TECHNICAL UNIVERSITY OF MOMBASA

FACULTY OF APPLIED AND HEALTH SCIENCES

DEPARTMENT OF PURE & APPLIED SCIENCES

UNIVERSITY EXAMINATION FOR:

BTMB

PAPER 1

ABT 4405: BIOINFORMATICS

END OF SEMESTER EXAMINATION

SERIES: APRIL 2016

TIME:2 HOURS

DATE: May 2016

Instructions to Candidates

You should have the following for this examination
-*Answer Booklet, examination pass and student ID*
This paper consists of **FIVE** questions. Attemptquestion ONE (Compulsory) and any other TWO questions.
Do not write on the question paper.

Question ONE

a. Define the following terms.

i.	Identifier	1 mark
ii.	Open reading frame	1 mark
iii.	Bootstrapping	1 mark

b. Give the meaning of the following terms;

i.	SMART	1 mark
ii.	COG	1 mark
iii.	BLAST	1 mark

iv. SAM 1 mark

c. You are given the following pair of sequences:

1. CCJKCEKDVSTT
2. CCJ I CEKAVSFY

3. CCMMSDRSSTVC
4. CCM E SCASSTVC

i. Calculate similarity scores between the pair of sequences using an identity matrix (assume mismatch=0, and match=1). 4 marks

d. Outline the merits of maximum likelihood (ML) method in phylogenetic analysis. 4 marks

e. Describe the three primary nucleotide sequence databases. 6 marks

f. Describe the FASTA file format. 5 marks

g. Distinguish between unigene and locus link. 4 marks

h. Explain the progressive alignment principle in ClustalW. 4 marks

Question TWO

Describe the following protein databases.

i. KEGG 8 marks

ii. FSSP 6 marks

iii. CATH 6 marks

Question THREE

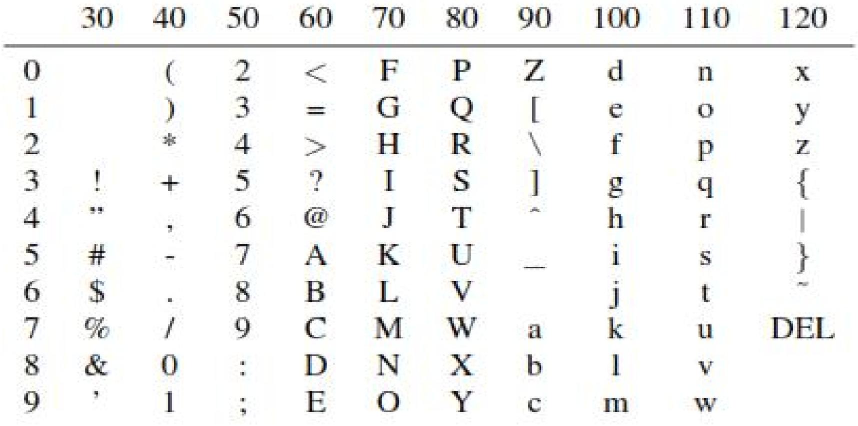
a. Describe three bioiformatic tools you would use to translate nucleic acid sequences. 6 marks

b. Explain the Dotplot method of sequence comparison. 8 marks

c. Distinguish between MEGABLAST and BLASTn bioinformatics tools. 6 marks

Question FOUR

a. 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 c

i. Calculate the corresponding Phred quality score based on the fastq-sanger format. 3 marks

ii. Calculate the corresponding Phred quality score based on Solexa/Illumina read format. 3 marks

b. Describe the use of the following bioinformatic tools. 8 marks

i. ESTcan

ii. Reverse

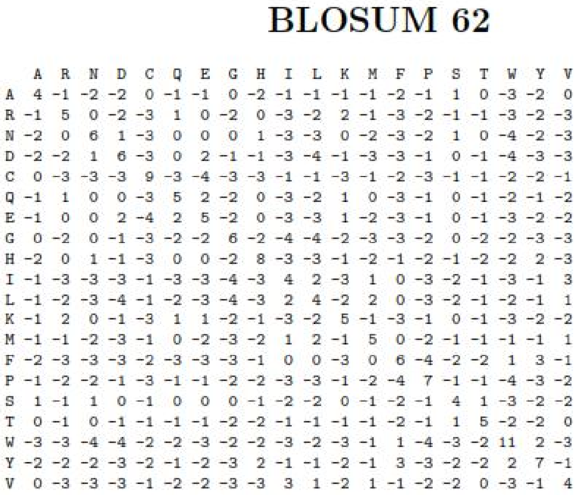
iii. GENSCAN

iv. GeneWise

c. Explain the use of tBLASTn program 6 marks

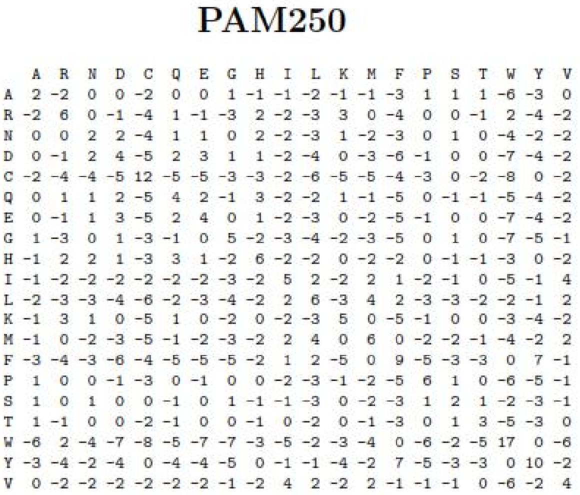
Question FIVE

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



BLOSUM 62

	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



PAM250

	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	-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	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4

Given the pairs of sequences below;

1. AWSADVKY
2. DWSAEVKK

3. KKAWVDVA
4. KYWSAVYA

a. Calculate similarity scores between these sequences using:

   i. PAM250 matrix 4 marks

   ii. BLOSUM 62 matrix 4 marks

   iii. Identity matrix (assume match = 2, and mismatch = 1) 4 marks

b. Describe any FOUR bioinformatic tools you would use to perform multiple sequence alignment.
8 marks