**Roll. No. ........................**

**Question Booklet Number**

**O.M.R. Serial No.** ☐☐☐☐☐☐☐☐

# B.Sc. (PART-II) EXAMINATION, 2021
# BIOTECHNOLOGY (NEW COURSE)
## [ PAPER : Fifth (BBT-205) ]
## ( Computational Biology and Bioinformatics )

| Paper ID | | |
|---|---|---|
| 5 | 0 | 7 |

**Question Booklet Series**
**A**

**Time : 1 : 30 Hours** **Max. Marks : 150**

**11**

**Instructions to the Examinee :**

1. Do not open this Booklet untill you are told to do so.
2. Candidates should fill their roll number, subject and series of question booklet details correctly, otherwise, in case of any discrepancy in the evaluation, it will be the responsibility of the examinee himself.
3. There are 100 questions in the booklet. Examinee is required to answer only 75 questions in the OMR Answer Sheet provided. Four alternative answer to each question are given below the question, out of these four only one answer is correct. The answer which you think is correct or most appropriate, completely fill in the circle containing its letter in your answer sheet (O.M.R. Answer Sheet) with black or blue ball point pen.

***(Remaining instructions on last page)***

**परीक्षार्थियों के लिए निर्देश :**

1. जब तक कहा न जाये, इस प्रश्नपुस्तिका को न खोलें।
2. परीक्षार्थी अपने अनुक्रमांक, विषय एवं प्रश्नपुस्तिका की सिरीज का विवरण यथास्थान सही-सही भरें, अन्यथा मूल्यांकन में किसी भी प्रकार की विसंगति की दशा में उसकी जिम्मेदारी स्वयं परीक्षार्थी की होगी।
3. प्रश्न-पुस्तिका में 100 प्रश्न हैं। परीक्षार्थी को केवल 75 प्रश्नों का उत्तर दी गई OMR उत्तर-पत्रक में देना है। प्रत्येक प्रश्न के चार वैकल्पिक उत्तर प्रश्न के नीचे दिये गये हैं। इन चारों में से केवल एक ही उत्तर सही है। जिस उत्तर को आप सही या सबसे उचित समझते हैं, अपने उत्तर-पत्रक (O.M.R. Answer Sheet) में उसके अक्षर वाले वृत्त को काले या नीले बॉल प्वाइंट पेन से पूरा भर दें।

*(शेष निर्देश अन्तिम पृष्ठ पर)*

# ROUGH WORK

1. Computational biology involves development and application of ____ to study biological systems.

(A) Graphs

(B) Equation

(C) Text

(D) Theoretical Methods

2. GWAS stands for :

(A) Global Wise Alignment Sequence

(B) Genome Wide Association Studies

(C) Global Wide Association Sequence

(D) Gene Wide Association Studies

3. How many genes does the Human genome contain ?

(A) 12000

(B) 9000

(C) 30000

(D) 60000

4. The part of genome that is common between various species following comparative genomics analysis is called :

(A) Core genome

(B) Cot plot

(C) Pan genome

(D) Whole genome

5. Which of the following are not macromolecule visualization platforms :

(A) Cn3D

(B) PyMol

(C) GWAS

(D) Jmol

6. ENCODE stands for :

(A) Encyclopedia of DNA Elements

(B) Encyclopedia of Complete DNA Expression

(C) Encyclopedia of Omics DNA Expression

(D) None of the above

7. Which of the following VISTA servers can be used to align nucleotide sequence to whole genome assemblies ?

(A) VVg VISTA

(B) gVISTA

(C) mVISTA

(D) phyloVISTA

8. Which of the following is true regarding ARTEMIS software ?

(A) Developed by Sanger Institute

(B) Genome browser and annotation tool

(C) None of the above

(D) All of the above

9. The size of draft human genome sequence was :

(A) $3x10^9$ bp

(B) $3x10^9$ kbp

(C) $3x10^9$ bases

(D) $3x10^9$ Mbp

10. A parameter that describes number of hits expected by chance when searching a database is :

(A) Gap value (G)

(B) Expect value (E)

(C) Score value (S)

(D) Probability value (P)

