

ENTRANCE EXAMINATIONS – 2019

(Ph.D. Admissions - January 2020 Session)

Ph.D. Systems and Computational Biology

HALL TICKET NUMBER

Maximum Marks: 70

Time: 2 hours

INSTRUCTIONS: Please read the instructions carefully before answering the questions

1. Answers are to be marked on the OMR answer sheet.
2. Hand over the OMR answer sheet at the end of the examination to the invigilator.
3. The question paper contains 70 questions of multiple choices, printed in 13 pages (last three pages to be used for rough work), including this page.
4. OMR answer sheet provided separately.
5. All questions carry one mark each.
6. In case the candidates have equal marks, preference will be given towards the candidate who has obtained higher marks in Part-A.
7. **There is negative marking for wrong answer. Each wrong answer carries -0.33 mark.**
8. Non-programmable scientific calculators are permitted.
9. Cell,Mobile Phones are strictly prohibited in the examination hall.

PART-A

1. Consider two normal distributions with same mean. The alternate hypothesis that a random variable x doesn't belong to the population is rejected in the case of the first (distribution) but accepted in the other at a confidence level of 99%. As a statistician, your correct assessment would be that
 - A. The two normal distributions cannot be compared in such a way
 - B. The two distributions have very different variance
 - C. The median values should have been used as a measure of central tendency
 - D. The data have inherent noise

2. In next generation sequencing, the base calling step also provides quality scores (Q), also called *phred* scores, computed by formula $Q = -10 * \log_{10} p$, where p is the p-value, probability of calling the given base by chance. If the phred score is 40, the chances of a wrong base call would be
 - A. 1 out of 10
 - B. 1 out of 100
 - C. 1 out of 1000
 - D. 1 out of 10000

3. In a differential gene expression analysis, a researcher decides to select those genes with \log_2 fold change (w.r.t. control) ≥ 2 in treated group. A gene qualifying this criterion showed expression of 200 units in control group, its expression in treated group would be _____ units
 - A. 300
 - B. 400
 - C. 600
 - D. 800

4. A plant researcher working on Sorghum decides to re-sequence the genome of one of the varieties of interest. The size of Sorghum genome is approximately 740 Mb (million base pairs). After sequencing, the researcher receives sequence data of approximately 20 Gb (Giga bases). What is the range of genome coverage?
 - A. 50 and above
 - B. 35-40
 - C. 25-30
 - D. 15-20

5. A researcher interested in finding co-expressed genes typically uses clustering of gene expression profiles. One of the distance measures used for this purpose is Euclidean distance. Considering profile (20,1000,0) for the first and (80,100,50) for the second gene, the Euclidean distance between these two genes is _____
 - A. 903
 - B. 1010
 - C. 800

D. 790

6. If you are asked to use a computer program to construct all possible codons from a set of nucleotides (A, U, C, G), how many *for* loops would you be using optimally?
- A. 4
 - B. 3
 - C. 2
 - D. 64
7. Shared infrastructure; dynamic provisioning; network-based access and managed metering are the characteristics of _____
- A. Cloud computing
 - B. High-performance computing
 - C. Grid computing
 - D. Quantum Computing
8. A researcher wants to precipitate DNA from a sample using vortex. What kind of force acts on DNA during its precipitation?
- A. Gravitational force
 - B. Electrostatic force
 - C. Centrifugal force
 - D. Centripetal force
9. An outlier is a data point whose value _____
- A. Lies between the first and the second quartile values
 - B. Lies between the first and the third quartile values
 - C. Lies between the second and the third quartile values
 - D. None of the above
10. Let $y = x^2 - 2x$, then the number of minima and the value of global minimum of this function are _____ and at $x = _____$ respectively.
- A. 1 and 0
 - B. 1 and 1
 - C. 1 and -1
 - D. 2 and 1
11. Efficacy of a blood pressure (BP) lowering drug was tested on a number of BP patients as well as on healthy individuals and it was found that the drug administered was efficient in lowering the BP. What could be the most appropriate null hypothesis (H_0) used and the associated p-value, to reject H_0 ?
- A. drug has no effect on lowering BP ($p << 0.05$)
 - B. drug lowers BP in controls than in patients ($p >> 0.05$)
 - C. drug lowers BP in patients than in controls ($p << 0.05$)

