UNIVERSITY OF ESWATINI

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF BIOLOGICAL SCIENCES

MAIN EXAMINATION PAPER 2020/2021

COURSE CODE: BIO451

TITLE OF PAPER: BIOINFORMATICS

TIME ALLOWED: **THREE** (3) HOURS

INSTRUCTIONS: NUMBERS IN BRACKETS DENOTE THE NUMBER OF MARKS

THIS PAPER COMPRISES OF **TWO SECTIONS and ONE ANNEX.**

SECTION A COMPRISES OF EIGHT (8) QUESTIONS.

SECTION B COMPRISES OF SEVEN (7) QUESTIONS.

ANSWER ALL QUESTIONS IN SECTIONS A AND B.

TOTAL MARKS AVAILABLE: 70

NO ADDITIONAL MATERIAL (E.G. NOTES, BOOKS, PHONES, SMART WATCHES ETC) MAY BE TAKEN INTO THE EXAMINATION.

THE USE OF A SCIENTIFIC CALCULATOR IS PERMITTED.

DO NOT OPEN THIS PAPER UNTIL PERMISSION HAS BEEN GRANTED BY THE CHIEF INVIGILATOR

Section A

Total marks available: 35

A poacher has been arrested by the police in possession of several processed animal parts. Unfortunately, it is not possible to identify the animal(s) just by looking at them. The police have sent samples of the processed animal parts to you for forensic analysis. You have extracted and sequenced the samples using your next-generation sequencer.

AQ1: State what **type of molecule** and **file format** of the data you expect to receive from the sequencer.

[2]

AQ2 Describe what you would do with the sequence data to help identify the organism(s). In your answer, include what databases you would use.

[4]

AQ3: The results have come back for one of the samples. Explain what the four columns inside the black oval (Query cover, E value, Per. Ident and Accession) tells you about your search.

Descriptions Graphic Summary   Alignments   Taxonomy

Sequences producing significant alignments

Download ∨   **New Manage Columns** ∨   Show 100 ∨ ?

☐ select all *0 sequences selected*

Description	Scientific Name	Common Name	Taxid	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Tragelaphus strepsiceros isolate PHC11 mitochondrion, complete genome	Tragelap...	greater ...	9946	3195	3195	100%	0.0	99.77%	16382	JH632708.1
Tragelaphus spekii mitochondrion, complete genome	Tragelap...	Sitatunga	69298	2809	2809	100%	0.0	95.75%	16408	EF536357.1
Tragelaphus oryx isolate PHC13 mitochondrion, complete genome	Tragelap...	eland	9945	2776	2776	100%	0.0	95.42%	16397	JH632704.1
Taurotragus derbianus mitochondrion, complete genome	Taurotra...	giant el ...	303930	2776	2776	100%	0.0	95.46%	16399	EF536354.1
Tragelaphus eurycerus isolate SUN8 mitochondrion, complete genome	Tragelap...	bongo	69297	2771	2771	100%	0.0	95.36%	16349	JH632703.1
Tragelaphus scriptus isolate MBP11 mitochondrion, complete genome	Tragelap...	bushbuck	66440	2748	2748	100%	0.0	95.18%	16384	JH632706.1
Tragelaphus scriptus isolate PhC17 mitochondrion, complete genome	Tragelap...	bushbuck	66440	2732	2732	100%	0.0	95.01%	16402	JH632707.1
Tragelaphus buxtoni mitochondrion, complete genome	Tragelap...	mountai...	69296	2723	2723	100%	0.0	94.80%	16601	NC_038064.1
Tragelaphus scriptus isolate CAR31 mitochondrion, complete genome	Tragelap...	bushbuck	66440	2553	2553	99%	0.0	93.26%	16401	JH632705.1
Tragelaphus angasii isolate MBP15 mitochondrion, complete genome	Tragelap...	nyala	66437	2514	2514	100%	0.0	92.77%	16399	JH632702.1
Tragelaphus imberbis mitochondrial complete genome	Tragelap...	lesser k...	9947	2507	2507	100%	0.0	92.62%	16401	LN873135.1
Tragelaphus imberbis mitochondrion, complete genome	Tragelap...	lesser k...	9947	2507	2507	100%	0.0	92.62%	16405	EF536356.1
Tragelaphus imberbis mitochondrial 12S and 16S ribosomal RNAs and Phe-, Val- and Leu-tRNA genes	Tragelap...	lesser k...	9947	2507	2507	100%	0.0	92.62%	2686	M86493.1
Boselaphus tragocamelus mitochondrion, complete genome	Boselap...	nilgai	9917	2318	2318	100%	0.0	90.73%	16322	EF536350.1

