UNIVERSITY
OF SKÖVDE

School of Bioscience

WRITTEN EXAMINATION

Course: Bioinformatics – Concepts and Methods

Examination: Module 5

Course code: BI760A

Credits for written examination: 1.5

Date: 19 December 2025

Examination time: 2 hours

Examination responsible: Zelmina Lubovac

Teachers concerned: 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: 0-15 = F; 16-18 = E; 19-21 = D; 22-24 = C; 25-27 = B; 28-30 = A

Examination results should be made public within 18 working days

Good luck!

Total number of pages: 6

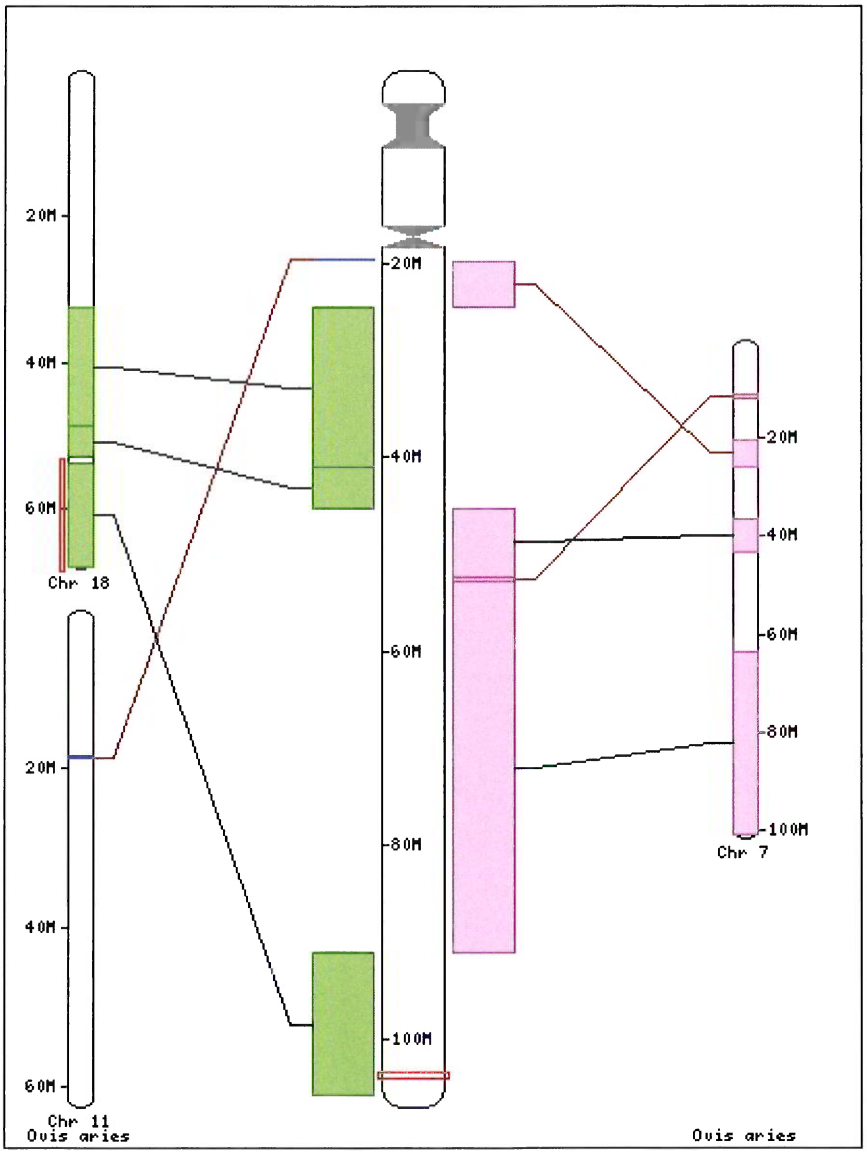
UNIVERSITY
OF SKÖVDE

Question 1 (6p)

The figure below shows a synteny map from Ensembl. The synteny map was generated by finding the gene entry for the human AKT1 gene, which is located on chromosome 14, and then creating a synteny map for the region where AKT1 is located. The comparison species is sheep, *Ovis aries*.