- D. drug has no effect on lowering BP ($p >> 0.05$)
12. A coin is tossed and a die is rolled simultaneously. The probability of getting "Head" and "6" together is equal to _____
- $1/2 - 1/6$
 - $1/2 \times 1/6$
 - $1/2 + 1/6$
 - $1/2 \div 1/6$
13. A painted wall upon flashing a monochromatic light appeared black. The situation in which this **cannot** happen is when:
- The light flashed is of blue color and the wall is painted red
 - The light flashed is of red color and the wall is painted green
 - The light flashed is of orange color and the wall is painted yellow
 - The light flashed is of blue color and the wall is also painted blue
14. If R is a set such that $R = \{(x,y) \in N \times N : y = x^2 \wedge x < 5\}$. Then cardinality of R (i.e. $|R|$) is _____
- 10
 - 5
 - 4
 - 8
15. In a circle of radius of 2 units, the angle subtended by an arc of length 8 units is equal to _____ radians
- 4
 - 1
 - 8
 - 2
16. Ten years ago, Sachin was half of Meena's age. If the ratio of their present ages is 3:4, what will be the total of their present ages?
- 45
 - 40
 - 35
 - 30
17. A man rows to a place, which is at a distance of 48 km and returns. The entire journey takes 14 hours. He finds that he could cover a distance of 4 km with the stream and 3 km against the stream per hour. The flow rate of the stream is _____
- 1 km/hr
 - 1.5 km/hr
 - 2.0 km/hr

D. 2.5 km/hr

18. The ratio between the speeds of trains P and Q is 4:5. If the train Q runs 100 km in 2 hours, What is the speed of the train P ?

- A. 45 km/hr
- B. 40 km/hr
- C. 35 km/hr
- D. 25 km/hr

19. $\log_7 0 = ?$

- A. 0
- B. 1
- C. 7
- D. Cannot be defined

20. A Head of an institution decided to form a committee to look into a matter, which is gender sensitive. The head made a list of 7 men and 3 women who would be part of that committee. How many ways can a committee of 5 men and 2 women be made out of his list?

- A. 1
- B. 126
- C. 63
- D. 64

21. From a deck of 52 cards, a card is drawn. The probability of the drawn card to be "Queen" is

- A. $1/2$
- B. $1/13$
- C. $1/6$
- D. $1/3$

22. Type I error is _____

- A. A null hypothesis is not rejected when it is actually false
- B. A null hypothesis is rejected when it is actually true
- C. An alternate hypothesis is rejected when it is actually true
- D. None of the above

23. How many recombinant clones of 20kb size need to be screened to cover 99% of the human genome?

- A. 6.9×10^5
- B. 6.9×10^4
- C. 6.9×10^6
- D. 6.9×10^8

24. Gene X of size 2.5kb has three exons and two introns, all of same size. What is the molecular weight of the protein after translation (assume that there is only constitutive splicing and no intron retention)?

- A. 91 kDa
- B. 84 kDa
- C. 55 kDa
- D. 66 kDa

25. If you dissolve 0.02 moles of acetic acid (CH_3COOH ; $\text{pK}_a=4.8$) in water to give 1 liter of solution. What is the pH of the solution?

- A. 3.25
- B. 4.55
- C. 2.85
- D. 3.45

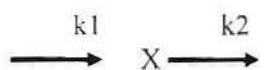
26. In order to prepare 400 ml of a 0.24M NaCl solution, which of the following is correct?

- A. 3.25 g of NaCl dissolved in 400ml of water
- B. 3.5 g of NaCl dissolved in 400ml of water
- C. 5.61g of NaCl dissolved in 400ml of water
- D. 11.23 g of NaCl dissolved in 400ml of water

27. In 2019 Nobel prize in medicine is given to William G. Kaelin, Gregg L. Semenza and Peter J. Ratcliffe for their pioneer work on?

- A. Hypoxia and oxygen sensing
- B. Induced pluripotent stem cells (iPS)
- C. Immunotherapy
- D. Circadian Rhythm

28. Consider the following reaction



Which of the following ODEs represents the rate of change of concentration of X (dX/dt)

- A. $dX/dt = k_1 * X - k_2 * X$
- B. $dX/dt = -k_1 * X + k_2$
- C. $dX/dt = k_1 * X - k_2$
- D. $dX/dt = k_1 - k_2 * X$

29. Which of the following is used to solve an ordinary differential equation numerically

- A. Runge-Kutta Method
- B. Finite Element Method
- C. Simpson's rule
- D. Trapezoidal rule

30. Which one of the following equations is considered as an ordinary differential equation?

- A. $dy/dx + 5yx + 10 = 0$
- B. $d^2y/dx^2 + 5yx + 10 = 0$
- C. $d^2y/dx^2 + 5x + 10 = 0$
- D. $\partial y/\partial x + 2 \partial y/\partial t + 5xy^2 t + 10 = 0$

31. Consider two variables x & y with the variances 4 and 9 respectively. If the correlation between the two variables is 0.84, then their covariance is equals to

- A. 4.20
- B. 5.04
- C. 0.14
- D. 30.04

32. Find the product of $(7.23 \times 7.23 \times 7.23 + 2.77 \times 2.77 \times 2.77) / (7.23 \times 7.23 - 7.23 \times 2.77 + 2.77 \times 2.77)$?

