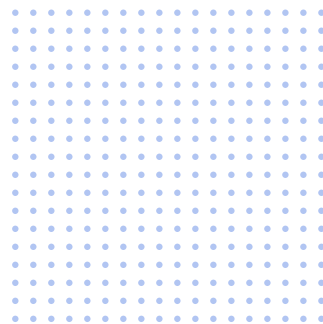




FOR JNU EE EXAM PREPARATION

JNU EE 2022

Question Paper · Ph.D

Computational Biology

and Bioinformatics -

Track 2 Bioinformatics

EXAM	YEAR	TYPE
JNU EE	2022	Question Paper

SUBJECT

Ph.D Computational Biology and Bioinformatics - Track 2 Bioinformatics

Test Date : 10 Dec 2022**Test Slot : Slot 1****Subject : TRBB-Computational Biology & Bioinformatics Track -2 Bio Informatics****Sl. No.1****QBID:1112141**

The term 'N50' in genome assembly refers to

- 1) The largest value of N for which 50% of the base pairs in the bin is in supercontigs with length N base pairs or longer.
- 2) The smallest value of N for which 50% of the base pairs in the bin is in supercontigs with length N+10 base pairs or longer.
- 3) 50% of bases in the whole genome.
- 4) 50% of correct bases in the whole genome.

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17361]

2[Option ID=17362]

3[Option ID=17363]

4[Option ID=17364]

Sl. No.2**QBID:1112142**

Pathway enrichment means

- 1) More genes than expected in a gene list coming from that pathway.
- 2) Addition of new genes to a given pathway
- 3) Many pathways being activated.
- 4) Over-expression of all genes from that pathway in an experiment.

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17365]

2[Option ID=17366]

3[Option ID=17367]

4[Option ID=17368]

Sl. No.3**QBID:1112143**

What does RPKM counts obtained in RNA sequencing experiment stand for?

- 1) Reads per K-mers
- 2) Reads per kilobase per million
- 3) Reads per kilobase per machine
- 4) Reads per kilobyte and megabyte

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17369]

2[Option ID=17370]

3[Option ID=17371]

4[Option ID=17372]

Sl. No.4

QBID:1112144

You have sequenced a genome. Following assembly, the draft genome has a size of 5,000,000 bp. The raw output from the assembly is 2,500,000 reads. The average read length is 100 bp. The depth of coverage is

- 1) 200x
- 2) 50x
- 3) 125x
- 4) 150x

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17373]

2[Option ID=17374]

3[Option ID=17375]

4[Option ID=17376]

Sl. No.5

QBID:1112145

Which of these projects would be best suited for Next Generation Sequencing?

- 1) To determine if a tumour sample contains a common missense mutation
- 2) To find the transcriptome of a tumour sample.
- 3) To genotype ten genomic DNA samples for a known single nucleotide polymorphism
- 4) To identify restriction sites

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17377]

2[Option ID=17378]

3[Option ID=17379]

4[Option ID=17380]

Sl. No.6

QBID:1112146

The microRNA and lncRNA can be best defined (OR) categorized in terms of

1. Stem loop structure
2. Base composition
3. Coding potential
4. Canonical base

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17381]

2[Option ID=17382]

3[Option ID=17383]

4[Option ID=17384]

Sl. No.7

QBID:1112147

In the context of phylogenetic tree, Let S be set of points. Any distance measure on the set S that satisfies the requirement that of the three distances between any three members of the set; two are equal and one greater than the third is called as

1. additive
2. ultrametric
3. parametric
4. normal

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17385]

2[Option ID=17386]

3[Option ID=17387]

4[Option ID=17388]

Sl. No.8

QBID:1112148

In BLAST software tool-based sequence homology searches, one of the parameters is word size. What happens when it is reduced from 20 to 12?

