



**RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES, MIHINTALE**

**B.Sc. (Special) Degree in Applied Biology
Fourth Year – Semester II Examination – April/May 2016**

MIB 4204 – BIOINFORMATICS

Time: Two (02) hours

Answer all (04) questions.

1. I. Circle "T" if you think the statement is always true; circle "F" if you think the statement can sometimes or always be false. **(14 marks)**
 - (a) Two sequences are called homologous if they are significantly similar (T / F)
 - (b) Use of two different substitution matrices to perform the same BLAST search may result in different scores and e values, but the alignments will be identical (T / F)
 - (c) A region of the genome that is conserved between only rats and mice is less likely to be functionally important than a region that is conserved among humans, mice, rats, chicken, and puffer fish (T / F)
 - (d) BLOSUM80 matrix will always yield more significant matches than BLOSUM50 matrix (T / F)
 - (e) Non-protein-coding regions of the genome do not contain functionally important elements (T / F)
 - (f) Increasing the BLAST seed (word) size will allow BLAST to return more alignments (T / F)
 - (g) The NCBI protein and nucleotide databases contain about the same number of sequences because there is a one-to-one correspondence between genes and protein products (T / F)

II. Consider two proteins with weak and remote similarity. Why would you expect a better alignment using the protein sequences rather than the DNA sequences?

(10 marks)

III.

	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

(The gap opening costs -8 and gap extension costs -1)

Deduce the total scores of following alignments using the above substitution matrix. (09 marks)

(a) AACDQRST

A-CD-RST

(b) AACDQRST

A-CD-SST

(c) AACDQRST

A-DDQ-YT

IV. (a) Describe what pairwise sequence alignment is. (05 marks)

(b) Compare and contrast between the local and global alignments. (10 marks)

(c) In a homology search why is it necessary to filter the low complexity regions?
(02 marks)

2. (a) Describe the type of characters that can be used in molecular phylogenetics. **(10 marks)**
- (b) Explain the importance of DNA sequence alignment in molecular phylogenetics. **(10 marks)**
- (c) Human Immunodeficiency Virus (HIV) originated from Simian Immunodeficiency Virus (SIV) that infects apes and monkeys in Africa. There are two hypotheses proposed to explain its spread to humans from monkeys and apes. One suggests that humans were infected through the hunting of SIV infected monkeys for food. The alternative hypothesis is that HIV spread to humans through oral polio vaccine produced in SIV infected chimpanzee kidney tissue in a lab in Kisangani, Democratic Republic of Congo (DRC). You have been provided with SIV infected blood samples from monkeys and apes in Africa and blood samples from chimpanzees from Kisangani, DRC. Explain how you plan to test these hypotheses using your knowledge in molecular phylogenetics. **(30 marks)**
3. (a) Many scientific journals require the deposition of the data in to one of the three most prominent online databases GenBank, DRYAD and TreeBASE. Evaluate the methods of submission of your sequence data with respect to these databases. **(20 marks)**
- (b) Write short notes on the following.
- I. BLAST **(10 marks)**
 - II. Orthologous and Paralogous genes **(10 marks)**
 - III. PAM and BLOSSUM matrices **(10 marks)**
4. (a) Assume you obtained an E value of 2 in a databank search using a DNA sequence as your query. What is your opinion of the homology of the sequences? Justify your answer. **(05 marks)**
- (b) Zika virus disease (Zika), is a disease caused by the Zika virus, which is spread to people primarily through the bite of an infected *Aedes* species mosquito. Although this disease doesn't pose very specific symptoms, recently a patient from Maharashtra, India was admitted to the hospital due to fever, rash, joint pain, conjunctivitis and muscle pain which are common symptoms of Zika. Many scientists claim that this is a possible case of Zika. Using your background of molecular biotechnology and bioinformatics, formulate a proposal to examine this claim. **(45 marks)**