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: 3 June 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 for your answer, just writing (for each claim) "true" or "false" is sufficient.

1a) PROSITE is an example of a primary database because it stores raw DNA sequences. **(1p)**

1b) Secondary databases are built by analysing data from primary databases and storing derived information. **(1p)**

1c) Secondary databases cannot contain pathway or protein family information. **(1p)**

1d) Primary databases store raw biological data, such as nucleotide sequences or gene expression values. **(1p)**

1e) GenBank and ENA are examples of nucleotide sequence databases. **(1p)**

1f) Gene Ontology is described as metadata because it provides "data about data," such as annotation terms. **(1p)**

UNIVERSITY
OF SKÖVDE

Question 2 (6p)

Part 1 – BLAST general (4 points)

For BLAST program blastx, 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) Compares a protein query against a protein database **(1p)**

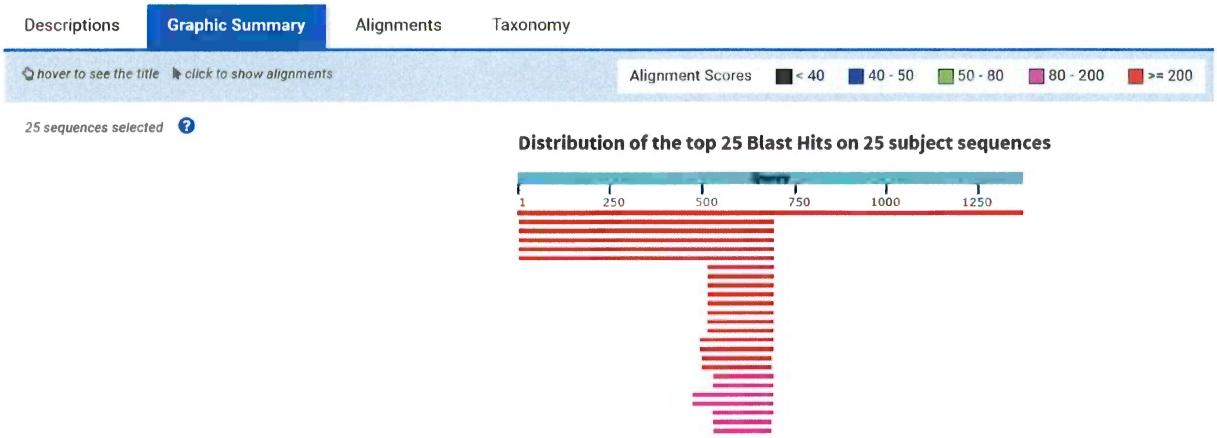
2b) Compares a nucleotide query against a nucleotide database **(1p)**

2c) Compares a translated nucleotide query against a protein database **(1p)**

2d) Compares a translated nucleotide query against a translated nucleotide database **(1p)**

Part 2 – BLAST specific interpretation (2 points)

Analyse the "Graphic Summary" below for the search results for human HOXB7 gene (NM_004502.4) 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)

ClustalW employs a standard three-step approach to produce a multiple sequence 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) Iterative refinement, tree building, and gap penalization **(1p)**

3b) Pairwise alignment, guide tree construction, and progressive alignment. **(1p)**

3c) Hidden Markov Modeling, mBed clustering, and profile alignment. **(1p)**

3d) Circular mapping, neighbor-joining, and heuristic search. **(1p)**

Part 2 (2p)

3e) Explain why multiple protein sequence alignment is useful in bioinformatics.? **(2p)**

UNIVERSITY
OF SKÖVDE

Question 4 (6p)

Which of the following statements about Differential Gene Expression (DGE) analysis are true. For each claim, 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.

4a) It measures the total amount of DNA in a cell. **(1p)**

4b) It can be measured by Fold Change (FC). **(1p)**

4c) It identifies genes whose expression levels change significantly due to an experimental variable. **(1p)**

4d) It determines the 3D structure of the ribosomes. **(1p)**

4e) It compares transcript abundance between two or more conditions (e.g., healthy vs. diseased). **(1p)**

4f) It is commonly used as input to annotation enrichment analysis. **(1p)**

UNIVERSITY
OF SKÖVDE

Question 5 (6p)

Choose two of the following annotation enrichment approaches: Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network-Based Enrichment Analysis (NEA). For each of the two approaches you choose, describe the main idea of the method and state one strength and one weakness.

UNIVERSITY
OF SKÖVDE

Question 6 (6p)

6a) Explain the role of an adjacency matrix in biological network analysis. In your answer, describe what the rows, columns, and numerical values represent, and give one example of a network property that can be calculated from the matrix. **(3p)**

6b) Complete the adjacency matrix for the undirected graph in Figure 1. **(3p)**

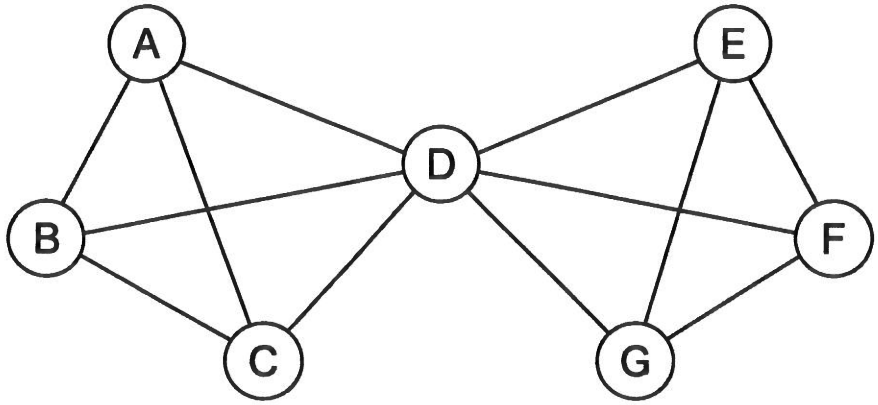


Figure 1: Undirected graph

	A	B	C	D	E	F	G
A							
B							
C							
D							
E							
F							
G							