[2]

AQ4: Based on the results above, what accession number would you pursue? [1]

AQ5: After further investigation, using GenBank, you come across this result (see Annex 1) for a second sample. Identify the following:

    a) The size and type of molecule [2]
    b) The full name of the GenBank Division [2]
    c) The protein name [1]
    d) Organism (common and scientific) [3]
    e) Name of the gene encoding for this protein [2]

f) The accession number of the gene encoding for this protein [2]

AQ6: You have sent your results to the legal teams handling the case. Unfortunately, there is some dispute over the results from the second sample (AQ5). Both the prosecution and defense teams have requested some further analysis. Specifically, they request you to analyse the sample that they gave you against the very same organism identified in from Annex 1. **Explain how you could go about getting a tissue sample from the organism (in Annex 1) to run your investigation.**
[5]

AQ6: Describe **how** you would use **bioinformatics** to determine the structure and function of the protein (from AQ5), including the database(s) you would use. [5]

AQ7: List two reasons for determining the structure of proteins. [2]

AQ8: Describe the key feature(s) of the protein below. [2]
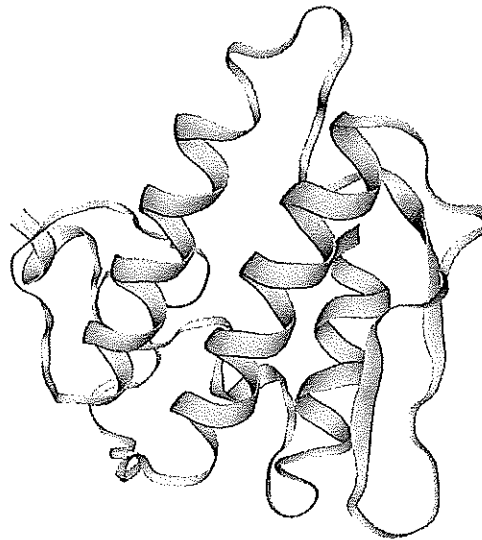
Section B   Total marks available: 35

BQ1: Which BLAST application would you use to identify the following sequences:

    a) AGAAGTTCCAGATCATTCCAAGA [1]
    b) AUCCCAGGGUUCACCAGAGCCAGGAGA [1]
    c) KAGTAKCATVADDCHUABCCAGAAGNMWSS [1]
    d) VLPWGQMSFWGATVITNLLSAIPYIGNTLV [1]

BQ2: Identify which of the following sequence(s) would you enter into the tBLASTn database: [1]

    a) MGGNTAKCRRADDCHUABCCAGAAGNMWSS
    b) CGAGAUUAUCCCAGGGUUCACCAGAGA
    c) TAGGTTCCCAGATCATTCCAAGA
    d) ATPYISSAFVGYVLPWGPTVQMSFWGATVITNLLS

BQ3: What is the default scoring matrix for BLASTp? [1]

BQ4: Assuming a scoring method: match = +1, mismatch = -2, and gap = −1, using the Needleman-Wunsch algorithm:

a) Initialize the scoring matrix below. [2]

		T	C	T	G	A	T	T	C	T	C
T											
C											
G											
C											
A											
T											
C											
A											
C											
C											

b) Calculate the score of the matrix and fill the traceback matrix. Use arrows to indicate how you determined the path forward. Use circles and arrows to determine your trace back. [5]

c) Deduce the best alignment based on the traceback matrix. [3]

BQ5: Assuming a scoring method: match = +1, mismatch = -2, and gap = -2, using the Smith and Waterman algorithm:

a) Initialize the scoring matrix below. [2]

		T	C	T	G	A	T	T	C	T	C	C
T												
A												
G												
C												
A												
T												
T												
A												
G												
C												

b) Calculate the score of the matrix and fill the traceback matrix. Use arrows to indicate how you determined the path forward. Use circles and arrows to determine your trace back. [5]

c) Deduce the best alignment based on the traceback matrix. [3]

d) Were there any sub-optimal alignments? Explain your answer. [2]

BQ6: Align the following two sequences using a dot plot [3]

- TCTGATTCTC
- TCGCATCAC

BQ7: You have four sets of nucleic acid sequences of very different lengths to align. Which type of algorithm(s) would be most appropriate? Explain your answer and give ONE advantage for your algorithm(s). [4]

END OF EXAMINATION