Computational Biotechnology / Bioinformatics

Answer Key

27/06/2021

Gujarat Biotechnology Research Centre

Computational Biotechnology / Bioinformatics (Scientist B)

This question booklet contains 32 pages

Application No: ______________________

Time: 2 Hours

Total Marks: 200

Total Questions: 200

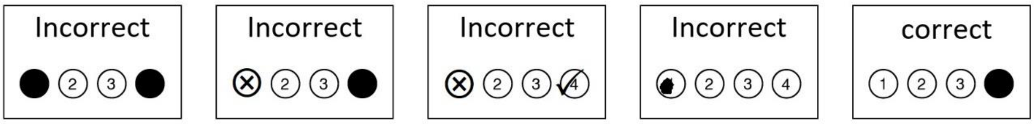
Candidate Signature : ....

Invigilator Signature : ....

Instructions for Candidate

---

1. This question booklet contains 200 questions.
2. Each correct answer carries 1 mark.
3. Use only Black Ball Point Pen to darken the appropriate circle in OMR.
4. Please darken the complete circle.
5. Darken ONLY ONE CIRCLE for each Question as shown below:



6. Answer once marked cannot be changed.
7. Please do not make any stray marks on the Question Booklet.
8. Rough works must be done on the blank page of Question Booklet.
9. Mark your answer in the appropriate space in the Answer Sheet against the Number corresponding to the question.
10. The Candidate is NOT allowed to carry Question booklet and OMR response sheet with him/her on conclusion of examination.

1 Which of the following is not used for visualizing Protein Interactions?

(A) VisANT
(B) Cytoscape
(C) BIND
(D) Osprey

2 Which is the first crop that was sequenced?

(A) Arabidopsis thaliana
(B) Oryza sativa
(C) Triticum aestivum
(D) Sorghum bicolor

3 A QTL is a

(A) Stretches of DNA (linking to a gene) underlying a quantitative trait
(B) Stretches of mapping regions of gene underlying a qualitative trait
(C) Stretches of molecular tags
(D) Qualified Trait at the Locus

4 Which of the following is linked to poor man's genome?

(A) CDNAs
(B) ESTs
(C) miRNAs
(D) LncRNAs

5 What is NGS

(A) Next Generation Sequencing
(B) Now Generation Sequencing
(C) New Generation Sequencing
(D) Novel Generation Sequencing

6 Statistical significance is arbitrary. Please justify with the following point

(A) P-value <0.05
(B) P-value >0.05
(C) P- value >0.01
(D) P-value <0.01

7 Who proposed a three-kingdom system of classification?

(A) Ernst Haeckel
(B) David H. Bergey
(C) Robert Whittaker
(D) Monera

8 Which of the following is a Cyanobacteria?

(A) Eubacteria
(B) Blue-green algae
(C) Blue fungus
(D) Black fungus

9 SARS-CoV-2 is

(A) Double stranded RNA virus
(B) Double stranded DNA virus
(C) Single stranded RNA virus
(D) Single stranded DNA virus

10 Gram staining is based on

(A) Chemical and physical properties of cell walls.
(B) Chemical and physical properties of their cytoplasm.
(C) Chemical and physical properties of cell metabolism.
(D) Chemical and physical properties of their cell matrix.

11 What is LMO

(A) lovely modified organism
(B) lonely modified organism
(C) living modified organism
(D) lively modified organism

12 Provirus is

(A) a virus genome which can integrate into host cell
(B) a virus genome which can integrate into DNA of host cell
(C) a virus genome which can integrate into RNA of host cell
(D) a virus genome which can integrate into mRNA of host cell

13 What is a Hypothetical Protein?

(A) Known knowns
(B) known unknowns
(C) unknown knowns
(D) unknown unknowns

14 What is Systems Biology?

(A) Systematic Biology
(B) System Integrated Science
(C) Biology of systems with special focus on interactions
(D) Biology of Systematic taxonomy

15 what is KDD

(A) Knowledge Driven Databases
(B) Knowledge Drug discovery
(C) Knowledge Discovery in Drugs
(D) Knowledge Discovery in Databases

16 Pharmacogenomics is

(A) identification of molecules through the genomic study
(B) identification of drugs through the genomic study
(C) identification of drugs through the Pharmacy
(D) identification of drugs through the Pharma industries

17 What is the organelle that produces proteins

(A) Golgi complex
(B) ER
(C) Mitochondria
(D) Ribosomes

18 What are the two approaches of systems biology?

(A) Dynamic and reductionist
(B) Dynamic and Scalar
(C) Reductionist and Reactive
(D) Reductionist and Pragmatic

19 What is SBML

(A) Systems Biology Make-up Language
(B) Systems Biology Mark-up Language
(C) Systems Biology Mask-up Language
(D) Systems Biology Maintaining-up Language

20 Ks/Ka score detects

(A) Conservation
(B) Non conservation
(C) Evolution
(D) Estimation

21 Which has the smallest genome

(A) Mycobacterium
(B) Mycoplasma
(C) Mycorrhiza
(D) Mycoses

22 How many protein coding genes are there in humans (approximately)

(A) 20,000
(B) 25,000
(C) 13,000
(D) 35,000

23 COPE is

(A) Committee on Publication Ethics
(B) Complex Pathway Engineering
(C) Container of Productive Encyclopaedia events
(D) Container of Publication Enhancement

24 Expasy protein database was launched in

(A) 1991
(B) 1992
(C) 1993
(D) 1994

25 2D Gel Electrophoresis is for

(A) separating DNA fragments
(B) separating RNA fragments
(C) separating different proteins
(D) separating ncRNAs

26 What if you want to compare two or more sequences? Which tool would you use ?

(A) Phylogeny/Parsimony
(B) BLAST
(C) PROTPARAM
(D) Unix

27 ANOVA is

(A) Analysis of Variance
(B) Analysis of Variability
(C) Analysis of Verification
(D) Analysis of Villifying genes

