

UNIVERSITY OF ESWATINI
FACULTY OF SCIENCE AND ENGINEERING
DEPARTMENT OF BIOLOGICAL SCIENCES
SUPPLEMENTARY 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 EIGHT (8) QUESTIONS and ONE ANNEX.
ANSWER ALL THE QUESTIONS.

TOTAL MARKS AVAILABLE: 75

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

Total marks available: 75

You are working in a molecular laboratory and have a student intern working with you. Unfortunately, when the student was converting the abi files into fasta formats, the student mixed up all the results from today's analysis carried out on the next generation sequencer. The sequences are as follows:

Sequence 1: ATGACAAACATTCGAAAAAC

Sequence 2: ATGACCAACATCCGAAAAAC

Sequence 3: GCCGCCTTTTCCTCCGTTGC

Sequence 4: ATGAAAAACATACGAAAAAC

Sequence 5: TTGACAAACATTCGAAAATC

Sequence 6: TCTCATCCTCTACTTAAAAT

The electricity has also switched off, and you need to urgently analyse your results because your laboratory manager needs to present these results at an important meeting. Your only choice is to analyse the sequences without the aid of a computer. You look through your computer and you find sequences for cytochrome b *Rattus norvegicus* and *Acomys russatus* that you had previously identified.

Acomys Russatus: ATGAAAAACATACGAAAATC*Rattus Norvegicus*: ACGCGGGATGACAAACATCC

Q1: If electricity were available, describe what you would do with the sequence data to help identify the organism(s). In your answer, include what databases you would use. [5]

Q2: Unfortunately, the electricity is not back, and you are forced to analyse the data without a computer. The lab manager wants to know which sequences (including the cytochrome B sequences) are related to each other for taxonomic purposes. Prepare a phylogenetic tree using the six sequences unknown sequences and the two known sequences. Your answer must detail (show) each step of your process you took in constructing your tree, including identifying the name and type of the algorithm you used. [20]

Q3: From your phylogenetic tree, identify which sequences are most closely related? [3]

Q4: Name one advantage and one disadvantage of the method you used. [2]

The student intern has come to you and presented you with another 6 sequences.

Sequence 7: MTNIRKTHPLA

Sequence 8: MTNILRKTHPL

Sequence 9: AAFSSVAHICM

Sequence 10: LAAMKNMRKTH

Sequence 11: MTNIRKSHPLS

Sequence 12: SHPLLKIINHS

Q5: Your lab manager would like to know

- a) if there is any homology between these 6 sequences, and if so, [1]
- b) calculate the alignment score for each pair of sequences using the PAM250 matrix (see Annex below) [15]
- c) which sequences are homologous (explain your answer) [4]

Q6: Your lab manager would like to know the similarity between sequences 10 and 11 only. Using the following scoring matrix (match = +3, mismatch = -2, gap = -1),

- a) initialize the scoring matrix [1]
- b) score the matrix [5]
- c) calculate the optimal alignment score [2]
- d) indicate the traceback [2]
- e) deduce the optimal alignment [3]
- f) if appropriate, deduce any sub-optimal alignments [2]

Q7: With the aid of ONE example, explain the difference between local and global alignment. [3]

Q8: Describe what are hidden Markov models, and when and why we use them in bioinformatics. [7]