11. BLAST program was developed by :

(A) Pearson and Lipman

(B) Stephen Altschul

(C) Frederick Sanger

(D) Margaret Dayhoff

12. BLAST and FASTA are similarity searching programs based on :

(A) Heuristic algorithms

(B) Word algorithm

(C) Exact algorithm

(D) All of the above

13. For perfectly aligned protein sequences, dot plot will show :

(A) Diagonal formation

(B) No diagonals

(C) Two intersecting diagonals

(D) Parallel diagonals

14. In dot plots, insertion and deletion of nucleotides will be represented respectively by :

(A) Horizontal, vertical frameshift

(B) Vertical, horizontal frameshift

(C) Small crosses

(D) None of the above

15. Which of the following is true for PAM matrices ?

(A) Point Altered Mutation

(B) Percent Altered Mutations

(C) Point Advanced Matrix

(D) Point Accepted Mutation

16. Global alignment is useful for :

(A) Aligning closely related sequences

(B) Based on dynamic programming

(C) Alignment for every residue in every sequence

(D) All of the above

17. PAM Matrix is used for :

(A) Multiple sequence alignment

(B) Local alignment

(C) Global alignment

(D) All of the above

18. In FASTA program, scoring matrix is formatted to include :

(A) gap addition

(B) no gap penalties

(C) gap penalties

(D) units

19. In affine gap penalty :

(A) gap opening and gap extension receives same score

(B) gap opening has higher score than gap extension

(C) Gap opening has lower score than gap extension

(D) None of the above

20. Which of the following is incorrect about BLOSUM ?

(A) Stands for BLOCKS Substitution Matrix

(B) Based on global alignment

(C) Used for sequence alignment of proteins

(D) Useful for evolutionary divergent protein sequences

21. BLOSUM 80 is useful for comparing :

(A) Distant sequences

(B) Unrelated sequences

(C) Closely related sequence

(D) Divergent sequences

22. What is the color key for alignment scores > = 200 in NCBI blast search ?

(A) Green

(B) Red

(C) Pink

(D) Blue

23. In which year was NCBI BLAST program developed ?

(A) 1990

(B) 1989

(C) 1992

(D) 1991

24. In BLAST, word from query sequence contains ______ residues for proteins and _____ residues for DNA respectively.

(A) three, four

(B) one, three

(C) three, three

(D) three, eleven

25. Which format is used for storing assembled genome sequence data ?

(A) FASTA

(B) FAST Q

(C) ILLUMINA

(D) BLAST

26. The method of constructing genomes from large sequence reads without a reference genome sequence is known as :

(A) comparative genomics

(B) de novo sequence assembly

(C) none of the above

(D) all of the above

27. Which of the following queries protein sequence to nucleotide sequence database translated in all six reading frames ?

(A) tblast x

(B) tblast n

(C) pblast n

(D) blast x

28. Dot plots to visualize similarity between 2 protein or nucleotide sequence was introduced by :

(A) Smith Waterman

(B) Needleman Wunsch

(C) Gibbs and McIntyre

(D) Lipman

29. Which of the following is not variant of NCBI BLAST ?

(A) tblast x

(B) pblast x

(C) blast x

(D) blast n

30. Fluorescent labelled nucleotides are used for identifying sequence in which of the following techniques ?

(A) Illumina

(B) PacBio

(C) Oxford Nanopore

(D) All of the above

31. Which of the following is not a gene prediction tool ?

(A) GLIMMER

(B) GeneMark

(C) FGENESH

(D) GOAT

32. Two proteins sharing high level of sequence identity are likely to :

(A) differ in function

(B) be closely linked on chromosome

(C) share same function

(D) be in non-coding region of genome

33. Functional genomics studies utilize which of the following technologies ?

(A) Transposon

(B) CRISPR Cas

(C) RNAi

(D) All of the above

34. Gene ontology permits description of genes and its products on the basis of :

(A) Genome, organism, species

(B) Molecular function, biological process in which it participates, cellular location

(C) Molecular structure, cellular location and transcript

(D) Molecular function, gene loci, map

35. Probablistic functional gene networks are used for :

(A) RNA quantification

(B) Creating phylogenetic analysis

(C) Constructing interactions between genes of metabolic pathways

(D) Determining macromolecule structure

36. In genetic network, genes are represented by ______, and pairwise relationship between genes are represented by _____.