1a) Describe the concept of "synteny" in general, including the definition of the term "synteny" and a description of what we can learn from studying synteny between different genomes. **(3p)**

1b) Analyze the synteny map between human chromosome 14 and sheep. The analysis should mention which chromosomes in the sheep genome that the different regions of human chromosome 14 have synteny with, and the probable location of the sheep ortholog of the human AKT1 gene. **(3p)**



UNIVERSITY
OF SKÖVDE

Question 2 (6p)

For each of the following claims about Ensembl, 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) The Ensembl genome database holds only data about the human genome. **(1p)**

2b) There are a number of different types of entries in the Ensembl database, such as gene entry, protein entry, transcript entry, etc. **(1p)**

2c) There is only one way to access the contents of the Ensembl database - by viewing it in the genome browser. **(1p)**

2d) You can find data about single-nucleotide polymorphisms (SNPs) in Ensembl. **(1p)**

2e) The Ensembl database was founded in the 1970s. **(1p)**

2f) Ensembl includes the genomes of model species, such as mouse, fruitfly and zebrafish. **(1p)**

UNIVERSITY
OF SKÖVDE

Question 3 (6p)

Gene prediction can be a very difficult or relatively easy task, depending in various factors. Explain the following:

3a) Gene prediction in prokaryote genomes is generally easier than gene prediction in eukaryote genomes. Why? (**3p**)

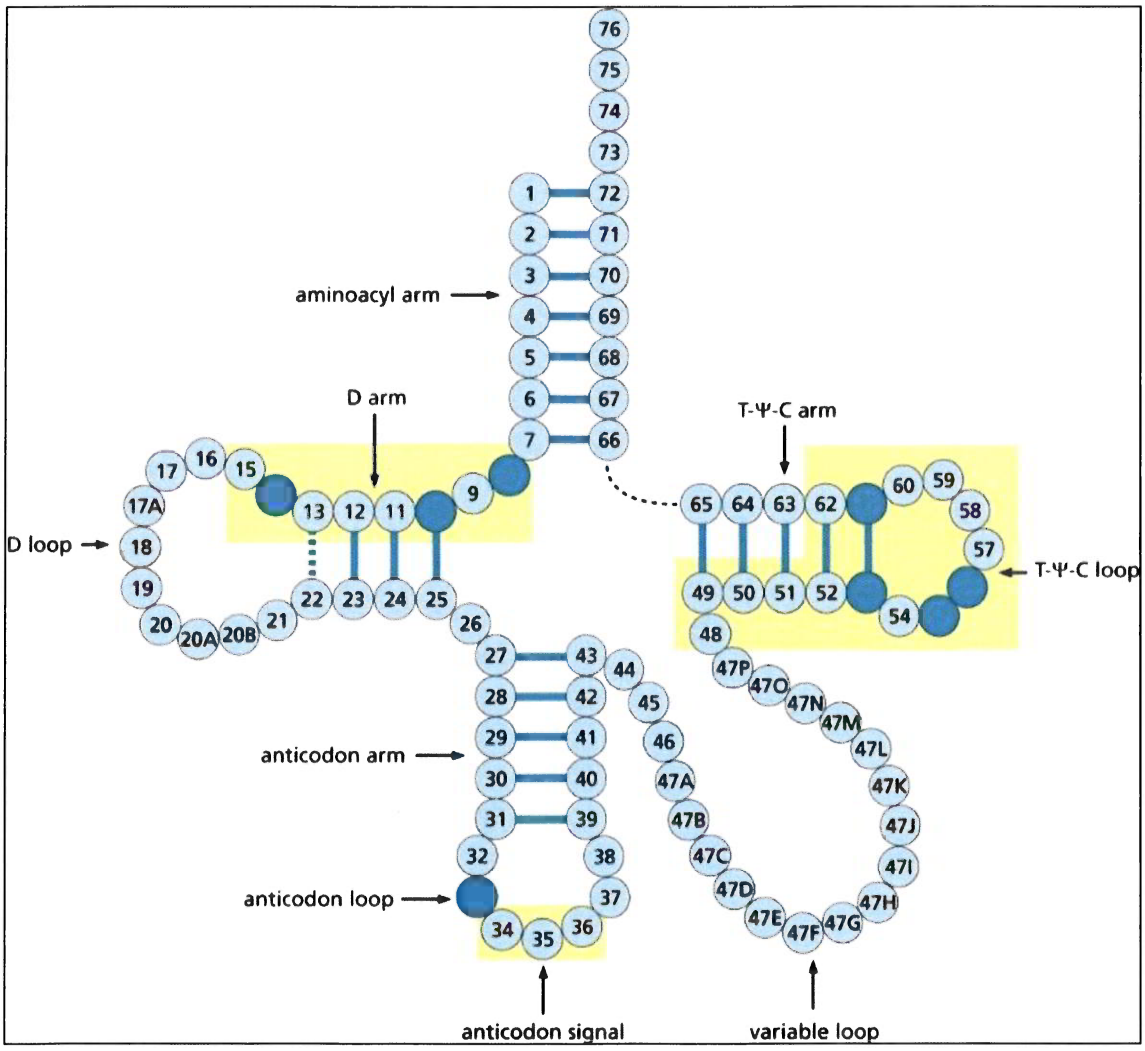
3b) Predicting different types of genes can be easier or more difficult. It is, for example, easier to predict tRNA genes than protein coding genes. Why? **(3p)**

