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: 28 October 2024

Examination time: 2 hours

Examination responsible: Zelmina Lubovac

Teachers concerned: Björn Olsson

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

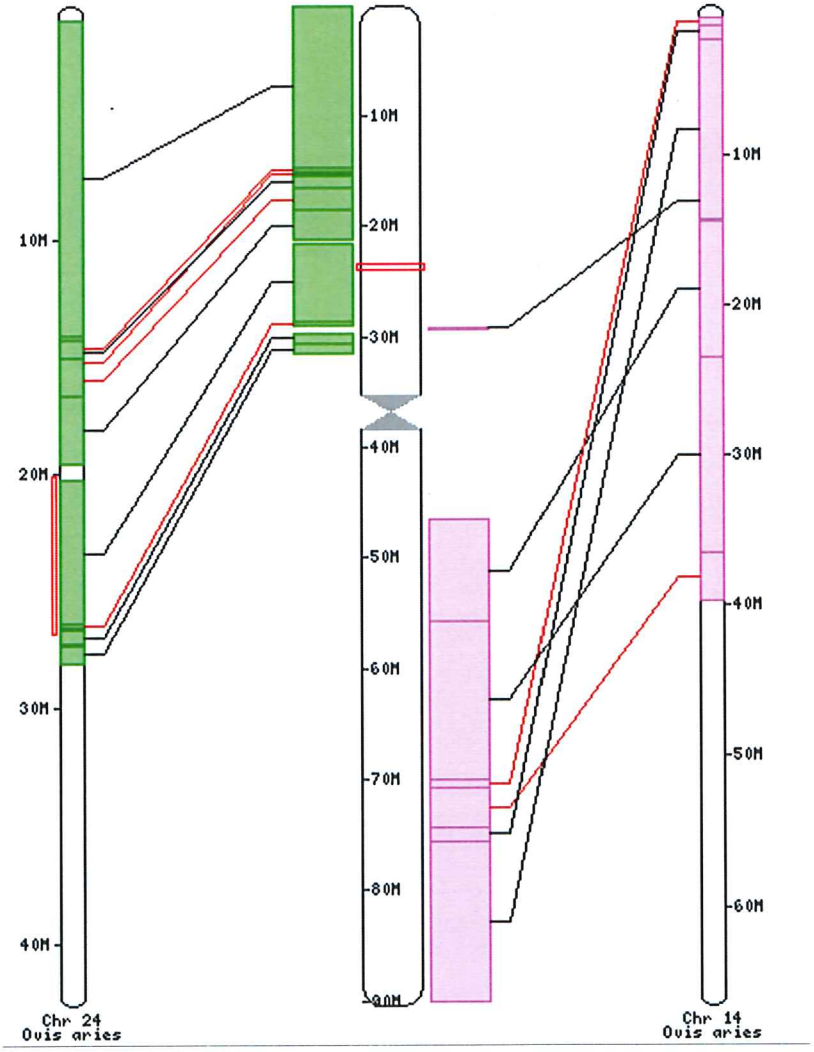
UNIVERSITY
OF SKÖVDE

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 PALB2 gene, which is located on chromosome 16, and then creating a synteny map for the region where PALB2 is located. The comparison species is sheep, *Ovis aries*.

1a) Define the term "synteny" and explain its importance in comparative genomics. What can we learn from studying synteny between different genomes? **(3p)**

1b) Analyze the synteny map between human chromosome 16 and sheep. Your analysis should identify which chromosomes in the sheep genome show synteny with the different regions of human chromosome 16. Additionally, mention the probable location of the sheep ortholog of the human PALB2 gene, based on the synteny map provided. **(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) Ensembl includes genome information for model organisms such as mouse, fruitfly, and zebrafish. **(1p)**

2b) BioMart is a tool within Ensembl that allows users to perform custom queries and export data. **(1p)**

2c) The Ensembl database can be accessed solely through the Ensembl genome browser. **(1p)**

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

2e) The Ensembl Variant Effect Predictor (VEP) is a tool in Ensambl that only works with human genome variants. **(1p)**

2f) Ensembl provides access to multiple types of entries, such as gene, transcript, and protein entries. **(1p)**

UNIVERSITY
OF SKÖVDE

Question 3 (6p)

ORF Finder is a tool used to identify potential Open Reading Frames (ORFs) within a nucleotide sequence.

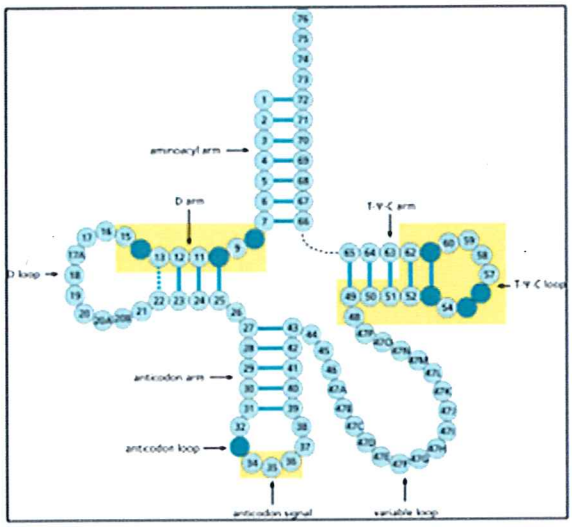
3a) Explain how the ORF Finder algorithm works, and discuss how changing the minimal ORF length affects its predictions. In your answer, include a description of how ORF Finder identifies Open Reading Frames (ORFs), and the trade-offs between setting a high and low minimum ORF length. **(3p)**