(A) nodes, edges

(B) edge, node

(C) node, internode

(D) node, interedge

37. ____ is the study of how genes and intergenic regions of genome affect biological processes.

(A) Structural genomics

(B) Functional genomics

(C) Biostatistics

(D) Biochemistry

38. Which of the following statement is correct ?

(A) Predicted set of proteins for a genome is proteome

(B) Predicated set of RNA for a genome is transcriptome

(C) Entire set of protein encoded by genome is proteome

(D) None of the above

39. During genome assembly, sequences created from alignment of highly similar single reads are known as :

(A) Contigs

(B) Pair end reads

(C) Concatemers

(D) Scaffolds

40. Whole Exome Sequencing refers to :

(A) Sequencing of all RNA genes

(B) Sequencing of all protein coding genes in a genome

(C) Sequencing of all genome

(D) Sequencing of heterochromatin portion of genome

41. Functional genomics provides indepth analysis of :

(A) Coding region of genome

(B) Non-coding region of genome

(C) Both (A) and (B)

(D) None of the above

42. Gene prediction tool GLIMMER stands for :

(A) Gene Locater and Interpolated Markov Modeler

(B) Genome Locater and Interpolated Markov Modeler

(C) Gene Location and Interpretation Machine Model

(D) None of the above

43. Study of effect of genome to find a link between genotype and phenotype is termed as :

(A) Structural genomics

(B) Biostatistics

(C) Functional genomics

(D) Gene engineering

44. Tilling arrays are used in genome annotation for :

(A) Gene sequencing

(B) Genome wide mapping of transcript positions

(C) Proteome identification

(D) Array designing

45. Application of informatics techniques to obtain, store, analyze and interpret large biological data is known as :

(A) Literature

(B) Statistics

(C) Bioinformatics

(D) Mathematics

46. Sequencing errors defined as 'probability of base being incorrect' is given by :

(A) Phred score

(B) PAM score

(C) Was score

(D) $N_{50}$

47. Method used for predicting binding conformation of ligand to appropriate target binding site is known as :

(A) Signal prediction

(B) Molecular systems

(C) Networking

(D) Molecular docking

48. What are the challenges associated with whole genome sequencing assembly ?

(A) Repeat sequences

(B) Polymorphism

(C) Sequencing errors

(D) All of the above

49. Next Generation Sequencing cannot be used for :

(A) Genome sequencing

(B) Detection of sequence variation

(C) Molecular docking

(D) SNP detection

50. The first NGS platform commercially available was :

(A) Roche/454 Genome sequencer

(B) Illumina/Solexa Genome sequencer

(C) ABI Solid Genome Sequencer

(D) Heliscope Genome Sequencer

51. Global sequence alignment uses which type of dynamic programming ?

(A) Smith-Waterman algorithm

(B) Needleman-Wunsch algorithm

(C) None of the above

(D) All of the above

52. VMD stands for :

(A) Visual Molecular Device

(B) Visual Molecular Database

(C) Video Machine Design

(D) Visual Molecular Dynamics

53. Gene prediction of introns and exons need not be performed for :

(A) Mouse

(B) Fungi

(C) Algae

(D) Bacteria

54. Which of the following statements is incorrect ?

(A) Phylogenetic profiles shows genes not commonly present in organisms

(B) Linked genes are present on same chromosome

(C) Set of genes whose expression pattern in similar are likely to have related function

