(*University of Choice*)

MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY (MMUST)

MAIN CAMPUS

UNIVERSITY EXAMINATIONS
2019/2020 ACADEMIC YEAR

THIRD YEAR SECOND SEMESTER MAIN EXAMINATIONS

FOR THE DIPLOMA
OF
MEDICAL BIOTECHNOLOGY & LABORATORY SCIENCES (MAIN)

COURSE CODE: BBD 323

COURSE TITLE: FUNDAMENTALS OF BIOINFORMATICS & BIOCOMPUTING

DATE: 7th December 2020 **TIME: 8.00 – 10.00 AM**

INSTRUCTIONS TO CANDIDATES

TIME: 2 Hours

MMUST observes ZERO tolerance to examination cheating

This Paper Consists of 4 Printed Pages. Please Turn Over.

MULTIPLE QUISTIONS 60 (MARKS)

1).A single piece of information on a data base is called
A).File
B).Field
C).Record
D).Dataset

2).Which of the following is a nucleotide sequence data base?
A).EMBL
B).SWISPROT
C).PROSITE
D).PROEMBL

3).Operating systems are?
A).A collection of hardware components
B).A collection of input and output devices
C).A collection of software routines
D).All of the above

4).A data base of current sequence map of the human genome is called?
A).OMMIM
B).HGMD
C).GOLD PATH
D).GENE CARDS

5).BLAST programme is used in?
A).DNA sequencing
B).Amino acid sequence
C).DNA bar coding
D).Bioinformatics

6).SWISPROT is related to?
A).Portable data bank
B).Swiss bank data
C).Sequence data bank
D).Sequence /sequence data

7) .BLOSUM matrices are used for?
A) .Multiple sequence alignment
B).Pairwise sequence alignment
C).Phylogenetic analysis
D).A of the above

8). Phylogenetic relationships can be shown by?
A). Dendrogram
B) Gene bank
C).Data retrieval tool
D).Data search tool

9). Margret Dayhoff developed the first protein database called?
A).SWISPROT
B).PDB
C).Atlas of protein sequence and structure
D).Protein sequence data bank

10).Each record in databases Is called?
A).An entry
B). A file
C). Record
D). Ticket

11) Literature data base include?
A).Medline and PubMed
B).MEDLINE and PDP
C).PubMed and PDP
D).MEDLINE and PDS

12). Fly base is?
A).Biodiversity of database
B). Literature of database
C).Mode organism database
D).Biomolecular database

13).Which of the following is an E coli model organism database
A).Eco gene
B).Eco base
C).Eco sequence
D). Colgene

14). which of the following is a protein sequence database?
A).DDBJ
B).EMB L
C).Gene bank
D). PIR

15).Gene ban the nucleic acid sequence is maintained by?
A).Brook have laboratory
B).DNA database of japan- DDBJ
C).European molecular laboratory - EMBL
D).National Center for Biotechnology Informatics-NCBI

16) Submission to gene ban are made using?
A).Ban kit and sequin
B). Ban kit and Ban kin).
C).Sequin and ban kin
D).Entrez

17).A comprehensive data base for the study of human genetics and molecular biology is?
A).PDB
B).STAG
C).OMIM
D) .PSD

18).The information retrieval tool of NCBI Gene bank is?
A).Enraz
B). STAG
C).Sequin
D). Text search

19).The first secondary database developed was?
A). BlAST
B).PRINT
C).PROSITE
D).PIR

20).Which of the following is a sequence alignment tool
A).BAST
B), PRINT
C).PROSITE
D).(a) 1988

.

21).Which of the following scientists created the first Bioinformatics database?
(a) Dayhoff
(b) Pearson
(c) Richard Durbin
(d) Michael.J.Dunn

22. The human genome contains approximately_____.

(a) 6 billion base pairs

(b) 5 billion base pairs

(c) 3 billion base pairs

(d) 4 billion base pairs

.

23). Which of the following tools is used for the identification of motifs?

(a) BLAST

(b) COPIA

(c) PROSPECT

(d) Pattern hu

24). The first molecular biology server expasy was in the year _____.

(a) 1992

(b) 1993

(c) 1994
(d) 1995

25).The first biological data base was created by?
A).Richard Durbin
BDayhoff
C). Michael Dumn
D).Pearson
26). WhatC).Michaaael Dumn is the deposition of cDNA into the inert structure called?

(a) DNA probes

(b) DNA polymerase

(c) DNA microarrays

(d) DNA finger

27 The identification of drugs through the genomic study is called_____.

(a) Genomics

(b) Pharmacogenomics

(c) Pharmacogenetics

(d) Cheminformatics
28). Which of the following compounds has desirable properties to become a drug?

(a) Fit drug

(b) Lead

(c) Fit compound

(d) All of the above

29. Proteomics refers to the study of _____.

(a) Set of proteins in a specific region of the cell

(b) Biomolecules

(c) Set of proteins

(d) The entire set of expressed proteins in the cell

30. The process of finding the relative location of genes on a chromosome is called _____.

(a) Gene tracking

(b) Genome walking

(c) Genome mapping

(d) Chromosome walkin

31. The computational methodology that tries to find the best matching between two molecules, a receptor and ligand are called _____.

(a) Molecular fitting

(b) Molecular matching

(c) Molecular docking

(d) Molecule affinity

32). Which of the following are not the application of bioinformatics?

(a) Drug designing

(b) Data storage and management

(c) Understand the relationships between organisms

(d) None of the above

33). The term “invitro” is the Latin word which refers to_____.