28 Which of the following hybrids will be most stable?

(A) AGTGTC
TCACAG
(B) ATAGGC
TATCCG
(C) GTTGGC
CAACCG
(D) ATTGGC
TTTCCG

29 Which algorithm is used by FASTA for similarity between query and any group of sequences of the same type

(A) Needleman–Wunsch algorithm
(B) Pearson-Lipman algorithm
(C) Smith–Waterman algorithm
(D) FASTA is only used to store nucleotide sequences

30 INSDC stands for

(A) International Nucleotide Sequence Database Collaboration
(B) International Nucleic Acid Database Centre
(C) International Nucleotide Sequence Data Centre
(D) Indian Nucleotide Sequence Data Copy

31 RefSeq records are owned by _____

1. NCBI
2. EMBL
3. DDBJ

(A) 1, 2 and 3
(B) 1 and 2
(C) Only 1
(D) Only 2

32 You are provided with assembly of unidentified prokaryotic organism and asked to do functional annotation. Which of the following you will not perform.

(A) Triming and filtering of poor quality bases
(B) diamond blastx against latest nr database
(C) upload on rast server
(D) blastx agaist KEGG database

33 Which of the following search program/database is not found at NCBI website

(A) LocusLink
(B) OMIM
(C) PSIPRED
(D) dbSNP

34 Which of the following best describes the scoring pattern by homology search program such as blast?

(A) Identical residue = 10 pts, gap = 0 pts
(B) Identical residue = 10 pts, conservative substitution = 5 pts, gap = 0 pts
(C) Identical residue = 10 pts, conservative substitution = 1-9 pts, gap = -3 pts
(D) Identical residue = 10 pts, conservative substitution = 10 pts, gap = -3 pts

35 Which of the following databases would you select for a blast search if you would like to obtain the most homologous protein sequence to your search sequence that has its 3D structure solved?

(A) NR
(B) PDB
(C) Swisprot
(D) PAT

36 A unit of measurement on physical maps is:

(A) Kilobases
(B) Centimorgans
(C) Cytological bands
(D) Centimeters

37 The "structural dogma"of molecular biology states that:

(A) DNA is transcribed to RNA, which is then translated to protein
(B) Sequence Determines three-dimensional structure, which determines function
(C) Protein structure can be grouped into primary (sequence), secondary (local structure), tertiary (three-dimensional structure) and quaternary (multi-pro complexes)
(D) Alpha helices and beta sheets can be determined by bond angles as displayed on a Ramachandran plot.

38 Consider the two DNA sequences CAGCAT and CGACAT. These sequences are aligned using a global alignment algorithm where the score is 1 for a match and 0 for a mismatch and there is a penalty of 0.2 for each gap. Which statement is true?

(A) There are no gaps in the optimal alignment
(B) The score of the optimal alignment is 4.0
(C) There are two optimal alignments with equal score
(D) The score of the optimal alignment is 4.8

39 The number of possible distinct unrooted trees with 5 species is

(A) 15
(B) 25
(C) 5!
(D) 4!

40 Which of these statements about phylogenetic methods is correct

(A) Bootstrapping can be used as a measure of confidence of the evolutionary model used in the phylogeny
(B) Bootstrapping cannot be done if the rate of substitution varies across sites
(C) Transition substitutions will usually saturate at a smaller divergence time than transversion substitutions
(D) Long branch attraction arises when the lengths of the sequences in the analysis vary a great deal

41 What of the following is not sequence data filtering tool?

(A) Blast
(B) PRINseqLite
(C) FastXTool kit
(D) CLC genomics Workbench

42 ____ is the OLC sequence assembler

(A) SOAPdenovo
(B) Velvet
(C) MIRA
(D) SPAdes

43 What is the average quality score you should keep if you are provided with ion torrent data

(A) 20
(B) 15
(C) 25
(D) 30

44 FASTQ file starts with______

(A) >
(B) @
(C) #
(D) $

45 What is the difference between BAM and SAM file

(A) BAM file is blasted sequence aligned map file while SAM is known sequence alignment file
(B) BAM is binary version of SAM file
(C) SAM is binary version of BAM file
(D) BAM file contains information about reference sequence while SAM file dose not

46 CIGAR stands for _______

(A) Compact Ideal Gap Penalty Alignment Report
(B) Compact Idiosyncratic Gapped Alignment Report
(C) Compact Integrated Alignment Report
(D) Computer Integrated Gapped Alignment Report

47 Co-assembly is _______

(A) Samples are first assembled individually and the the resulting sets of contigs are merged by combining contigs with semi-global identity
(B) Reads from all samples are pooled and a single assembly is performed
(C) Reads of each sample assembled individually and reference mapping performed
(D) Reads from each sample is assembled using different tools and merged together

48 Which of the following cannot be a python variable?

(A) __init_
(B) it
(C) on
(D) in

49 What will be the output of this expression, 3*1**3? if passed to the python interpreter

(A) 1
(B) 3
(C) 9
(D) 27

50 Which function is used to write a single character to console in C++?

(A) cout.put(ch)
(B) cout.putline(ch)
(C) write(ch)
(D) printf(ch)

51 Which of the following escape sequence represents carriage return?

(A) \r
(B) \n
(C) \n\r
(D) \c

52 Creator of C++?

(A) Ken Thompson
(B) Dennis Ritchie
(C) Brian Kernighan
(D) Bjarne Stroustrup

53 Which command is used to save and close the vi editor?

(A) q
(B) wq
(C) both q and wq
(D) alt + F4

54 What is /root?

(A) root filesystem
(B) the directory which contains all the directories of the filesystem
(C) home directory of the root user
(D) super user

55 What is the meaning of $ sign in awk programming?

(A) we are refering to a field or column in the current line
(B) $ sign is used for comment
(C) $ sign is used for variable
(D) the word following is the name of variable

56 What is sed?

(A) a hex editor
(B) an IDE
(C) a non-interactive stream editor
(D) a python interpreter

57 Which command is used to replace word ‘cat’ (already present in the file) with ‘mouse’ at all places in a file ‘old.txt’ and save the result in a new file ‘new.txt’?