- A. 20.02
- B. 11
- C. 10
- D. 4.46

33. The distribution of number of mutations in a DNA sequence, after every replication can best be described by _____ distribution

- A. Poisson
- B. Gaussian
- C. Gumbel
- D. Hyper-geometric

34. $\lim_{x \rightarrow 3} \frac{x^2 - 9}{x - 3}$?

- A. 0
- B. 6
- C. does not exist
- D. 9

35. A straight line is such that y changes 5 times with change in x by 1. The value of y at $x=0$ is 2. The equation for straight line would be

- A. $5y = x + 2$
- B. $y = 5x + 2$
- C. $5y + 2 = x$
- D. $5y + x = 2$

PART-B

36. Which PCR is used to amplify the DNA of unknown sequences flanking a known DNA sequence

- A. Touch Down PCR
- B. Nested PCR
- C. inverse PCR
- D. RT-PCR

37. Which enzyme is used to digest template DNA during site directed mutagenesis after completion of PCR?

- A. EcoR1
- B. Sal1
- C. Bc1
- D. Dpn1

38. In order to study the role of glycosylation in protein sorting which of the following chemical is used?

- A. Tunicamycin
- B. Actinomycin D
- C. α -Amanitin
- D. Rapamycin

39. Which of the following vector is used in order to insert more than 1000 kb DNA insert

- A. Cosmid
- B. Phagemid
- C. YAC (Yeast Artificial Chromosome)
- D. BAC (Bacterial Artificial Chromosome)

40. Which of the following method is used to identify strength of a promoter?

- A. CHIP (Chromatin Immunoprecipitation Assay)
- B. EMSA (Electrophoretic Mobility Shift Assay)
- C. Foot printing
- D. Luciferase Assay

41. In immune check-point inhibition, which signaling axis is targeted for immune therapy?

- A. JAK/STAT signaling axis
- B. PD-1 and PD-L1 axis
- C. MHC II and TCR axis
- D. None of the above

42. The distinguishing genomic/genetic feature in plants as compared to animals is ____
- plant genomes are typically larger in size
 - plant genomes have more number of repeats
 - plant genomes show variation in terms of ploidy
 - plant genomes have relatively more sequence variation than that in case of animals
43. Low complexity regions in a genome comprised of:
- non-coding sequences
 - sequence repeats
 - genes without introns
 - genes with splice sites
44. Next-generation sequencing is characterized by ____
- massively parallel sequencing chemistry
 - digital sequencing
 - improved Sanger sequencing chemistry
 - improved Edman degradation chemistry
45. The term **contig** refers to ____
- gaps in a draft assembly
 - a fragment of DNA during library preparation
 - a consensus sequence of overlapping reads
 - unique regions in a chromosome
46. Which statement is *incorrect* about transcriptome sequencing? It is used ____
- for finding differentially expressed genes between pair of conditions/treatment
 - to create gene expression atlas
 - to obtain sequences of expressed regions of genome
 - to create mRNA looping
47. Which statement is *correct* about haploid genome size ____
- the size increases linearly with complexity of organism
 - the size remains static
 - the size is not related to the complexity of organism
 - the size increases exponentially with complexity of organism
48. For prediction of microRNA in a genome, which factor is *NOT* important?
- the miRNA precursor sequence can be of any length
 - the transcribed sequence can fold into energetically stable hairpin-loop
 - conservation across related species
 - presence of inverted repeats

49. RNA-sequencing and microarray are the two most common approaches for transcriptome profiling. Advantage of RNA sequencing technology over microarray is _____

- A. former can detect spliced variants while latter cannot
- B. former does not require much data analysis
- C. former can detect lowly expressed genes
- D. former is highly specific in transcript profiling while latter has very poor specificity

50. The ϕ_i and ψ_i angles that describe conformation of an amino acid residue "i" in a polypeptide chain are described by _____ respectively

- A. $C_{i-1}-N_i-C\alpha_i-C_i$ and $N_i-C\alpha_i-C_i-N_{i+1}$
- B. $C_{i-1}-N_i-C\alpha_i-C_i$ and $C\alpha_i-C_i-N_{i+1}-C\alpha_{i+1}$
- C. $C\alpha_i-C_i-N_{i+1}-C\alpha_{i+1}$ and $N_i-C\alpha_i-C_i-N_{i+1}$
- D. $N_i-C\alpha_i-C_i-N_{i+1}$ and $C_i-N_{i+1}-C\alpha_{i+1}-C_{i+1}$

51. The ϕ and ψ angles of a Leu residue in a polypeptide chain are $+20^\circ$ and -40° respectively. This corresponds to _____ conformation for that amino acid residue.