END OF EXAMINATION

Annex 1

| | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V | B | Z |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| A | 2 | -2 | 0 | 0 | -2 | 0 | 0 | 1 | -1 | -1 | -2 | -1 | -1 | -3 | 1 | 1 | 1 | -6 | -3 | 0 | 0 | 0 |
| R | -2 | 6 | 0 | -1 | -4 | 1 | -1 | -3 | 2 | -2 | -3 | 3 | 0 | -4 | 0 | 0 | -1 | 2 | -4 | -2 | -1 | 0 |
| N | 0 | 0 | 2 | 2 | -4 | 1 | 1 | 0 | 2 | -2 | -3 | 1 | -2 | -3 | 0 | 1 | 0 | -4 | -2 | -2 | 2 | 1 |
| D | 0 | -1 | 2 | 4 | -5 | 2 | 3 | 1 | 1 | -2 | -4 | 0 | -3 | -6 | -1 | 0 | 0 | -7 | -4 | -2 | 3 | 3 |
| C | -2 | -4 | -4 | -5 | 12 | -5 | -5 | -3 | -3 | -2 | -6 | -5 | -5 | -4 | -3 | 0 | -2 | -8 | 0 | -2 | -4 | -5 |
| Q | 0 | 1 | 1 | 2 | -5 | 4 | 2 | -1 | 3 | -2 | -2 | 1 | -1 | -5 | 0 | -1 | -1 | -5 | -4 | -2 | 1 | 3 |
| E | 0 | -1 | 1 | 3 | -5 | 2 | 4 | 0 | 1 | -2 | -3 | 0 | -2 | -5 | -1 | 0 | 0 | -7 | -4 | -2 | 3 | 3 |
| G | 1 | -3 | 0 | 1 | -3 | -1 | 0 | 5 | -2 | -3 | -4 | -2 | -3 | -5 | 0 | 1 | 0 | -7 | -5 | -1 | 0 | 0 |
| H | -1 | 2 | 2 | 1 | -3 | 3 | 1 | -2 | 6 | -2 | -2 | 0 | -2 | -2 | 0 | -1 | -1 | -3 | 0 | -2 | 1 | 2 |
| I | -1 | -2 | -2 | -2 | -2 | -2 | -2 | -3 | -2 | 5 | 2 | -2 | 2 | 1 | -2 | -1 | 0 | -5 | -1 | 4 | -2 | -2 |
| L | -2 | -3 | -3 | -4 | -6 | -2 | -3 | -4 | -2 | 2 | 6 | -3 | 4 | 2 | -3 | -3 | -2 | -2 | -1 | 2 | -3 | -3 |
| K | -1 | 3 | 1 | 0 | -5 | 1 | 0 | -2 | 0 | -2 | -3 | 5 | 0 | -5 | -1 | 0 | 0 | -3 | -4 | -2 | 1 | 0 |
| M | -1 | 0 | -2 | -3 | -5 | -1 | -2 | -3 | -2 | 2 | 4 | 0 | 6 | 0 | -2 | -2 | -1 | -4 | -2 | 2 | -2 | -2 |
| F | -3 | -4 | -3 | -6 | -4 | -5 | -5 | -5 | -2 | 1 | 2 | -5 | 0 | 9 | -5 | -3 | -3 | 0 | 7 | -1 | -4 | -5 |
| P | 1 | 0 | 0 | -1 | -3 | 0 | -1 | 0 | 0 | -2 | -3 | -1 | -2 | -5 | 6 | 1 | 0 | -6 | -5 | -1 | -1 | 0 |
| S | 1 | 0 | 1 | 0 | 0 | -1 | 0 | 1 | -1 | -1 | -3 | 0 | -2 | -3 | 1 | 2 | 1 | -2 | -3 | -1 | 0 | 0 |
| T | 1 | -1 | 0 | 0 | -2 | -1 | 0 | 0 | -1 | 0 | -2 | 0 | -1 | -3 | 0 | 1 | 3 | -5 | -3 | 0 | 0 | -1 |
| W | -6 | 2 | -4 | -7 | -8 | -5 | -7 | -7 | -3 | -5 | -2 | -3 | -4 | 0 | -6 | -2 | -5 | 17 | 0 | -6 | -5 | -6 |
| Y | -3 | -4 | -2 | -4 | 0 | -4 | -4 | -5 | 0 | -1 | -1 | -4 | -2 | 7 | -5 | -3 | -3 | 0 | 10 | -2 | -3 | -4 |
| V | 0 | -2 | -2 | -2 | -2 | -2 | -2 | -1 | -2 | 4 | 2 | -2 | 2 | -1 | -1 | -1 | 0 | -6 | -2 | 4 | -2 | -2 |
| B | 0 | -1 | 2 | 3 | -4 | 1 | 3 | 0 | 1 | -2 | -3 | 1 | -2 | -4 | -1 | 0 | 0 | -5 | -3 | -2 | 3 | 2 |
| Z | 0 | 0 | 1 | 3 | -5 | 3 | 3 | 0 | 2 | -2 | -3 | 0 | -2 | -5 | 0 | 0 | -1 | -6 | -4 | -2 | 2 | 3 |
| . | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V | B | Z |