(A) sed ‘s/cat/mouse’ old.txt new.txt
(B) sed ‘/s/cat/mouse/g’ old.txt new.txt
(C) sed ‘/s/cat/mouse’ old.txt > new.txt
(D) sed ‘s/cat/mouce/g’ old.txt > new.txt

58 The command “sed -n ‘/trinity/p’ abc.fasta” will

(A) print the lines containing the word ‘trinity’ in file abc.fasta
(B) delete the lines containing the word ‘trinity’ in file abc.fasta
(C) will generate an error message
(D) will remove all bad sequences from fasta file

59 Which command can create environment variable?

(A) set
(B) export
(C) read
(D) env

60 ‘kill %s’ command will

(A) terminate the last job
(B) terminate the first job
(C) terminate all jobs
(D) terminate the job whose command line starts with s

61 Indicate the right option to search for BOB, Bob, BOb or BoB?

(A) grep ‘[BOB]’ files
(B) grep –i Bob files
(C) grep -v ‘Bob’ files
(D) grep ‘B[oO][bB]’ files

62 BASH shell stands for?

(A) Big and Advanced Shell
(B) Basic to Advanced Shell
(C) Basic Access Shell
(D) Bourne-again Shell

63 Filesystem for CDROM is:

(A) Ext2
(B) Ext3
(C) Isofs
(D) Profs

64 Which file contains the filesystems to be automatically mounted during boot?

(A) /etc/mount
(B) /etc/mnt
(C) /etc/fstab
(D) /etc/boot

65 Which of the following represents the user home directory?

(A) /
(B) ~
(C) !
(D) ..

66 Which commands will give you information about how much disk space each file in the current directory uses?

(A) ls -l
(B) ls -la
(C) ls -a
(D) du

67 pwd command displays

(A) user password
(B) displays active directory path
(C) offers options to change current password
(D) shows path of current users home directory.

68 Which tar command option is used to list the files in a tape archive format?

(A) xvf
(B) ovf
(C) cvf
(D) tvf

69 Which command is used to extract intermediate result in a pipeline

(A) tee
(B) extract
(C) exec
(D) grep

70 Which command is used to perform backup in unix?

(A) backup
(B) zip
(C) gzip
(D) cpio

71 Which option of rm command is used to remove a directory with all its subdirectories

(A) -a
(B) -b
(C) -r
(D) -p

72 Which of the following are concatenation operators in perl script?

(A) +
(B) *
(C) .
(D) and

73 Which of the following expression matches a line which begins with a?

(A) /^a/
(B) /@a/
(C) /$a/
(D) /~a/

74 They primary R system is available from the ______

(A) GNU
(B) CRWO
(C) CRDO
(D) CRAN

75 In R language, a vector is defined that it can only contain objects of the _______

(A) Similar class
(B) Different Class
(C) Same Class
(D) Any class

76 A conservative mutation means:

(A) An amino acid change in a conserved region of the protein
(B) An amino acid change to an amino acid with similar size and chemical properties
(C) An amino acid change from acidic to basic
(D) An amino acid change from hydrophobic to hydrophilic

77 GFF is a file format for storing…

(A) Primary sequence data
(B) Three-dimensional molecular structures
(C) Genome annotations
(D) Results of microarray experiments

78 ENCODE stands for:

(A) Encyclopedia of database entries
(B) Encrypted oracle database encyclopedia
(C) Encyclopedia of DNA elements
(D) Eukaryotic normalized collection of DNA elements

79 A haplotype is:

(A) The set of polymorphic nucleotides found together on a single chromosome
(B) A genotype that is unique to non-African populations
(C) A genotype that is only found in a single individual in a population
(D) A set of diploid genotypes at two or more loci in an individual

80 Which of the following statements concerning the BLOSUM62 matrix below is correct ?

(A) Alanine is aligned with arginine more often than expected by chance.
(B) Alanine never changes to cysteine.
(C) Tryptophan evolves the slowest.
(D) The off-diagonal elements are proportional to the rates of substitution from one amino acid to another.

81 In a _____ protocol, bacteria with engineered abilities to detoxify pollutants are intentionally released in an area.

(A) Recombinant DNA
(B) Bioremediation
(C) Rhizosecretion
(D) Transgenic

82 Well-conserved regions in multiple sequence alignments denote:

(A) Areas of structural importance.
(B) Areas of functional importance.
(C) Areas of both functional and structural importance.
(D) Areas likely to be of functional and/or structural importance.

83 The transfer of genes from one phylogenetic group to another is called

(A) Endosymbiosis
(B) Southern hybridization
(C) Vertical transfer
(D) Horizontal transfer

84 Which of the following algorithm is useful in RNA structure prediction ?

(A) Needleman–Wunsch algorithm
(B) Nussinov algorithm
(C) Hirschberg's algorithm
(D) Viterbi algorithm

85 Which of the following tool/software replaces N ambiguty nucleotide code to Adenine before performing assembly?

(A) Velvet
(B) BWA
(C) SPAdes
(D) MIRA

86 Sequences obtained from impure nucleic acid preparations may contain DNA from sources other than the sample. Those sequence contaminations are a serious concern to the quality of the data used for downstream analysis, possibly causing erroneous conclusion

(A) Dinucleotide odds ratio
(B) Tetranucleotide odds ratio
(C) Pentanucleotide odds ratio
(D) Hexanucleotide odds ratio

87 Which of the following institute serves as a nodal point for Genome India initiative launched by DBT ?

(A) IISC-CBR
(B) GBRC
(C) CSIR-IGIB
(D) CSIR-CCMB

88 If you are provided with unknown assembly with more than 5000 contigs/scaffolds and asked to use it as a reference genome for transcriptome analysis using RNA STAR tool which of the following option will be helpful to reduce RAM consumption?