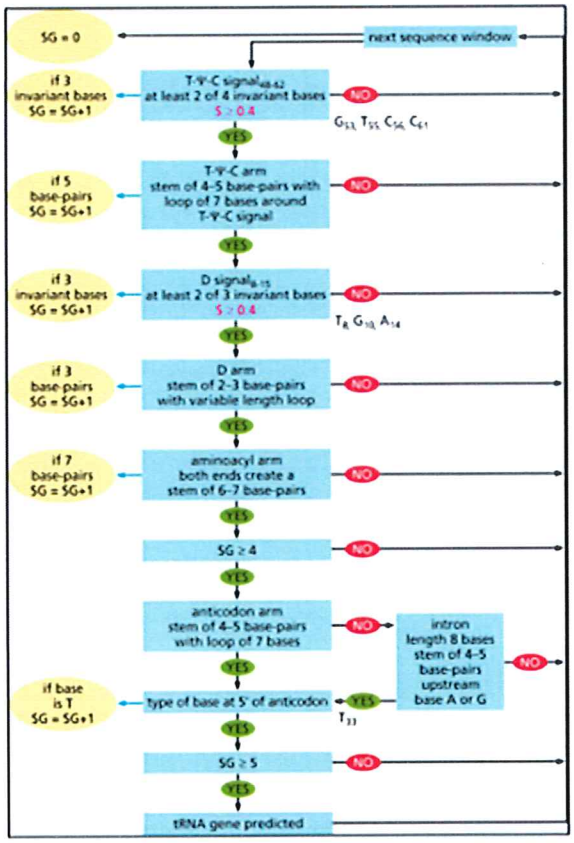
3b) Describe how you could confirm an ORF prediction made by ORF Finder using additional tools. Your answer should include examples of two approaches or tools that can be used to strengthen the prediction from ORF Finder. The accuracy and level of detail in your explanation will influence the points awarded. **(3p)**

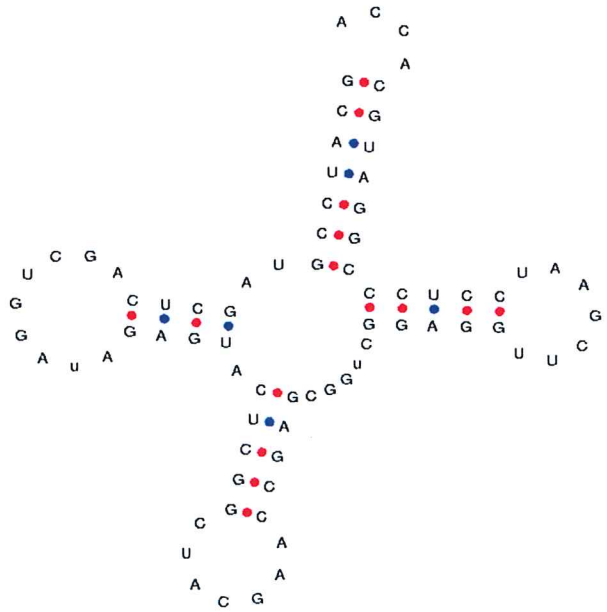
UNIVERSITY
OF SKÖVDE

Question 4 (6p)

The figures show a schematic representation of a tRNA molecule's general structure, the tRNAscan algorithm's classification rules, and a predicted tRNA molecule. Describe **three** rules that tRNAscan applies to determine whether the sequence is a tRNA gene or not. Refer to how these rules are applied to the predicted tRNA in the figure, using it to support your explanations. You can earn up to **2p** for each described rule (based on accuracy and detail), for a total of up to **6p**.





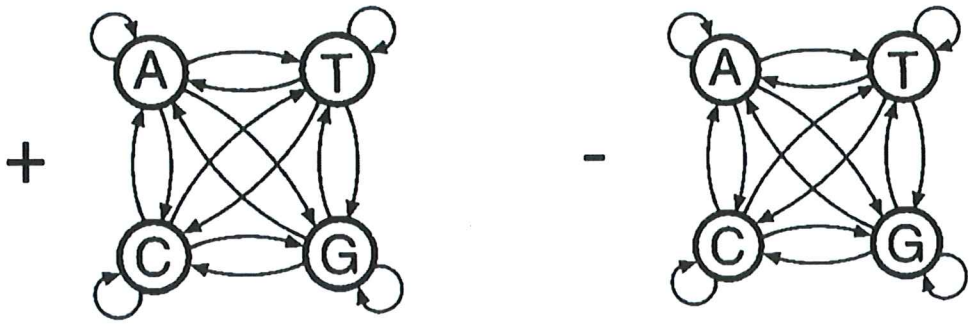


UNIVERSITY
OF SKÖVDE

Question 5 (6p)

5a) Explain what a Markov model is and describe its key components. Refer to the figures below to illustrate and explain the fundamental "building blocks" of a Markov model. **(3p)**

5b) The CpG island model is a so-called first-order Markov model. Explain what this means. Additionally, discuss the advantages and disadvantages of using higher-order Markov models for gene prediction. **(3p)**



+	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