1. Search hits may increase
2. Search hits may decrease
3. Search hits won't change
4. Search hits will be ordered

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17389]

2[Option ID=17390]

3[Option ID=17391]

4[Option ID=17392]

Sl. No.9

QBID:1112149

BLAST search algorithm uses _____ strategy

1. Dynamic programming
2. Word based heuristic search
3. N - dimensional search
4. Genetic algorithm based search

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17393]

2[Option ID=17394]

3[Option ID=17395]

4[Option ID=17396]

Sl. No.10

QBID:1112150

The commonly used target database for BlastP search is

1. PDB
2. Swiss-prot
3. GenBank
4. EMBL

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17397]

2[Option ID=17398]
 3[Option ID=17399]
 4[Option ID=17400]

Sl. No.11
 QBID:1112151

Assume that for a given DNA sequence i.e

ACTACTATACTAATACTAGCTT

how many dinucleotides can be considered if we slide the window in an overlapping fashion?

1. 20
2. 21
3. 19
4. 11

- (1) 1
- (2) 2
- (3) 3
- (4) 4

1[Option ID=17401]
 2[Option ID=17402]
 3[Option ID=17403]
 4[Option ID=17404]

Sl. No.12
 QBID:1112152

For a typical Markov Chain with "m" states and "k" th order there are $p = m^k (m-1)$ independent parameters to be estimated. In that case for a typical DNA sequence of A, T, G, C and for the first order Markov Chain, there are ____ independent parameters

1. 3
2. 48
3. 12
4. 4

- (1) 1
- (2) 2
- (3) 3
- (4) 4

1[Option ID=17405]
 2[Option ID=17406]
 3[Option ID=17407]
 4[Option ID=17408]

Sl. No.13
 QBID:1112153

Position weight matrix can be derived or constructed from

1. Pairwise alignment
2. Phylogenetic tree
3. Multiple sequence alignment
4. Prosite database search hits

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17409]

2[Option ID=17410]

3[Option ID=17411]

4[Option ID=17412]

Sl. No.14

QBID:1112154

For the given four pairwise sequence alignments, consider the score for identities = 6; mismatches = 3; gap opening penalty = 2 and gap extension penalty = 0.5.

You may use the Scoring function "S" for the alignment as

$S = [\text{Number of matches/mismatches}] - [\text{opening gap penalty} + (\text{no of gaps} - 1) \times \text{gap extension penalty}]$

Using the above scoring function, which pair shows the highest similarity?

1) A G T C T A C
C G T - - A C

2) C T - - A A T
C T G T A A -

3) G T A C G T T
G - - - G T T

4) A C C - - T A
A C C T T T A

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17413]

2[Option ID=17414]

3[Option ID=17415]

4[Option ID=17416]

Sl. No.15

QBID:1112155

In a typical Neural Network framework, the major objective is to minimize

1. the weights
2. the input values
3. the error due to (output-target) values
4. learning rate

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17417]

2[Option ID=17418]

3[Option ID=17419]

4[Option ID=17420]

Sl. No.16

QBID:1112156

In the context of protein structures, which one of the following is not considered

1. tertiary
2. Quaternary
3. Hexamery
4. secondary

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17421]

2[Option ID=17422]

3[Option ID=17423]

4[Option ID=17424]

Sl. No.17

QBID:1112157

If we consider a true random sequence of dsDNA, then which one of the following nucleotide compositions is correct

1. $n(A) = n(T) \text{ \& } n(C) \neq n(G)$
2. $n(A) \neq n(T) \text{ \& } n(C) = n(G)$
3. $n(A) = n(T) \text{ \& } n(C) = n(G)$
4. $n(A) \neq n(T) \text{ \& } n(C) \neq n(G)$

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17425]

2[Option ID=17426]

3[Option ID=17427]
4[Option ID=17428]

Sl. No.18
QBID:1112158

A statistical or machine learning procedure that does require annotated training data to analyse new (or) unannotated data is called as

1. Incremental learning
2. Supervised learning
3. Linear learning
4. Unsupervised learning

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17429]
2[Option ID=17430]
3[Option ID=17431]
4[Option ID=17432]

Sl. No.19
QBID:1112159

Which one of the following is not gene prediction program (or) computational tool?