(A) genomeSAindexNbases
(B) genomeChrBinNbytes
(C) genomeChrBinNbits
(D) genomeSAindexNbase

89 With --quantMode ______ option STAR will count number reads per gene while mapping.

(A) ReadCounts
(B) MapCounts
(C) BaseCount
(D) GeneCounts

90 The combined mixture of all labeled DNA fragments is electrophoresed to _____ the fragments by______ and the ladder of fragments is scanned for the presence of each of the four labels.

(A) separate, size
(B) separate, pH
(C) assimilate, pH
(D) assimilate, size

91 Which of the given statements is untrue?

(A) There is no detailed check of sequence accuracy prior to submission to GenBank and other databases
(B) Often, a sequence is submitted at the time of publication of the sequence in a journal article, providing a certain level of checking by the editorial peer review process
(C) No sequence is submitted without being published or prior to publication
(D) In laboratories performing large sequencing projects, such as those engaged in the Human Genome Project or the genome projects of model organisms, the granting agency requires a certain level of accuracy of the order of 1 possible error per 10 kb

92 According to standard amino acid code letters which of the given pair is not right?

(A) K- lysine
(B) Y- tyrosine
(C) Q- glutamine
(D) R- serine

93 Genes that are transcribed are located in the ______

(A) euchromatin
(B) heterochromatin
(C) heterochromatin and euchromatin
(D) tightly bound DNA

94 Eukaryotic genes that encode proteins are interrupted by ________

(A) exons of varying length and number
(B) introns of varying length and number
(C) exons of varying length but same number
(D) introns of varying number but same length

95 In Protein 2D gel Electrophoresis, Individual proteins produced by the genome can be separated to _____ by this method and specific ones identified by various _______

(A) smaller extent, biochemical and immunological tests
(B) a large extent, biochemical and immunological tests
(C) a large extent, biochemical tests only
(D) smaller extent, purely mechanical tests

96 What is the length of a motif, in terms of amino acids residue?

(A) 30- 60
(B) 10-20
(C) 70- 90
(D) 44470

97 In regular expressions, which of the following pair of pattern is wrongly matched with its significance?

(A) [ ] – Or
(B) { } – Not
(C) ( ) – Repeats
(D) Z – Any

98 In terminologies related to regular expressions which of the following is false about terms and operators?

(A) Terms are strings or substrings
(B) Operators combine terms and expressions
(C) Operators do not have precedence
(D) Operators have precedence like arithmetic operators

99 What does this representation mean- R.L.[EQD]?

(A) An arginine- Amino acid- Leucine- Amino acid- Either Apartic acid, glutamic acid or glutamine
(B) An arginine- Leucine- Either Apartic acid, glutamic acid or glutamine
(C) An arginine- Leucine- Amino acid- Either Apartic acid, glutamic acid or glutamine
(D) An arginine- Leucine- Apartic acid and glutamic acid and glutamine

100 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

101 When did Needleman-Wunsch first describe the algorithm for global alignment?

(A) 1899
(B) 1970
(C) 1930
(D) 1950

102 Local alignments are more used when _____________

(A) There are totally similar and equal length sequences
(B) Dissimilar sequences are suspected to contain regions of similarity
(C) Similar sequence motif with larger sequence context
(D) Partially similar, different length and conserved region containing sequences

103 MEME stands for _____________

(A) Multiple Expectation Maximization for Motif Elicitation
(B) Multiple Expectation Maximization for Motif Extraction
(C) Mega Expectation Maximization for Motif Elicitation
(D) Micro Expectation Maximization for Motif Extraction

104 For significantly aligning sequences what is the resulting structure on the plot?

(A) Intercrossing lines
(B) Crosses everywhere
(C) Vertical lines
(D) A diagonal and lines parallel to diagonal

105 The matrices PAM250 and BLOSUM62 contain _______

(A) positive and negative values
(B) positive values only
(C) negative values only
(D) neither positive nor negative values, just the percentage

106 By whom and when were the Bayesian methods applied first?

(A) Smith-Waterman, 1981
(B) Agarwal and States, 1996
(C) Smith-Waterman, 1996
(D) Agarwal and States, 1981

107 CLUSTALW is a more recent version of CLUSTAL with the W standing for ________

(A) weakening
(B) winding
(C) weighting
(D) wiping

108 In Smith–Waterman algorithm, in initialization Step, the _________ row and ________ column are subject to gap penalty.

(A) first, first
(B) first, second
(C) second, First
(D) first, last

109 In pharmacokinetics, clearance refers to:

(A) Approval by the Food and Drug Administration
(B) Elimination of drugs from the body
(C) Marketing for the sale of pharmaceutical agents
(D) Removal of expired drugs from the pharmacy

110 What is the largest amino acid, often found in the hydrophobic core, and usually very conserved

(A) W, tryptophan
(B) G, glycine
(C) A, alanine
(D) L, leucine

111 Which of the following cloning vectors is designed to accommodate the largest size DNA insert ?

(A) Plasmid
(B) Phage
(C) Cosmid
(D) YAC

112 The bulk of RNA in cells is

(A) Non-coding RNA
(B) Coding RNA
(C) mRNA
(D) Transfer RNA

113 What is the maximum quality score value PacBio read encoding supports?

(A) 33
(B) 40
(C) 41
(D) 93

114 Diamond blast uses _______ as a database

(A) Nucleotide fasta
(B) Protein Fasta
(C) Nucleotide alignment
(D) Protein alignment

