

UNIVERSITY OF ESWATINI
FACULTY OF SCIENCE & ENGINEERING
DEPARTMENT OF BIOLOGICAL SCIENCES
MAIN EXAMINATION PAPER 2018/2019

COURSE CODE: BIO451

TITLE OF PAPER: BIOINFORMATICS

TIME ALLOWED: THREE HOURS

INSTRUCTIONS:

1. ANSWER ALL QUESTIONS
2. CANDIDATES MAY USE SCIENTIFIC CALCULATORS

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

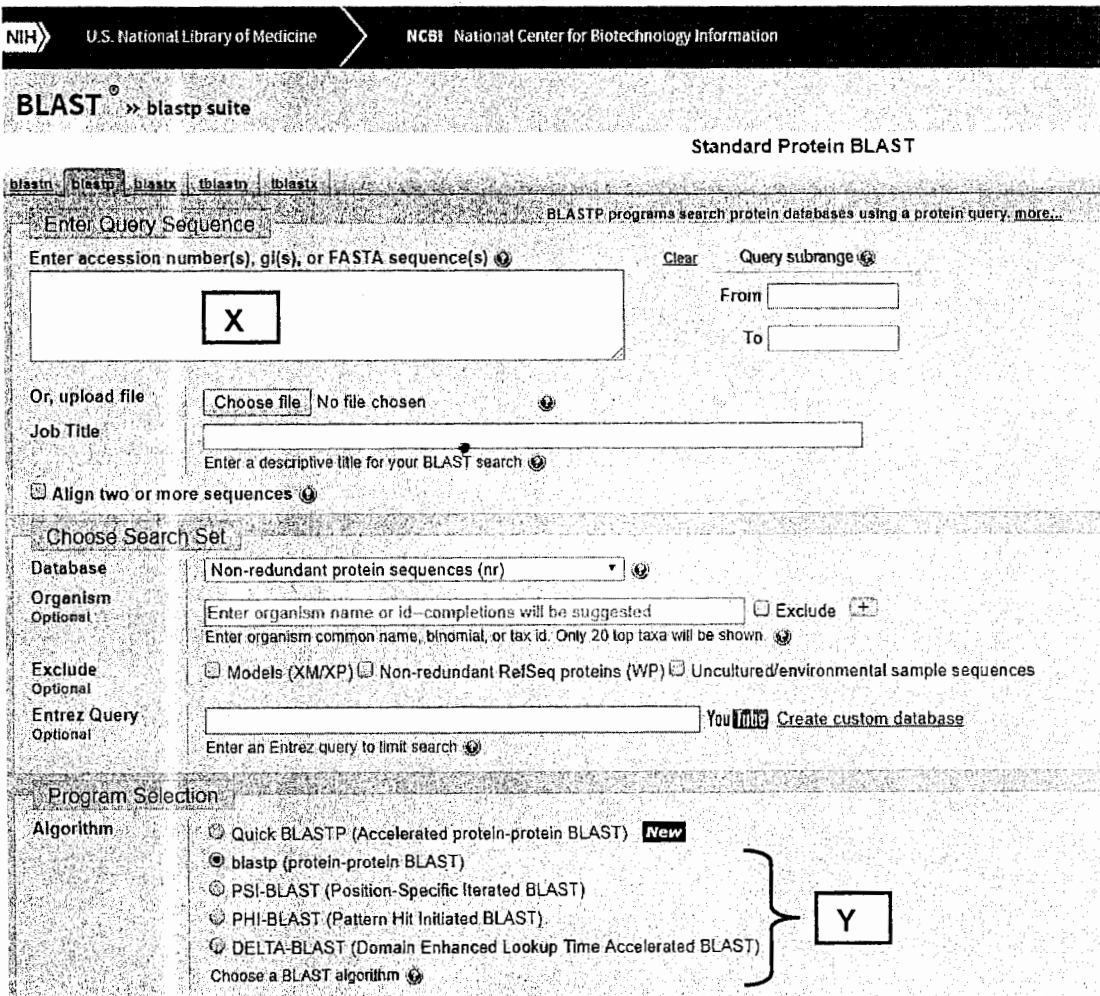
QUESTION 1:
BIOINFORMATICS APPLICATION & DATABASES

- (a) Briefly explain the role of public biological databases. [2 marks]
- (b) Distinguish between primary and secondary databases. [3 marks]
- (c) Describe the following databases:
 - (i) GenBank, ENA and DDBJ [2 marks]
 - (ii) UniprotKB, [2 marks]
 - (iii) Refseq, [2 marks]
 - (iv) OMIM [2 marks]
 - (v) MEDLINE [2 marks]
- (c) A recent enteric disease epidemic in the Manzini region of Eswatini has been reported to be caused by an unknown species of bacteria whose treatment is not yet known. A pure culture of this pathogen is brought to your consultancy lab for identification with a view to suggesting probable treatment for this disease. Describe the molecular biology and bioinformatics tools you can use to identify the species and/or genus this bacteria species may belong to. [10 marks]