(D) Microarray detect global transcript changes

55. In a typical GWAS study, common variants in individuals with or without a common disease are identified using :

(A) CRISPR editing

(B) Genome wide SNP arrays

(C) Restriction mapping

(D) Ultracentrifugation

56. Which of the following genomic library screening strategies does not depend on expression ?

(A) Screening using antibody

(B) Functional screening

(C) Using degenerate oligonucleotide primers

(D) Using EST

57. GWAS studies can help in crop improvement by :

(A) analyzing inheritance modes of traits

(B) studying association between SNP and desired phenotypic traits

(C) identifying quantitative trait loci for stress responses

(D) all of the above

58. Which of the following is false regarding HUGO ?

(A) Established as non-profit organization in 1988

(B) Goal to build genetic and physical map of human and chimpanzee genome

(C) Sequence human genome

(D) Map and identify all human genes present in nucleotide sequence

59. Ab initio based programs for gene prediction, algorithms rely on :

(A) Gene signals

(B) Gene content

(C) Both gene signal and gene content

(D) Type of organism

60. Genome wide association mapping measures :

(A) correlation between marker alleles and allele in a population

(B) measures recombination between marker and unknown gene

(C) both (A) and (B)

(D) none of the above

61. As E value of a BLAST search becomes smaller, the score tends to be :

(A) Smaller

(B) Larger

(C) Same

(D) Zero

62. In prokaryotic gene prediction, start codon used include :

(A) ATG

(B) ACG

(C) ATG, GTG, TTG

(D) UAG

63. Codon usage bias in functional genomics refers to :

(A) genome structure

(B) differences in frequency of occurrence of synonymous codons in coding DNA

(C) computational tool

(D) none of the above

64. Shotgun sequencing strategy for sequencing human genome was given by :

(A) Craig Venter at Celera Genomics

(B) J. Watson at Celera Genomics

(C) Lipman at NIH

(D) Crick at NIH

65. In polymorphisms in DNA sequences, VNTR stands for :

(A) Very Numerous Transposon Repeat

(B) Variable Number Tandem Repeat

(C) Variable Number Transposon Repeat

(D) Variable Nature Tandem Repeat

66. By default, PyMol will display molecule(s) in PDB file as :

(A) Wireframe

(B) Cylindrical helices

(C) Ribbons

(D) Smooth loops

67. The most common way to evaluate ligand protein alignment using docking is :

(A) RMSD value

(B) RSMD value

(C) Cluster value

(D) Van der walls force

68. Which of the following are parameters in ligand protein docking ?

(A) Calculating binding affinity

(B) Considering degress of freedom

(C) Ligand flexibility

(D) All of the above

69. Sequencing method where base sequences are called on the basis of luciferin mediated reaction is :

(A) Sanger's dideoxy method

(B) Pyrosequencing

(C) Maxam-Gilbert chemical reaction

(D) Microarray hybridization

70. BLAST tools can be used for the determination of which of the following options ?

(A) Identify coding regions of genome

(B) Identify intron-exon boundaries

(C) Establish phylogeny

(D) All of the above

71. What are the uses of Paired End Sequencing ?

(A) Detection of repetitive sequences

(B) Detection of genomic rearrangements

(C) Detection of gene fusions

(D) All of the above

72. Epigenome can be defined as :

(A) Set of chemical modifications to DNA and DNA associated proteins in a cell

(B) Mixture of genome

(C) Episomal portion of genome

(D) Mitochondrial genome

73. Which of the following is incorrect regarding VMD ?

(A) Several molecular rendering and coloring methods

(B) Does not include multiple sequence alignment plugin

(C) Built on Tcl/Tk and Python

(D) Read protein data bank files

74. The most commonly used large insert clone for genome sequencing is :

(A) Cosmid

(B) Fosmid

(C) Bacterial Artificial Chromosome

(D) Human Artificial Chromosome

75. Building sub-databases can be done using :

(A) Web data extraction system

(B) Using microarrays

(C) Next Generation Sequencing

(D) Using PyMol

76. What is a profile in bioinformatics ?

(A) It contains position-specific weights for amino acid substitution and penalties for gaps

(B) Address file

(C) Protein structure definition

(D) All of the above

77. DNA sequencing is part of :

(A) Structural genomics

(B) Functional genomics

(C) Genetics

(D) Forward genetics

78. EST stands for (in bioinformatics) :

(A) Exon Sequence Tags

(B) Expressed Sequence Tags

(C) Expression System Template

(D) Exon Sequence Template

79. The starting of FASTA format has _____ symbol.

(A) >

(B) <

(C) /

(D) *

80. Comparative genomics hybridization is a technique that can :

(A) detect genetic abnormalities in tumors

(B) detect chromosomal changes without need for cell culturing

(C) none of the above

(D) Both (A) and (B)