115 Which of the following sign can be used at the end of the command to send command in background?

(A) >
(B) |
(C) and
(D) $

116 Which of the following tool/software is used to assess genome completness?

(A) BUSCO
(B) MIRA
(C) Velvet
(D) TCOFEE

117 Bioedit is used for

(A) Protein and DNA sequence editing
(B) Powerful structure analysis program
(C) Biological system tool
(D) Computational language

118 What are Degenerate Primers?

(A) Primers of consensus sequence
(B) Primers of Unidentified nucleotide regions
(C) Primers from Untranslated regions
(D) Primers from amino acid sequence

119 What is exon shuffling?

(A) The transfer of the exons
(B) A mechanism, which joins exons from different genes to generate more transcripts
(C) The shuffling of gene sequences
(D) The change in structure of exons

120 What will heterozygous single nucleotide substitution look like on your chromatogram?

(A) Two peaks of equal height at the same position
(B) One peak twice the height of those around it
(C) Two peaks in the same position, one twice the height of the other
(D) Three peaks of equal height at the same position

121 _______ is linked with conformational searching?

(A) DOCK
(B) LUDI
(C) Monte Carlo method
(D) CoMFA

122 RMSD is used for ______________________

(A) Quantitative measure of similarity and differences between two or more protein structures
(B) Quantitative measure of similarity between two or more protein structures
(C) Qualitative measure of similarity between two or more protein structures
(D) Qualitative measure of similarity and differences between two or more protein structures

123 Which of the following is untrue about LAGAN?

(A) regions that have high density of words are selected as anchor
(B) web based program designed for pairwise alignment of small fragments of genomes only
(C) Limited Area Global Alignment of Nucleotides
(D) it is first anchors between two genomic sequences using algorithm that identifies short, exactly matching words

124 Which is model organism databases?

(A) PomBase
(B) Ensembl
(C) ArrayExpress
(D) GenBank

125 Which of the following tools is used for Analysis of Gene Expression (SAGE)?

(A) SAGEmap
(B) SAGExProfiler
(C) GonadSAGE
(D) DNAarray

126 Which version of Pfam was released in March, 2021 ?

(A) Pfam 34.0
(B) Pfam 33.1
(C) Pfam 32.0
(D) Pfam 31.0

127 In CpG island p refers to____________?

(A) phosphodiester bond connecting the two nucleotides (C and G)
(B) P value between C and G
(C) Required energy to breakdown of phosphodiester bond between C and G
(D) strength between C and G

128 In R, missing values are represented as _________

(A) NA
(B) Na_N
(C) 0
(D) MV

129 To test whether perl is in your PATH, use ____

(A) perl -i
(B) perl -l
(C) perl -el
(D) perl -e

130 ________ is a collection of perl modules that facilitate the development of Perl scripts for bioinformatics application

(A) Bioperl
(B) Btperl
(C) infoperl
(D) Genomeperl

131 What is the command in linux to view all the tcp ports?

(A) Netstat -a
(B) netstat -at
(C) netstat --list
(D) netstat -t

132 What is the key combination in linux to quit the file without saving in vim?

(A) Ctrl + c
(B) Ctrl + .
(C) :qa
(D) :qa!

133 What is the command to create virtualenv in python?

(A) virtualenv projectname
(B) venv projectname
(C) python --venv projectname
(D) python -v projectname

134 What is the type of sysargs?

(A) Set
(B) list
(C) string
(D) object

135 When to use a stream editor in Linux? And how? Select right option:

(A) For editing large files use sad command
(B) Only for editing large python scripts use sed command
(C) For editing large files use sed command
(D) For editing files which are too large for computer use sad command

136 Which is not an Indian Super computer?

(A) Colour Boson
(B) PARAM YUVA II
(C) Theta
(D) Virgo

137 Chronogram is

(A) a phylogenetic tree that explicitly represents time through its branch lengths
(B) a diagram representing a cross section of a phylogenetic tree
(C) phylogenetic tree that has branch lengths proportional to the amount of character change
(D) only represents a branching pattern

138 A tool that combines different protein signature recognition methods.

(A) RADAR
(B) MSA
(C) InterProScan
(D) PfamScan

139 UniRef50 is generated by ___________

(A) clustering only UniProt Knowledgebase records
(B) clustering UniProt Knowledgebase records plus selected UniParc records
(C) clustering UniRef90 seed sequences.
(D) clustering UniRef100 seed sequences

140 Average Read length Pac Bio platform is

(A) 0.2-1kb
(B) 1-2Kb
(C) 2-5Kb
(D) 8-15kb

141 Orthologs are defined as

(A) Homologous sequences in different species that share an ancestral gene
(B) Homologous sequences that share little amino acid identity but share great structural similarity
(C) Homologous sequences in the same species that arose through gene duplication
(D) Homologous sequences in the same species which have similar and often redundant functions

142 You have two distantly related proteins. Which BLOSUM or PAM matrix is best to use to compare them?

(A) BLOSUM45 or PAM250
(B) BLOSUM45 or PAM10
(C) BLOSUM80 or PAM250
(D) BLOSUM80 or PAM10

143 Normalized BLAST scores (also called bit scores):

(A) Are unitless
(B) Are not related to the scoring matrix that is used
(C) Can be compared between different BLAST searches, even if different scoring matrices are used
(D) Can be compared between different BLAST searches, but only if the same scoring matrices are used

144 How does BLAT differ from BLAST?

(A) BLAT includes both global and local alignment.
(B) BLAT employs a database that is parsed into a set of words that are matched to the DNA query.
(C) BLAT only identifies genomic regions that exactly match a query sequence
(D) BLAT cannot accept a protein sequence as a query

145 Which one of the following is a character-based phylogenetic algorithm?

(A) Neighbor joining
(B) Kimura
(C) Maximum likelihood
(D) PAUP

146 In analyzing cDNA libraries, a pitfall is that:

(A) The libraries may be derived from different tissues.
(B) The libraries may contain thousands of sequences.
(C) The libraries may have been normalized differently
(D) The libraries may contain many rarely expressed transcripts.

