



TECHNICAL UNIVERSITY OF MOMBASA

FACULTY OF APPLIED AND HEALTH SCIENCES

DEPARTMENT OF PURE & APPLIED SCIENCES

UNIVERSITY EXAMINATION FOR:

BMLS

PAPER 2

ABT 4405: BIOINFORMATICS

END OF SEMESTER EXAMINATION

SERIES: AUGUST 2019

TIME: 2 HOURS

DATE: AUGUST 2019

Instructions to Candidates

You should have the following for this examination

-Answer Booklet, examination pass and student ID

This paper consists of **FIVE** questions. Attempt question ONE (Compulsory) and any other TWO questions.

Do not write on the question paper.

QUESTION ONE (30 Marks)

- a. Define the following terminologies as used in bioinformatics.
- | | |
|------------------------|--------|
| i. Biological database | 1 Mark |
| ii. Accession number | 1 Mark |
| iii. Biosis | 1 Mark |
- b. Give the meaning of the following terms:
- | | |
|-----------|--------|
| i. ASCII | 1 Mark |
| ii. PSSM | 1 Mark |
| iii. XML | 1 Mark |
| iv. UPGMA | 1 Mark |
- c. Describe the BSML file format. 7 Marks
- d. Outline any EIGHT limitations of bioinformatics. 8 Marks

- e. You are given the following pair of sequences:
1. DDBICEKDDATU
 2. DRBICEKADATM

Calculate similarity scores between the pair of sequences using an identity matrix (assume mismatch=0, and match=1). 2 Marks

- f. Describe the following primary protein sequence database. 6 Marks
- i. PDB
 - ii. TrEMBL

QUESTION TWO (20 Marks)

Explain the following BLAST programs as used in bioinformatics.

- i. tblastx 10 Marks
- ii. Blastx 10 Marks

QUESTION THREE (20 Marks)

- a. Explain the general guidelines to consider when designing primers. 8 Marks
- b. Describe FOUR bioinformatic tools you would use to design primers. 8 Marks
- c. Use the Standard Code for Information Interchange table below to solve the following.

	30	40	50	60	70	80	90	100	110	120
0		(2	<	F	P	Z	d	n	x
1)	3	=	G	Q	[e	o	y
2		*	4	>	H	R	\	f	p	z
3	!	+	5	?	I	S]	g	q	{
4	"	,	6	@	J	T	^	h	r	
5	#	-	7	A	K	U	_	i	s	}
6	\$.	8	B	L	V		j	t	~
7	%	/	9	C	M	W	a	k	u	DEL
8	&	0	:	D	N	X	b	l	v	
9	'	1	;	E	O	Y	c	m	w	

Given a character 'W', Calculate the corresponding Phred quality score based on fastq-solexa format. 4 Marks

QUESTION FOUR (20 Marks)

a. Using an illustration, explain the position specific iterative-BLAST (PSI-BLAST) approach. 8 Marks

b. Use the BLOSUM 62 and PAM250 substitution matrices below to solve the following questions

BLOSUM 62

PAM250

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

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

Given the two pairs of sequences below;

- 1. DWSDEVKA
- 2. YWSDEVKA

- 1. DEVKDWSA
- 2. DEKVDWSS

Calculate similarity scores between these sequences using:

- i. PAM250 matrix 4 Marks
- ii. Identity matrix (assume match =2, and mismatch = 0) 4 Marks
- iii. BLOSUM 62 matrix 4 Marks

QUESTION FIVE (20 Marks)

- a. Distinguish between ENZYME and REBASA databases. 14 Marks

- b. Describe THREE bioinformatic tools you would use to align two sequences. 6 Marks