UNIVERSITY
OF SKÖVDE

Question 4 (6p)

The tRNAscan algorithm predicts tRNA genes using a rules-based approach. The figure below is a schematic drawing of the structure of a tRNA molecule. When answering the question below, you can refer to this figure to illustrate and exemplify your explanations.

Describe <u>two</u> of the rules that tRNAscan applies when classifying a sequence as either being a tRNA gene sequence or not. You can get up to **3p** for each described rule (depending on correctness and level of detail), for a total of maximum **6p**.



UNIVERSITY
OF SKÖVDE

Question 5 (6p)

The figure shows two Markov models used to analyze DNA sequences. The model marked with + represents CpG island regions, and the model marked with – represents non-CpG island regions. The tables show transition probabilities between nucleotides (A, C, G, T).

Which of the following statements best explain how the models can be used to decide whether the sequence ACGTCG is more likely to come from a CpG island region? For each of the following claims, state if the claim is true or false. Just writing (for each claim) "true" or "false" is sufficient.

5a) The probability of the sequence is calculated by multiplying the transition probabilities for each consecutive nucleotide in the sequence. **(1p)**

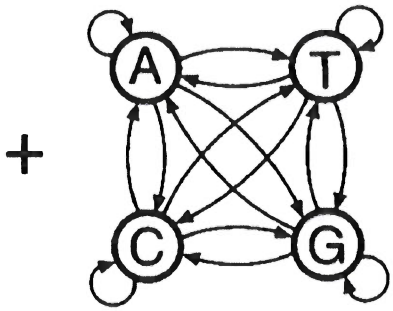
5b) The sequence is more likely to originate from the model in which transitions such as C → G have higher probabilities. **(1p)**

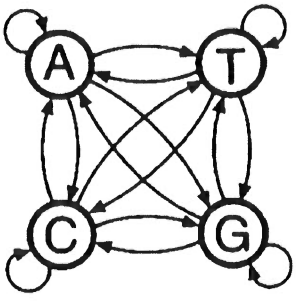
5c) Each nucleotide in the sequence is evaluated independently, without considering the previous nucleotide. **(1p)**

5d) The sequence is compared against both models, and the model giving the higher overall probability is chosen **(1p)**

5e) CpG island regions typically show higher transition probabilities between C and G than non-CpG regions. **(1p)**

5f) Only the first nucleotide of the sequence is needed to determine whether it comes from a CpG island. **(1p)**





+	A	C	G	T
A	0.180	0.274	0.426	0.120
C	0.171	0.368	0.274	0.188
G	0.161	0.339	0.375	0.125
T	0.079	0.355	0.384	0.182

–	A	C	G	T
A	0.300	0.205	0.285	0.210
C	0.322	0.298	0.078	0.302
G	0.248	0.246	0.298	0.208
T	0.177	0.239	0.292	0.292