



CENTRAL UNIVERSITY OF SOUTH BIHAR

Department of Bioinformatics

End-Term Open Book Examination

Session: 2020-2022

Semester: 2nd (April-July 2021)

Programme: **M.Sc. Bioinformatics**

Date: - - - - -

Course Code: MSBIS2001C04

Course Title: Algorithms in Bioinformatics

Duration: **2 ½ hours**

Course Credits: **4**

Maximum Marks: **50**

Name: Programme:..... Semester:
.....

Course Title: Course Code:

Total No. of pages used: Date: Signature:

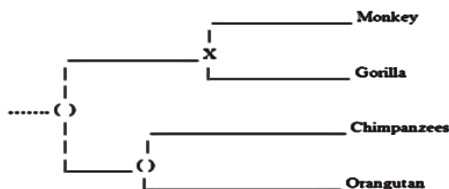
Instructions:

1. Preferably write your answers on A4 size plain paper (non-ruled) sheets.
2. Write your required details on the first page in the same order as specified below:
3. After completing the examination, write page number on the top right corner of each page in the format: 1/x, 2/x, , x/x where 'x' is the total number of pages used. If you have used total 6 pages then your page numbers will be 1/6, 2/6, 6/6.
4. The students have to write the answers on both side of the sheet (A4 size plain paper non ruled sheet).
5. The questions asked here are basically designed to assess the interpretation and application of knowledge, comprehension skills, and critical thinking skills rather than only recalling capacity.
6. Total twelve short answer questions of **five points each** are given covering the entire course content.
7. Answer **any ten** questions in total in maximum **two and half hours**.
8. The maximum limit to answer a question is 200 -300 words.
9. At the start of the examination all the questions will be released through e-mail and/or WhatsApp.
10. The total time limit to attempt the question paper is **two and half hours**. Along with the two and half hours, extra 30 minutes will be given for IT related activities such as downloading questions, scanning of answer sheets and uploading/emailing them.
11. After completing the examination within the stipulated time (two and half hours, scan your answer sheets or click pictures and submit it electronically in **one single file** (preferably PDF) to the course instructor through e-mail (ajaysingh@cusb.ac.in) strictly within stipulated time limit for submission (Three hours). Before submitting, rename your file and keep your name and enrolment number as file name.
12. **Please note** that do not use these extra 30 minutes for writing answers. Rather, finish writing as soon as possible within two and half hours and immediately submit your answers in the prescribed way given below. **Due to any reason, if a student is unable to submit the answer sheet file within the time limit, the university will not consider this examination and conduct another examination in the conventional mode whenever the conditions return to normal and circumstances permit or the university deems suitable. No other option or reason shall be entertained.**

13. In case you feel difficulty in submitting the answer sheet file through e-mail, then you are required to submit it to the concerned course instructor through WhatsApp **within the stipulated time only** and email it later on (within 48 hours) along with the screen shots of WhatsApp submission.

.....

Q1. In below diagram, where () indicates speciation and X indicates gene duplication. Looking the diagram, how you will explain homologous, orthologous and paralogous?



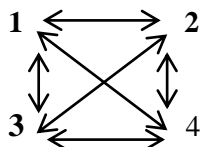
Q2. How FASTA and BLAST algorithm differs in their algorithms for sequence searches in Bioinformatics? What is the role of selection of substitution matrices in BLAST algorithm?

Q3. Calculate and construct the Global sequence alignment matrix for below given sequence with following parameter

AGTTC **Match = +1, Mismatch = -1 and Gap Score = -2**
 AGTC

Q4. When any Bioinformatician will prefer iterative or progressive method for sequence alignment approach for Multiple Sequence Alignment? Justify with an example.

Q5. For given weight matrix solve the travel salesman problem as shortest route?



0	4	10	4
4	0	5	6
4	5	0	3
6	3	3	0

Q6. Calculate the probability of Sequence = CAGGTCAGT in below protein sequence with PSSM Algorithm?

GAGGTAAC
TCCGTAAGT
CAGGTTGGA
ACAGTCAGT
TAGGTCATT
TAGGTAAGT

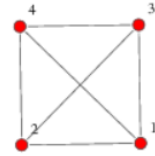
Q7. Which information you can retrieve with whole genome alignment? How Next generation sequencing is preferable to Sanger method and Di-deoxy method of sequencing? Justify.

Q8. Gene finding for a newly sequenced genome of an organism can be achieved by aligning RNA sequencing reads (from the same organism) to newly sequenced genome. Does the Bioinformatical method of gene

prediction still hold its relevance in current era of, cost effective and easily available, next generation sequencing technologies? Justify your answer.

Q9. Describe the characteristic of promoter sequences? Give four tools to predict promoter region of any given sequence?

Q10. Discuss adjacency matrix. Write down the adjacency matrix of the opposite graph.



Q11. Cluster these data points – {2, 3, 4, 10 11, 12, 20, 25, 30} using k-means clustering.

Q12. Attempt any two of the following.

- (a) Write down the kernel function that can separate these two-dimensional data points - (0,3), (1, 2), (2,1) and (3, 0). What is kernel trick in SVM?
- (b) Write down the master equation of a neural network along with the illustration describing Weights, bias and activation function.
- (c) Explain differences in graph centrality measures- betweenness, eigenvector, degree & closeness centrality.