(a) Within the lab

(b) Within the glass

(c) Outside the lab

(d) Outside the glass

34. The stepwise method for solving problems in computer science is called_____.

(a) Flowchart

(b) Algorithm

(c) Procedure

(d) Sequential design

35). The term Bioinformatics was coined by _____.
(a) J.D Watson
(b) Pauline Hogeweg
(c) Margaret Dayhoff
(d) Frederic Sang
36).The laboratory work using computers and associated with web-based analysis generally online is referred to as _____.
(a) In silico
(b) Dry lab
(c) Wet lab
(d) All of the
37).. Which of the following is the first completed and published gene sequence?
(a) ΦX174
(b) T4 phage
(c) M13 phage
(d) Lambda phage
Sol: (a) ΦX174
38 The laboratory work using computers and computer-generated models generally offline is referred to as _____.
(a) Insilico
(b) Wet lab

(c) In silico
(c) Dry lab
(d) All of the above.
39 The computer simulation refers to _____.
(a) Dry lab
(b) Invitro
(d) Wet lab
40). Which of the following sets contains all aromatic residues?
A. G, D, N, E
B. I, V, L, M
C. R, K, H
D. F, Y, W
41). What is the Twilight Zone?.
A. Where alignments appear plausible and are statistically significant.
B. Where alignments may appear plausible to the eye, but are no longer statistically significant.
C. Where alignments neither appear plausible nor statistically significan
D. Where alignments share 30% identity.
42 Two sequences are said to be homologous if:
A. they have diverged from a common ancestor.
B. their alignments share 30% identity or more.
C. they belong to the same fold family.
D. they have converged to share similar functional properties.
43 When performing a database search, what is the definition of an E-value?
A. The chance that a random sequence could achieve a better score than the query.
B. The chance that a homologous sequence could achieve a similar score to the query.
C. The chance that a random sequence could achieve a worse score than the query.
D. The chance that a homologous sequence could achieve a better score than the query.

Protein Family Analysis

44 Which of the following regular expressions would be matched by sequence DWILKDG?
A. D-M-x-[ILV]-x{2}-G
B. [DN]-W-x-[ILV]-[RKH]-x-G
C. [DN]-W-x{2}-[ILV]-G
D. D-W-I-[ILMV]-x-K-[GA]

45 Well-conserved regions in multiple sequence alignments:
A. reflect areas of structural importance.
B. reflect areas of functional importance.
C. reflect areas of both functional and structural importance.

D. reflect areas likely to be of functional and/or structural importance.

46. Why are colour schemes important in creating and analyzing sequence alignments?
A. They look pretty
B. To make clearer printouts and presentations
C. To allow you to distinguish conserved residues and residue groups more easily
D. To allow you to detect active sites of proteinsSequence property profiles

47Hydropathy plots are usually used to predict:
A. beta secondary structures
B. transmembrane domains
C. alpha secondary structures

48 Databases such as CATH and SCOP are used to identify:
A. the structural family to which a protein belongs
B. the genic family to which a protein belongs
C. homologous proteins
D. analogous proteins

49). Coordinates for known protein structures are housed in?
A. CATH
B. SCOP
C. PDBsum
D. PDB

50). Homology modelling is a procedure whereby:
A. due to low sequence similarity between proteins of unknown and known structure, the structure is predicted from first principles (i.e., ab initio)
B. due to high sequence similarity between proteins of unknown and known structure, the same function is assumed for both
C. due to high sequence similarity between proteins of unknown and known structure, the structure of the latter is used as a template to model the former
D. a protein of unknown structure is compared against a library of fold templates to find the best match

51 Threading is a procedure whereby:
A. due to low sequence similarity between proteins of unknown and known structure, the structure is predicted from first principles (i.e., ab initio)
B. due to high sequence similarity between proteins of unknown and known structure, the same function is assumed for both
C. due to high sequence similarity between proteins of unknown and known structure, the structure of the latter is used as a template to model the former
D. a protein of unknown structure is compared against a library of fold templates to find the best match

52). With homology modelling, if there are minor errors in the template, the model will:
A. be very good
B. be just as good as the template
C. be unable to be built using current modelling programs
D. be completely wrong

53 With homology modelling, if there are major errors in the template, the model will:
A. be very good
B. be just as good as the template
C. be unable to be built using current modelling programs
D. be completely wrongInvestigating the human genome

54 Should one use several gene prediction packages, and why?
A. no, each is 100% accurate
B. yes, it's useful to see what is/isn't predicted by each package and, consensus is useful
C. yes as none of them work reliably
D. no, as the use of several packages complicates the analysis

55). Should one check gene predictions using sequence similarity searches, and why?
A. yes, because sequence similarity to a known gene implies the candidate open reading frame is likely to be part of a gene
B. no, because orphan genes make this approach unreliable
C. yes, because ab initio gene predictions are not reliable
D. no, because ab initio gene predictions are reliable

Investigating the inositol phosphatase family

56).Results from multiple motif database searches are biologically more meaningful than single motif searches because:
A. multiple motif databases store more motifs, and thus there is a greater chance of a hit

B. the alignments and the selection of motifs in single motif databases is performed using automatic rather than manual methods, adding more errors
C. multiple motif databases are annotated
D. a single motif may be present in several, otherwise unrelated, familes and it is only when put in context with other motifs that families of related proteins are apparent

57). Are profiles and hidden Markov models more potent discriminators than regular expressions?
A. no, because they are used for domain databases and, having a different area of application, they can't be compared
B. yes, as they contain probability data for each position in the motif, which can be used to calculate match statistics such as E-values
C. no, because even though they are more sensitive, they are less selective
D. yes, because they are fine tuned to individual protein domains

58 An example of homology similarity tool?

A.Prospect
B.EMROSS
C.RASMOL
D.BLAST
59). What is a fingerprint?
A. A protein family discriminator built from a set of regular expressions.

A).1985
B).1986
C).1987
D)1988

60).Swissprot protein sequence data base began in?
A).1085
B).1986
C).1987
D).1988