVDE

Question 4 (6p)

4a) What does the "Fold Change" (FC) represent in differential expression analysis? **(2p)**

4b) Explain "Multiple Testing Problem" in the context of transcriptomics? **(2p)**

4c) Which methods are commonly used to correct for the Multiple Testing Problem? **(2p)**

UNIVERSITY
OF SKÖVDE

Question 5 (6p)

State whether each claim about the data used in network-based enrichment analysis is true or false. You do not need to provide any motivation; writing "true" or "false" for each claim is sufficient.

5a) Network-based enrichment analysis uses only gene expression levels. **(1p)**

5b) It uses only information from literature-based experiments. **(1p)**

5c) Protein–protein interaction networks can be incorporated into the analysis. **(1p)**

5d) It considers only the number of genes shared between two pathways. **(1p)**

5e) Relationships between genes or proteins can be used to identify connected biological processes. **(1p)**

5f) Network-based enrichment analysis cannot incorporate information about pathway topology. **(1p)**

UNIVERSITY
OF SKÖVDE

Question 6 (6p)

6a) What does the "guilt by association" principle suggest about an unknown protein? **(3p)**

6b) Explain the role of modules in a biological network **(3p)**