147 A self-organizing map:

(A) Imposes some structure on the formation of clusters
(B) Is unstructured, like k-means clustering
(C) Has neighboring nodes that represent dissimilar clusters-
(D) Cannot be represented as a clustering tree

148 In general, if you compare the size of a pattern (also called a motif or fingerprint) and a domain:

(A) They are about the same size.
(B) The pattern is larger.
(C) The pattern is smaller
(D) The comparison always depends on the particular proteins in question.

149 An underlying assumption of the Gene Ontology Consortium is that the description of a gene or gene product according to three categories (molecular function, biological process, and cellular component):

(A) Is likely to be identical across many species, from plants to worms to human
(B) Is likely to vary greatly across many species, from plants to worms to human
(C) May or may not be identical across many species and thus must be assessed for each gene or gene product individually
(D) May or may not be identical across many species and thus must be assessed for each gene or gene product individually by an expert curator

150 Protein localization is described primarily in which Gene Ontology category?

(A) Molecular function
(B) Cellular component
(C) Cellular localization
(D) Biological process

151 Protein secondary structure prediction algorithms typically calculate the likelihood that a protein:

(A) Forms alpha helices
(B) Forms alpha helices and beta sheets
(C) Forms alpha helices, beta sheets and coils
(D) Forms alpha helices, beta sheets, coils and multimeric

152 An advantage of x-ray crystallography relative to NMR for structure determination is that using x-ray crystallography:

(A) It is easier to solve the structure of transmembrane domain containing proteins.
(B) It is easier to grow crystals than to prepare samples for NMR.
(C) It is easier to interpret diffraction data
(D) It is easier to determine the structures of large proteins

153 Homology modeling may be distinguished from ab initio prediction because:

(A) Homology modeling requires a model to be built
(B) Homology modeling requires alignment of a target to a template.
(C) Homology modeling is usefully applied to any protein sequence
(D) The accuracy of homology modeling is independent of the percent identity between the target and the template.

154 Cn3D is a molecular structure viewer at NCBI. It features

(A) A menu-driven program linked to automated homology modeling
(B) A command line interface useful for a variety of structure analyses
(C) A structure viewer that is accompanied by a sequence viewer
(D) A structure viewer that allows stereoscopic viewing of structure images

155 Hub proteins are

(A) Proteins that occur at nodes that are highly connected within a protein network.
(B) Proteins that occur at edges that are highly connected within a protein network.
(C) Proteins that occur at nodes that are sparsely connected within a protein network
(D) Proteins that occur at edges that are sparsely connected within a protein network.

156 The biggest problem in predicting protein-coding genes from genomic sequences using algorithms is that:

(A) The software is difficult to use
(B) The false negative rate is high: many exons are missed
(C) The false positive rate is high: many exons are falsely assigned.
(D) The false positive rate is high: many exons have unknown function

157 The C value paradox is that

(A) The nucleotide C is underrepresented in some genomes.
(B) The genome size of various eukaryotes correlates poorly with the number of protein-coding genes of the organism.
(C) The genome size of various eukaryotes correlates poorly with the biological complexity of the organism
(D) The genome size of various eukaryotes correlates poorly with the evolutionary age of the organism

158 You are sequencing the genome of a newly described organism (a slime mold). What is likely to happen if you use RepeatMasker to assess its repetitive DNA content? You set the default setting of RepeatMasker to the settings for human DNA

(A) RepeatMasker should successfully identify essentially all of the repetitive DNA. Various repetitive DNA elements are similar enough between organisms to allow this software to work on your slime mold DNA.
(B) RepeatMasker should identify most of the repetitive DNA. However, because some types of repeats are species specific, it is likely that there will be many false positive and false negative results
(C) RepeatMasker would fail to identify most of the repetitive DNA. Most types of repeats are highly species specific. It is necessary for you to train the RepeatMasker algorithm on your slime mold DNA in order for the program to work
(D) It is not possible to predict, because repetitive DNA may or may not be variable between organisms.

159 Single nucleotide polymorphism (SNP) arrays can reliably detect all of the following phenomena except_______________

(A) Deletions
(B) Duplications
(C) Inversions
(D) Uniparental isodisomy

160 How are the mouse and human genomes different?

(A) The mouse genome has a lower GC content
(B) The mouse genome has more protein-coding genes
(C) The mouse genome has undergone specific expansions of genes encoding particular protein families such as olfactory receptors
(D) The mouse genome has fewer telomeric repeats per chromosome, on average.

161 Many features distinguish the chimpanzee and human genomes, including all of the following except which one?

(A) Chimpanzees have more chromosomes.
(B) About 35 million nucleotide substitutions have been described
(C) There have been hundreds of pericentric inversions.
(D) Over 500 chimpanzee-human ortholog pairs may be under positive selection

162 Approximately what percentage of the human genome consists of repetitive elements of various kinds?

(A) 5%
(B) 25%
(C) 50%
(D) 85%

163 Online Mendelian Inheritance in Man (OMIM) includes entries that focus on:

(A) Particular diseases
(B) Particular genes
(C) Either genes or diseases
(D) Complex chromosomal disorders

164 The Twilight Zone is the region of sequence similarity

(A) above 50% identity
(B) where sequence alignments are not statistically significant, as the same alignment may have arisen by chance
(C) below 50% identity
(D) where sequences fail to be detected by even the most sensitive sequence-based search algorithms

165 Which of the following is not the Sequence Read Quality Control tool

(A) FastQC
(B) BBDuk
(C) GenomeQC
(D) Khmer

166 ___________Predicts genes from short reads incorporating a sequence error model and codon usage statistics.