1. GTRAIL
2. GenMark
3. Glimmer
4. Markov

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17433]
2[Option ID=17434]
3[Option ID=17435]
4[Option ID=17436]

Sl. No.20
QBID:1112160

Which one of the following is not related to eukaryotic gene structure?

1. Donor site
2. Splice site
3. Prime site
4. Acceptor site

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17437]

2[Option ID=17438]

3[Option ID=17439]

4[Option ID=17440]

Sl. No.21

QBID:1112161

Which one of the following data formats (file extension) is not linked to any gene or genome data?

1. gff
2. fna
3. gft
4. fasta

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17441]

2[Option ID=17442]

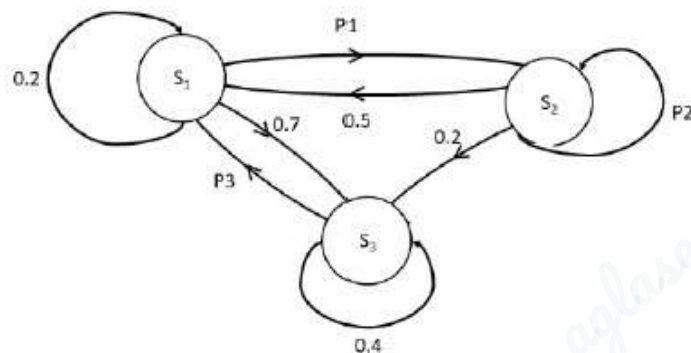
3[Option ID=17443]

4[Option ID=17444]

Sl. No.22

QBID:1112162

A Markov Chain model can be depicted in a graphical model as shown below with "states" as nodes and the transition probabilities with edges



The transition probabilities marked by p1, p2, p3 would be

1. 0.2, 0.4, 0.5
2. 0.1, 0.3, 0.6
3. 0.1, 0.3, 0.4
4. 0.2, 0.7, 0.1

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17445]

2[Option ID=17446]

3[Option ID=17447]

4[Option ID=17448]

Sl. No.23

QBID:1112163

Which one of the following is not a probability distribution?

1. Extreme value distribution
2. Geometric distribution
3. Trans normal distribution
4. Lognormal distribution

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17449]

2[Option ID=17450]

3[Option ID=17451]

4[Option ID=17452]

Sl. No.24

QBID:1112164

Which one of the following is not a palindromic DNA sequence?

1. T A T A T A
2. G A T A T C
3. C T A T A G
4. A T G C G C

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17453]

2[Option ID=17454]

3[Option ID=17455]

4[Option ID=17456]

Sl. No.25

QBID:1112165

Which software can be used to visualize and calculate properties of complex biological networks?

1. Runge-kutta method
2. Cobra
3. Cytoscape
4. Bionet

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17457]

2[Option ID=17458]

3[Option ID=17459]

4[Option ID=17460]

Sl. No.26

QBID:1112166

The algorithm used to simulate the direct set of randomly firing biochemical reactions to trace the trajectory of each biomolecule involved in the reactions is?

1. Euler method
2. Gillespie algorithm
3. Jacobi-eigen value algorithm
4. Metropolis algorithm

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17461]

2[Option ID=17462]

3[Option ID=17463]

4[Option ID=17464]

Sl. No.27

QBID:1112167

Which one of the following statements about docking algorithms is FALSE?

- 1) Docking methods will typically provide a unique solution
- 2) Docking methods incorporating ligand flexibility are common
- 3) Exploring search space is a major computational overhead
- 4) Crystal structures of complexes are still the best resource for validation of docking simulation methods

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17465]

2[Option ID=17466]

3[Option ID=17467]

4[Option ID=17468]

Sl. No.28

QBID:1112168

The most difficult type of RNA structure to predict is?

- 1) Hairpin
- 2) Pseudoknot
- 3) Bulges/loop
- 4) Stems

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17469]

2[Option ID=17470]

3[Option ID=17471]

4[Option ID=17472]

Sl. No.29

QBID:1112169

In a H-P model of protein folding, the native structure is identified which,

- 1) Shows the minimum energy.
- 2) Allows formation of the maximum number of H-H contacts.
- 3) Is opted by the maximum number of sequences as the lowest energy structure.
- 4) Features an equal proportion of hydrophobic and polar residues.