81. EcoTILLING, a procedure used for indentifying allelic variants for gene targeting in natural collections was first applied to the crop :

(A) Rice

(B) Pea

(C) Melon

(D) Wheat

82. HUGO International Mission Statement consists of :

(A) Promoting fundamental genomic research within nations and throughout the world

(B) Foster scientific exchange in genomics particularly in scientifically developing and emerging countries

(C) Support discourse in genetic and genomic ethics

(D) All of the above

83. Which of the following is incorrect regarding scaffolds in genome sequence assembly ?

(A) Consists of contigs separated by gaps

(B) Gap sequence between contig

(C) Contig sequence linked by pair end reads

(D) Contig sequence linked by mate pair libraries

84. Which of the following is NOT an probabilistic functional gene network ?

(A) NetGen

(B) BLAST

(C) BIomeNet

(D) YeastNet v.2

85. In molecule viewing platforms, ribbon diagrams use cylinders/spiral and broad flat arrows to represent ____ and ______ respectively.

(A) $\alpha$-helices, $\beta$-strands

(B) $\beta$-strands, $\alpha$-helices

(C) $\alpha$-helices, $\beta$-galactosidase

(D) Wires, backbone

86. RNA microarrays :

(A) monitor 1000s of gene simultaneously

(B) make use of ligand protein interaction

(C) perform DNA amplification

(D) visualize macromolecule structures

87. In Ball and Stick model, ball represents atoms. The size of the ball corresponds to :

(A) type of bond

(B) van der Waals radii of atom

(C) $\alpha$-helices

(D) all of the above

88. What is the difference between synonymous and non-synonymous single nucleotide polymorphism ?

(A) Synonymous SNP do not result in amino acid change in protein, but non-synonymous do

(B) Synonymous SNP change protein sequence while non-synonymous SNP do not

(C) Both cause no change

(D) None of the above

89. The process of determining the complete DNA sequence of an organism at a single time is known as :

(A) DNA profiling

(B) Whole Genome Sequencing (WGS)

(C) Whole Exome Sequencing (WES)

(D) Genome printing

90. Strategies used for sequencing of large genomes include :

(A) Whole-genome shotgun sequencing

(B) BAC-by-BAC sequencing

(C) Both (A) and (B)

(D) None of the above

91. Hypothesis free studies useful for identifying associations between genetic loci and traits are known as :

(A) Genome wide association studies

(B) Genetic with association sequences

(C) Gene wise association sequences

(D) Genome wide accelerated studies

92. Next Generation Ssequencing Technology refers to :

(A) Sequencing genome regions multiple times

(B) Massively parallel sequencing technology

(C) Creating cDNA library

(D) Creating amplicon library

93. The main features of Illumina NGS are :

(A) Tracking of labeled nucleotides as DNA chain is copied (SBS technology)

(B) Use of emulsion PCR

(C) None of the above

(D) Both (A) and (B)

94. Which of the following is used for protein structure validation ?

(A) Blast

(B) PROCHECK

(C) FASTA

(D) NGS

95. In ligand-protein interaction analysis, _____ amino acid is considered as most flexible and ____ the least flexible amino acid.

(A) glycine, proline

(B) proline, glycine

(C) atom, bond

(D) bond, atom

96. Which of the following is Genome Assembly Software ?

(A) PyMol

(B) SWISSMODEL

(C) PHYLIP

(D) CELERA

97. Which of the following are Genome Assembly Algorithms ?

(A) Overlap Layout Consensus (OLC)

(B) K-mer or de Bruijn Graph

(C) Both (A) and (B)

(D) None of the above

98. Depth coverage in genome sequencing can be calculated using formula $\mathrm{N}\times\frac{\mathrm{L}}{G}$ where N refers to :

(A) Number of read length

(B) Number of reads

(C) Average read length

(D) Genome size

99. Which of the following databases use profiles to classify proteins ?

(A) PROSITE

(B) SQL

(C) PYTHON

(D) PUBMED

100. Profiles are built by converting multiple sequence alignments into :

(A) Fingerprints

(B) Position-specific scoring Matrices

(C) Transcript profiles

(D) Biological databases

# ROUGH WORK

**Example :**

**Question :**

Q.1 Ⓐ ● Ⓒ Ⓓ
Q.2 Ⓐ Ⓑ ● Ⓓ
Q.3 Ⓐ ● Ⓒ Ⓓ

If more than 75 questions are attempted by candidate, then the first attempted 75 questions will be considered for evaluation.