(A) Glimmer-MG
(B) FragGeneScan
(C) Prodigal
(D) Prokka

167 DADA2 is used in QIIME 2 for

(A) cluster sequences
(B) dereplicate
(C) denoising
(D) Classify Sequences

168 ____________allows users to perform abundance estimation with Kraken results. It uses a bayesian formula to estimate species/genus-level abundance from Kraken classification results.

(A) Bracken
(B) Pavian
(C) BLAT
(D) Bowtie2

169 A lead compound is___________

(A) A compound from the research laboratory that is chosen to process forward for preclinical and clinical trials
(B) lead containg compund
(C) A molecule that shows some activity or property of interest and serves as the starting point for the development of a drug.
(D) The first compound of a structural class of compounds to reach the market.

170 Which of the following terms is used to describe a drug that has the same effect on a receptor as the endogenous chemical messenger?

(A) agonist
(B) antagonist
(C) partial agonist
(D) inverse agonist

171 Which of the following is one of the rules in Lipinski's rule of five?

(A) A molecular weight equal to 500
(B) No more than five hydrogen bond acceptor groups
(C) No more than 10 hydrogen bond donor groups
(D) A calculated logP value less than +5

172 A good way to increase total proteome penetration by gel-free LC-MS/MS methods is to:

(A) Use two, orthogonal types of chromatography
(B) Enrich for phosphopeptides only
(C) Analyze whole proteins
(D) Label the proteins with a chemical tag

173 The higher is the score in the alignment _________

(A) the more significant is the alignment
(B) or the less it resembles alignments in related proteins
(C) the less significant is the alignment
(D) the less it aligns with the related protein sequence

174 Write full form of NAG

(A) N-Acetyly Gluocsamine
(B) N-Amyl Glucosamine
(C) N-Adenly glucosamine
(D) N-Artero Glucosamine

175 Which of the following is correct

(A) Microarrays are arrays where DNA oligonucleotides of DNA sequences are spotted as a matrix.
(B) Microarrays are arrays where RNA oligonucleotides of RNA sequences are spotted as a matrix.
(C) Microarrays are arrays where RNA oligonucleotides of DNA sequences are spotted as a matrix.
(D) Microarrays are arrays where DNA oligonucleotides of RNA sequences are spotted as a matrix.

176 Match the following

P. GeneMark	A. a graphical analysis tool which finds all open reading frames
Q. Geneparser	B. Family of gene prediction programs
R. GLIMMER	C. finding genes in microbial DNA
S. ORF FINDER	D. Parse a DNA sequence

(A) P->B
Q->D
R->C
S->A
(B) P->C
Q->D
R->B
S->A
(C) P->D
Q->B
R->C
S->A
(D) P->C
Q->B
R->D
S->A

177 Which of the following problem could not be done with HMM?

(A) Gene coding sequence identification
(B) Maximum Likelihood phylogeny estimation
(C) Pairwise alignment of DNA sequences
(D) Transcription factor binding site prediction

178 In R language, a vector is defined that it can only contain objects of the _______

(A) Similar class
(B) Different Class
(C) Same Class
(D) Any class

179 Which of the given statements is incorrect?

(A) The simplest and newest way of submitting sequences is through the Web site on a Web form page called BankIt
(B) The sequence can also be annotated with information about the sequence, such as mRNA start and coding regions
(C) The submitted form is transformed into GenBank format and returned to the submitter for review before being added to GenBank
(D) Sequin does not run on UNIX

180 Knowing ________ should be enough to find the required entry quickly.

(A) publication date, protein name, journal name
(B) accession number, protein name, or name of gene
(C) publication date, protein name, or volume
(D) properties, protein name, or title word

181 Which of the following does not describe BLOSUM matrices?

(A) It stands for BLOcks SUbstitution Matrix
(B) It was developed by Henikoff and Henikoff
(C) The year it was developed was 1992
(D) These matrices are logarithmic identity values

182 Gaps are added to the alignment because it ______

(A) increases the matching of identical amino acids at subsequent portions in the alignment
(B) increases the matching of or dissimilar amino acids at subsequent portions in the alignment
(C) reduces the overall score
(D) enhances the area of the sequences

183 In a protein family, key amino acids like those involved in the catalytic sites are:

(A) Are highly conserved because of their importance
(B) Are not especially conserved or variable
(C) Are often mutated so the protein function can evolve
(D) Are Allowed only a very specific type of variation.

184 Which of the following statements about restriction endonucleases is not true

(A) The enzymes are natural products of bacteria.
(B) The enzymes cut DNA at defined sequences.
(C) The enzymes cut only single-stranded DNA molecules.
(D) The enzymes are named because they restrict the growth of bacteriophage.

185 What will be the output for the following command?

```
perl prinseq-lite.pl -fastq sample.fastq  -min_len 50 -trim_left 5 -out_good good.fastq
```

(A) All reads will be trimmed from left in the window size of 5 Until length of 50 not achived
(B) All reads larger than 50 bp will be trimed to achive length of 50bp from left side
(C) All reds will be trimmed for 5 bp from length and then reads shorter than 50 will be discarded
(D) All reads shorter than 50 will be discarded and then trimed from left end by 5 bp each

186 Which of the following event is not a Post Translational Modification?

(A) Proteolytic cleavage
(B) Methylation of the nascent polypeptide chain
(C) Phosphorylation of the nascent polypeptide chain
(D) Splicing

187 What is base calling?

(A) Identify the relative positions of genetic markers on a chromosome and are based on how frequent the markers are inherited together.
(B) refers to the analysis of global gene expression and gene functions in a genome
(C) Refer to banding patterns seen on stained chromosomes, which can be directly observed under a microscope.
(D) The fluorescent traces of the DNA sequences are read by a computer program that assigns bases for each peak in a chromatogram.