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17473]

2[Option ID=17474]

3[Option ID=17475]

4[Option ID=17476]

Sl. No.30

QBID:1112170

A homology model was built and its energy was calculated using some force-field. The total energy was -1830 Kcal/mole. It was declared by the researcher that it is the best model. Do you agree?

- 1) No. Empirical energy values are not at all the indicators of quality of structures without the knowledge of the potential energy surface of that molecule
- 2) Yes. Because the energy value is so low
- 3) Yes. This energy was found less than the energy of the template
- 4) No, the energy values should be less than -5000 Kcal/mole

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17477]

2[Option ID=17478]

3[Option ID=17479]

4[Option ID=17480]

Sl. No.31

QBID:1112171

Triple helical structure of DNA involves

- 1) binding of a 3rd strand in the major groove of a double helical structure adopted by poly-purine stretches.
- 2) binding of a 3rd strand in the minor groove of a double helical structure adopted by poly purine stretches.
- 3) binding of a 3rd strand in the major groove of a double helical structure adopted by alternating purine-pyrimidine stretches.
- 4) binding of a 3rd strand in the minor groove of a double helical structure adopted by alternating purine-pyrimidine stretches.

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17481]

2[Option ID=17482]

3[Option ID=17483]

4[Option ID=17484]

Sl. No.32

QBID:1112172

The information content (S) of three oligopeptides (a) AAAACCCC; (b) AACCAACC; and (c) AACACACC have the following trend

- 1) $S(c) > S(b) > S(a)$
- 2) $S(a) > S(b) > S(c)$
- 3) $S(b) = S(a) > S(c)$
- 4) $S(a) = S(c) = S(b)$

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17485]

2[Option ID=17486]

3[Option ID=17487]

4[Option ID=17488]

Sl. No.33

QBID:1112173

What did Watson and Crick used to solve the structure of

- 1) X-ray diffraction.
- 2) Optical gratings
- 3) Cryo EM spectroscopy
- 4) NMR spectroscopy

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17489]

2[Option ID=17490]

3[Option ID=17491]

4[Option ID=17492]

Sl. No.34

QBID:1112174

How do the sugars of RNA and DNA differ?

- 1) RNA has a six-carbon sugar; DNA has a five-carbon sugar.
- 2) The sugar of RNA has a hydroxyl group that is not found in the sugar of DNA.
- 3) RNA contains uracil; DNA contains thymine.
- 4) DNA's sugar has a phosphorus atom; RNA's sugar does not

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17493]

2[Option ID=17494]

3[Option ID=17495]

4[Option ID=17496]

Sl. No.35

QBID:1112175

Which of the following option is/are true for the following relations for the percentage of bases in double-stranded DNA?

A. $C + T = A + G$

B. $\frac{C}{A} = \frac{T}{G}$

- 1) A only
- 2) B only
- 3) Both A and B
- 4) Relation is invalid

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17497]

2[Option ID=17498]

3[Option ID=17499]

4[Option ID=17500]

Sl. No.36

QBID:1112176

A diploid plant cell contains 2 billion base pairs of DNA. The number of molecules of each type of histone protein associated with the genomic DNA are

1) 10^5 2) 10^6 3) 10^7 4) 10^8

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17501]

2[Option ID=17502]

3[Option ID=17503]

4[Option ID=17504]

Sl. No.37

QBID:1112177

Identify the odd one out from the Linux commands.

1) vi

2) vim

3) nano

4) cd

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17505]

2[Option ID=17506]

3[Option ID=17507]

4[Option ID=17508]

Sl. No.38

QBID:1112178

Which Linux terminal command can be used to delete a directory?