4. Each question carries equal marks. Marks will be awarded according to the number of correct answers you have.
5. All answers are to be given on OMR Answer Sheet only. Answers given anywhere other than the place specified in the answer sheet will not be considered valid.
6. Before writing anything on the OMR Answer Sheet, all the instructions given in it should be read carefully.
7. After the completion of the examination, candidates should leave the examination hall only after providing their question booklet and OMR Answer Sheet separately to the invigilator.
8. There will be no negative marking.
9. Rough work, if any, should be done on the blank pages provided for the purpose in the booklet.
10. To bring and use of log-book, calculator, pager & cellular phone in examination hall is prohibited.
11. In case of any difference found in English and Hindi version of the question, the English version of the question will be held authentic.

**Impt. On opening the question booklet, first check that all the pages of the question booklet are printed properly. If there is any discrepancy in the question Booklet, then after showing it to the invigilator, get another question Booklet of the same series.**

उदाहरण :

प्रश्न :

प्रश्न 1 Ⓐ ● Ⓒ Ⓓ
प्रश्न 2 Ⓐ Ⓑ ● Ⓓ
प्रश्न 3 Ⓐ ● Ⓒ Ⓓ

यदि परीक्षार्थी द्वारा 75 से अधिक प्रश्नों को हल किया जाता है तो प्रारम्भिक हल किये हुए 75 उत्तरों को ही मूल्यांकन हेतु सम्मिलित किया जाएगा।

4. प्रत्येक प्रश्न के अंक समान हैं। आपके जितने उत्तर सही होंगे, उन्हीं के अनुसार अंक प्रदान किये जायेंगे।
5. सभी उत्तर केवल ओ०एम०आर० उत्तर-पत्रक (OMR Answer Sheet) पर ही दिये जाने हैं। उत्तर-पत्रक में निर्धारित स्थान के अलावा अन्यत्र कहीं पर दिया गया उत्तर मान्य नहीं होगा।
6. ओ०एम०आर० उत्तर-पत्रक (OMR Answer Sheet) पर कुछ भी लिखने से पूर्व उसमें दिये गये सभी अनुदेशों को सावधानीपूर्वक पढ़ लिया जाये।
7. परीक्षा समाप्ति के उपरान्त परीक्षार्थी कक्ष निरीक्षक को अपनी प्रश्नपुस्तिका बुकलेट एवं ओ०एम०आर० शीट पृथक-पृथक उपलब्ध कराने के बाद ही परीक्षा कक्ष से प्रस्थान करें।
8. निगेटिव मार्किंग नहीं है।
9. कोई भी रफ कार्य, प्रश्न-पुस्तिका में, रफ-कार्य के लिए दिए खाली पेज पर ही किया जाना चाहिए।
10. परीक्षा-कक्ष में लॉग-बुक, कैल्कुलेटर, पेजर तथा सेल्युलर फोन ले जाना तथा उसका उपयोग करना वर्जित है।
11. प्रश्न के हिन्दी एवं अंग्रेजी रूपान्तरण में भिन्नता होने की दशा में प्रश्न का अंग्रेजी रूपान्तरण ही मान्य होगा।

महत्वपूर्णः प्रश्नपुस्तिका खोलने पर प्रथमतः जाँच कर देख लें कि प्रश्नपुस्तिका के सभी पृष्ठ भलीभाँति छपे हुए हैं। यदि प्रश्नपुस्तिका में कोई कमी हो, तो कक्षनिरीक्षक को दिखाकर उसी सिरीज की दूसरी प्रश्नपुस्तिका प्राप्त कर लें।