188 For Needleman-Wunsch algorithm Which of the following statement is not an advantage ?

(A) New algorithmic improvements as well as increasing computer capacity
(B) Make it possible to align a query sequence against a large DB in a few minutes
(C) This does not help in determining evolutionary relationship
(D) If you have two genes that are well understood as closely related, then this type of algorithm can be used to understand them in further details

189 Which of the following sentences best describes the difference between a global alignment and a local alignment between two sequences?

(A) Global alignment is usually used for DNA sequences, while local alignment is usually used for protein sequences.
(B) Global alignment has gaps, while local alignment does not have gaps
(C) Global alignment finds the global maximum, while local alignment finds the local maximum
(D) Global alignment aligns the whole sequence, while local alignment finds the best subsequence that aligns

190 Which output from a BLAST search provides an estimate of the number of false positives from a BLAST search?

(A) E value
(B) Bit score
(C) Percent identity
(D) Percent positives

191 In comparing two homologous but distantly related proteins:

(A) They tend to share more three-dimensional structure features in common than percent amino acid identity.
(B) They tend to share more percent amino acid identity in common than three-dimensional structure features.
(C) They tend to share three-dimensional structure features and percent amino acid identity to a comparable extent.
(D) It is not reasonable to generalize about the extent to which they share three-dimensional structure features and percent amino acid identity

192 Hundreds or thousands of sequence repeats, each of a unit of about four to eight nucleotides, are commonly found where?

(A) In interspersed repeats
(B) In processed pseudogenes
(C) In telomeres
(D) In segmentally duplicated regions

193 DIAMOND, an open-source algorithm based on double indexing that is ____________times faster than BLASTX on short reads and has a similar degree of sensitivity.

(A) 20
(B) 200
(C) 2000
(D) 20,000

194 Which of the following descriptions most accurately describes binding sites and binding regions?

(A) a binding site is part of a binding region
(B) a binding region is part of a binding site
(C) a binding region is the same as a binding site
(D) a binding region is on a drug whereas a binding site is on a macromolecular target

195 LUDI used for___________________

(A) Creating 3D models of molecules
(B) Docking of molecules into binding sites
(C) Conformational study
(D) Automated de novo drug design

196 ______________ is for the assembly of WGS reads that accepts base calls along with quality scores assigned by Phred as input

(A) VecScreen
(B) ARACHNE
(C) EULER
(D) Needleman-Wunsch

197 Which of the following is not true about Needleman-Wunsch algorithm?

(A) Invented by Richard Bellman in 1953
(B) Dynamic programming
(C) In 1970, first used by Saul Needleman and Christian Wunsch
(D) scored of gaps -2

198 Choose correct answer for Gen Threader.

(A) web-based program and it uses profile methods only
(B) It uses pairwise energy methods only
(C) query protein sequence resulting multiple sequence hits that used to generate a profile and its secondary structure is predicted using PSIPRED
(D) Initial step is quite dissimilar to 3D-PSSM

199 Which of the following is not right regarding expressed sequence tags (ESTs)?

(A) Short sequences obtained from cDNA clones
(B) One of the high throughput approaches of wide profiling of genome's gene expression
(C) They are typically in the range of 200 to 400 nucleotides in length
(D) Serve as long identifiers of full-length genes

200 Study of the organismal variation in phenotype as it changes during its life span known as_____________

(A) Metatranscriptomics
(B) Phenomics
(C) Metabolomics
(D) Genomics

Q.No	Option	Q.No	Option	Q.No	Option	Q.No	Option
1	C	51	A	101	B	151	C
2	B	52	D	102	A	152	D
3	A	53	B	103	A	153	B
4	B	54	C	104	D	154	C
5	A	55	B	105	A	155	A
6	A	56	C	106	B	156	C
7	A	57	D	107	C	157	C
8	B	58	A	108	A	158	B
9	C	59	B	109	B	159	C
10	A	60	D	110	A	160	C
11	C	61	D	111	D	161	C
12	A	62	D	112	A	162	C
13	B	63	C	113	D	163	C
14	C	64	C	114	B	164	B
15	A	65	B	115	C	165	C
16	B	66	D	116	A	166	B
17	D	67	B	117	A	167	C
18	A	68	D	118	D	168	A
19	B	69	A	119	B	169	C
20	C	70	D	120	A	170	A
21	B	71	C	121	C	171	D
22	A	72	C	122	B	172	A
23	A	73	A	123	B	173	A
24	C	74	D	124	A	174	A
25	C	75	C	125	D	175	A
26	A	76	B	126	A	176	A
27	A	77	C	127	A	177	B
28	C	78	C	128	A	178	C
29	B	79	A	129	D	179	D
30	A	80	C	130	A	180	B
31	C	81	B	131	B	181	D
32	A	82	D	132	D	182	A
33	C	83	D	133	A	183	A
34	C	84	B	134	B	184	C
35	B	85	A	135	C	185	C
36	A	86	A	136	C	186	D
37	B	87	A	137	A	187	D
38	C	88	C	138	C	188	C
39	A	89	D	139	C	189	D
40	C	90	A	140	D	190	C
41	A	91	C	141	A	191	A
42	C	92	D	142	A	192	C
43	A	93	A	143	C	193	D
44	B	94	B	144	B	194	B
45	B	95	B	145	C	195	D
46	B	96	B	146	C	196	B
47	B	97	D	147	A	197	D
48	D	98	C	148	C	198	C
49	B	99	A	149	C	199	D
50	A	100	D	150	B	200	B