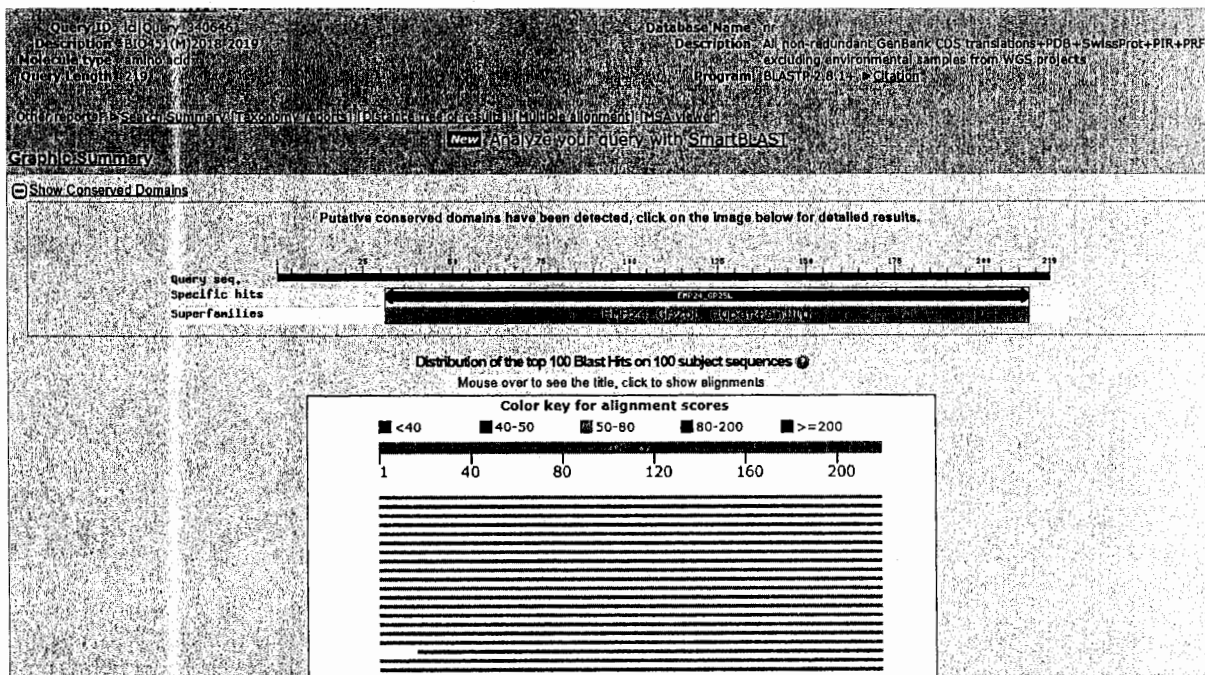
[Total marks = 25]

QUESTION 2:
NUCLEIC ACID/PROTEIN SEQUENCE ANALYSIS, MULTIPLE-SEQUENCE ALIGNMENT & HOMOLOGY/PATTERN-BASED SEARCH ALGORITHMS

- (a) Answer the following questions based on the screenshot of a BLAST page from NCBI website below.



- (i) Describe the nature and format of the sequence in the panel labelled X. [2 marks]
- (ii) Discuss the principle of the different algorithms in the section marked Y. [8 marks]
- (b) (i) Explain the difference between PAM and BLOSUM scoring matrices. [6 marks]
- (ii) Explain the significance of "62" on the BLOSUM62 matrix. [1 mark]
- (c) Explain the difference between the following sequence alignment concepts:
 - (i) Smith–Waterman and Needleman–Wunsch algorithms, [4 marks]
 - (ii) Progressive and iterative MSA, [2 marks]
 - (iii) Global and local MSA. [2 marks]
- (d) Shown below is a typical graphic summary of BLAST result. Identify important information can you extract from it. [3 marks]



- (e) Shown below is a screenshot of a typical BLAST HIT result. Identify and explain the labelled items. [6 marks]

Sequences producing significant alignments:

Select: All None Selected: 0

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/> transmembrane_emo24 domain-containing protein_10 precursor (Rattus norvegicus)	450	450	100%	4e-160	100%	NP_145319.1
<input checked="" type="checkbox"/> transmembrane_emo24 domain-containing protein_10 (Meriones unguiculatus)	444	444	100%	1e-157	99%	XP_021488318.1
<input checked="" type="checkbox"/> transmembrane_emo24 domain-containing protein_10 (Microtus ochrogaster)	439	439	100%	2e-155	98%	XP_025341425.1
<input type="checkbox"/> PREDICTED: transmembrane_emo24 domain-containing protein_10 isoform X1 (Citratellus orisicus)	436	438	100%	2e-155	98%	XP_007341584.2
<input type="checkbox"/> PREDICTED: transmembrane_emo24 domain-containing protein_10 (Peromyscus maniculatus bairdii)	437	437	100%	5e-155	98%	XP_056929632.1
<input checked="" type="checkbox"/> transmembrane_emo24 domain-containing protein_10 precursor (Mus musculus)	434	434	100%	6e-154	98%	NP_081951.3
<input checked="" type="checkbox"/> transmembrane protein Tmo21 precursor (Mesocricetus auratus)	434	434	100%	8e-154	97%	NP_601268517.1
<input checked="" type="checkbox"/> transmembrane_emo24-like trafficking protein_10 (yeast) (Hus musculus)	434	434	100%	1e-153	98%	EDL02295.1
<input checked="" type="checkbox"/> transmembrane_emo24 domain-containing protein_10 (Mus caroli)	433	433	100%	2e-153	97%	XP_021634520.1
<input checked="" type="checkbox"/> transmembrane_emo24 domain-containing protein_10 (Mus eabadi)	432	432	100%	4e-153	97%	XP_021627271.1
<input checked="" type="checkbox"/> PREDICTED: transmembrane_emo24 domain-containing protein_10 (Nannosialia gadii)	432	432	100%	7e-153	96%	XP_009347939.1
<input checked="" type="checkbox"/> transmembrane_emo24-like trafficking protein_10 (yeast) (Hus musculus)	431	431	100%	2e-152	97%	AAH64755.1

↑

A

↑

B

↑

C

(f) Shown below is a screenshot of a typical BLAST alignment result. Identify and explain the labelled items. [6 marks]

Download v GenPept Graphics

transmembrane emp24 domain-containing protein 10 precursor [Rattus norvegicus]
 Sequence ID: [NP_445919.1](#) Length: 219 Number of Matches: 1
 > See 3 more title(s)

Range 1: 1 to 219 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
450 bits(1158)	4e-160	Compositional matrix adjust.	219/219(100%)	219/219(100%)	0/219(0%)

Query 1 MSGLSGPLSWPGPLLSALLFLFLGPPSSVLGISFHLPVNSRKCLREEIHKDLLVTGAYEI 60
 MSGLSGPLSWPGPLLSALLFLFLGPPSSVLGISFHLPVNSRKCLREEIHKDLLVTGAYEI
 Sbjct 1 MSGLSGPLSWPGPLLSALLFLFLGPPSSVLGISFHLPVNSRKCLREEIHKDLLVTGAYEI 60

Query 61 TDQSGGAGGLRTHLKITDSAGHILYAKEDATKGF AFTTEDYDMFEVCFESKGTGRIPDQ 120
 TDQSGGAGGLRTHLKITDSAGHILYAKEDATKGF AFTTEDYDMFEVCFESKGTGRIPDQ
 Sbjct 61 TDQSGGAGGLRTHLKITDSAGHILYAKEDATKGF AFTTEDYDMFEVCFESKGTGRIPDQ 120

Query 121 LVILDMKHGVEAKNYEEIAKVEKLPLEVELRRLEDLSEIVNDFAYMKKREEMRDNE 180
 LVILDMKHGVEAKNYEEIAKVEKLPLEVELRRLEDLSEIVNDFAYMKKREEMRDNE
 Sbjct 121 LVILDMKHGVEAKNYEEIAKVEKLPLEVELRRLEDLSEIVNDFAYMKKREEMRDNE 180

Query 181 STNTRVLYFSIFSFMFLIGLATWQVYLRFFKAKKLIE 219
 STNTRVLYFSIFSFMFLIGLATWQVYLRFFKAKKLIE
 Sbjct 181 STNTRVLYFSIFSFMFLIGLATWQVYLRFFKAKKLIE 219

Labels: A points to the 'Identities' column; B points to the 'Positives' column; C points to the alignment sequence blocks.

[Total marks = 40]

QUESTION 3:

GENE & PROTEIN MOTIFS/DOMAIN PREDICTION AND GENOME ANNOTATION

- (a) Briefly discuss the advantages and disadvantages of having long versus short k-mers in genome assembly [6 marks]
- (b) Describe using any bioinformatics tool, how CpG islands in a genome can be detected. [5 marks]
- (c) Discuss the use of the Burrows-Wheeler transform in genome assembly, including its algorithmic complexity. [6 marks]

[Total marks = 17]

QUESTION 4:

MOLECULAR PHYLOGENETICS

- (a) Explain why a heuristic approach is required for constructing phylogenetic trees when the number of taxa is large. [6 marks]
- (b) Explain the difference between the Jukes-Cantor and Kimura 2-parameter models when inferring a phylogenetic tree from DNA sequences. [6 marks]
- (c) What is an outgroup and why is it important to include one or more outgroups in phylogenetic analysis? [6 marks]

[Total marks = 18]