- 1) `rm -f`
- 2) `rm -i`
- 3) `mv -r`
- 4) `rm -rf`

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17509]

2[Option ID=17510]

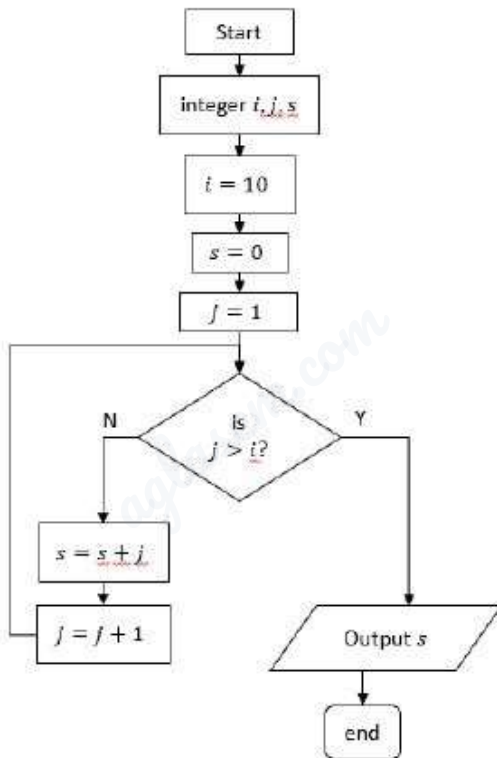
3[Option ID=17511]

4[Option ID=17512]

Sl. No.39

QBID:1112179

Determine the final output for the following flowchart



- 1) 45
- 2) 55
- 3) 65
- 4) 75

(1) 1

(2) 2

(3) 3

(4) 4

1[Option ID=17513]

2[Option ID=17514]

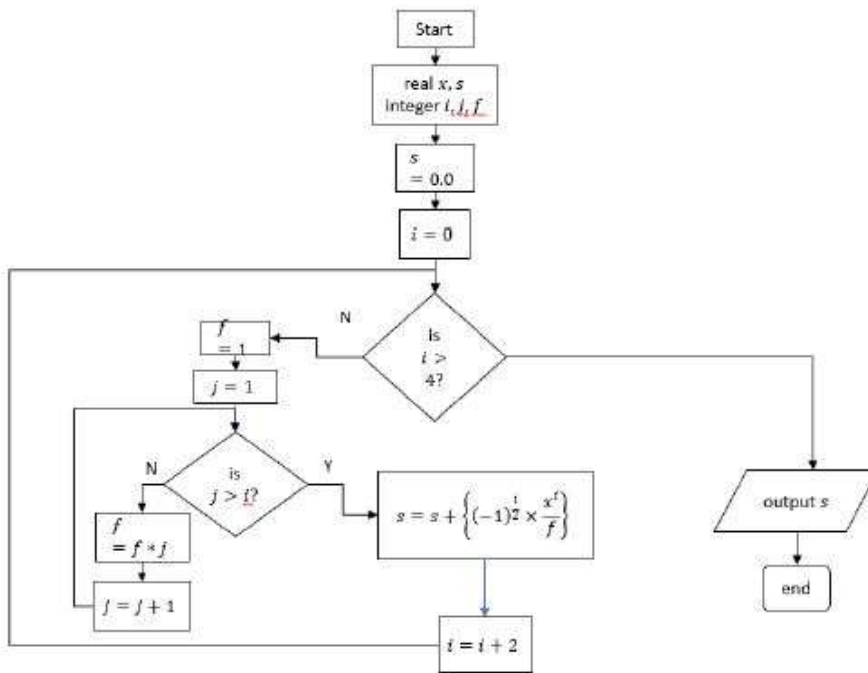
3[Option ID=17515]

4[Option ID=17516]

Sl. No.40

QBID:1112180

Find the correct output of the following flowchart.



- 1) .0998
- 2) 0.9950
- 3) 0.9900
- 4) 0.0900

- (1) 1
- (2) 2
- (3) 3
- (4) 4

1[Option ID=17517]
 2[Option ID=17518]
 3[Option ID=17519]
 4[Option ID=17520]