(University of Choice)

MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY (MMUST)

MAIN CAMPUS

UNIVERSITY EXAMINATIONS 2021/2022 ACADEMIC YEAR

FOURTH YEAR SECOND SEMESTER EXAMINATIONS

FOR THE DEGREE OF BACHELOR OF MEDICAL LABORATORY SCIENCES

MAIN EXAMINATION

COURSE CODE: BML 421

COURSE TITLE: BIO-INFORMATICS AND BIO-COMPUTING

DATE: 20/04/2022 **TIME:** 8.00 -10.00 AM

INSTRUCTIONS TO CANDIDATES

This paper is divided into three sections, **A B** and **C,** carrying respectively: Multiple Choice questions (**MCQs**), short answer questions (**SAQs**) and long answer questions (**LAQs**).

TIME: 2 Hours

MMUST observes ZERO tolerance to examination cheating

SECTION A: MULTIPLE CHOICE QUESTIONS (MCQs)

1. You are asked to go to the NCBI Human Genome page. What is the most probable task you will be asked to do?
   A. Determine what genes are around 'your' protein's gene on its chromosome.
   B. Identify a DNA sequence and see if it came from a human.
   C. Look up for papers about diseases caused by abnormalities in a certain protein.
   D. Look at colorful, rotating, 3-D pictures of the tertiary structure of a protein.

2. Which of the following cannot be related to multiple sequence alignment?
   A. Many conserved and functionally critical amino acid residues can be identified in a protein multiple alignment
   B. Multiple sequence alignment is also an essential prerequisite to carrying out phylogenetic analysis of sequence families and prediction of protein secondary and tertiary structures
   C. Multiple sequence alignment also has applications in designing degenerate polymerase chain reaction (PCR) primers based on multiple related sequences
   D. This method does not contribute much to degenerate polymerase chain reaction (PCR) primers creation.

3. Which of the following is untrue about DCA?
   A. It stands for Divide-and-Conquer Alignment
   B. It works by breaking each of the sequences into two smaller sections
   C. The breaking points during the process are determined based on regional similarity of the sequences.
   D. If the sections are not short enough, further divisions are restricted as well.

4. Pfam is available at four locations around the world. Which of the following is not one of them?
   A. UK
   B. Sweden
   C. US
   D. Japan

5. Which of the following is incorrect regarding pair wise sequence alignment?
   A. The most fundamental process in this type of comparison is sequence alignment
   B. It is an important first step toward structural and functional analysis of newly determined sequences
   C. This is the process by which sequences are compared by searching for common character patterns and establishing residue–residue correspondence among related sequences
   D. It is the process of aligning multiple sequences

6. If the two sequences share significant similarity, it is extremely _____ that the extensive similarity between the two sequences has been acquired randomly, meaning that the two sequences must have derived from a common evolutionary origin.
   A. Unlikely
   B. Possible
   C. Likely
   D. Relevant

7. Phylogenetic relationship can be shown by
   A. Dendrogram
   B. Gene Bank
   C. Data retrieving tool
   D. Data search tool

8. Which of the following is not a benefit or factual of FASTA over BLAST?
   A. FASTA scans smaller window sizes
   B. It gives more sensitive results
   C. It gives less sensitive results
   D. It gives results with a better coverage rate for homologs

9. The BLAST program was developed in _____
   A. 1992
   B. 1995
   C. 1990
   D. 1991

10. Which of the following is not a variant of BLAST?
   A. BLASTN
   B. BLASTP
   C. BLASTX
   D. TBLASTNX

11. Which of the following is not a correct about FASTA?
   A. Its stands for FAST ALL
   B. It was in fact the first database similarity search tool developed, preceding the development of BLAST
   C. FASTA uses a 'hashing' strategy to find matches for a short stretch of identical residues with a length of k
   D. The string of residues is known as blocks

12. Operating system is
   A. A collection of hardware components
   B. A collection of input-output devices
   C. A collection of software routines
   D. All of the above

13. A single piece of information in a database is called
   A. File
   B. Field
   C. Record
   D. Data set

14. Many scientists are interested in studying mitochondrial DNA because it.
   A. Is only present in vertebrates closely related to humans.
   B. Replicates by synthesizing an mRNA that then acts as a DNA polymerase.
   C. Contains over 50% of the gens in the human genome.
   D. Mutates rapidly and allows us to study evolution over short time scales.

15. The term bioinformatics was coined by
   A. JD Watson
   B. Margaret Dayhoff
   C. Pauline Hogeweg
   D. Frederic Sanger

16. Step wise method for solving problems in computer science is called
   A. Flowchart

B. Sequential design
C. Procedure
D. Algorithm

17. The term used to refer to something 'performed on computer or computer simulation'
A. Dry lab
B. Wet lab
C. *In vitro*
D. *In silico*

18. Application of bioinformatics include
A. Data storage and management
B. Drug designing
C. Understand relationships between organisms
D. All of the above

19. The computational methodology that tries to find the best matching between two molecules, a receptor and ligand is called
A. Molecular matching
B. Molecular docking
C. Molecular fitting
D. Molecular affinity checking

20. The process of finding relative location of genes on a chromosome is called
A. Gene tracing
B. Genome mapping
C. Genome walking
D. Chromosome walking

SECTION B: SHORT ANSWER QUESTIONS

1. Briefly describe physical and genetic maps and how they are generated in genome databases [8 Marks].
2. List types of bioinformatics databases giving examples of the databases in each case [8 Marks].
3. In the EMBL database format, what do the following lines represent; ID, AC, NI, DE, KW, DT, OS, OC, RN, FT, SQ, CC, RP, RX, RA, RL. [8 Marks].
4. Giving relevant examples, describe the use of databases in bioinformatics [8 Marks].
5. Name eight common web-based tools and their uses [8 Marks].

SECTION C: LONG ANSWER QUESTIONS (40 Marks)

1. Discuss the applications of bioinformatics in disease diagnosis [20 Marks].
2. Discuss the some of the common algorithms used in sequence analysis [20 Marks].
3. Compare and contrast BLAST and FastA platforms during sequence homology searches [20 Marks].