- A. Helical
- B. Allowed
- C. Disallowed
- D. Extended

52. "Disallowed" regions in the Ramachandran Plot have been so defined because _____

- A. At least one pair of bonded atoms in the polypeptide chain is having a covalent bond length far less than the characteristic length
- B. At least one pair of non-bonded atoms in the polypeptide chain has short contact
- C. The backbone does not form any secondary structure
- D. There are no intra-molecular hydrogen bonds

53. In a BLAST search _____ regions are usually masked

- A. Low complexity
- B. High complexity
- C. Highly conserved
- D. Highly diverged

54. Needleman-Wunsch method for pair-wise alignments is an implementation of _____ method

- A. BLAST
- B. Neighbour joining
- C. Genetic algorithm
- D. Dynamic programming

55. Among the following nodes, _____ is having high Eigen vector centrality value

- A. Node A (because it is connected to a number of nodes which are themselves highly connected in the network)
- B. Node B (because it is connected to a large number of different nodes in the network)
- C. Node C (because all its neighbors are highly interconnected)
- D. Node D (because it is associated with high betweenness value)

56. "Centrality-Lethality" rule as discerned from an analysis of protein-protein interaction networks, implies that _____

- A. proteins with high centrality values are associated with lethal mutations
- B. proteins having high centrality values are mostly the products of essential genes
- C. proteins having high centrality values usually are the products of lethal genes
- D. proteins having high centrality values usually undergo quick degradation

57. Michaelis-Menten Kinetics for the enzymatic reaction Substrate (S)+Enzyme (E)→ Product (P)+ Enzyme (E) is valid when _____

- A. $[E]=[S]$
- B. $[E] \gg [S]$
- C. $[E] \ll [S]$
- D. all $[E]$ and $[S]$

58. The rate of any chemical reaction $A+2B \rightarrow C$ with rate constant k (in some unit) can be given by

- A. $\frac{dA}{dt} = k [A][B]$
- B. $\frac{dA}{dt} = -k [A][B]$
- C. $\frac{dA}{dt} = k [A][B]^2$
- D. $\frac{dA}{dt} = -k [A][B]^2$

59. In order to find steady state concentration of A, B and C for any biochemical reaction $A + B \rightleftharpoons C$ we consider

- A. $dA/dt > 0, dB/dt > 0, dC/dt > 0$
- B. $dA/dt < 0, dB/dt > 0, dC/dt > 0$
- C. $dA/dt = 0, dB/dt = 0, dC/dt = 0$
- D. $dA/dt > 0, dB/dt = 0, dC/dt = 0$

60. The following behavior is observed in a synthetic genetic circuit called 'Toggle switch'
- A. Monostability
 - B. Bistability
 - C. Oscillations
 - D. Tristability
61. The Biobricks Foundation is for the _____
- A. collection of bio-models
 - B. collection of standard synthetic genetic modules
 - C. collection of biological network motifs
 - D. collection of biochemical reactions rate constants
62. The synthetic regulatory network referred to as Repressilator consists of three gene-promoter pairs and shows oscillations. The gene circuit is _____
- A. Gene 1 → Gene 2 → Gene 3 → Gene 1
 - B. Gene 1 → Gene 2 → Gene 3 → Gene 1
 - C. Gene 1 → Gene 2 → Gene 3 → Gene 1
 - D. Gene 1 → Gene 2 → Gene 3 → Gene 1
63. Which of this is both primary and secondary database
- A. ENA
 - B. Genbank
 - C. DDBJ
 - D. Uniprot
64. E-value **doesn't** directly depend on
- A. Size of the database
 - B. Length of the query
 - C. Low-complexity region
 - D. Scoring matrix
65. Which are global and local alignment tools, respectively?
- A. SSERACH and LALIGN
 - B. BLAST and WATER
 - C. NEEDLE and BLAST
 - D. tBLASTx and TFASTX
66. The method of scoring multiple-sequence alignment is _____
- A. Substitution scoring
 - B. Sum of pairs
 - C. Average of pairs scores
 - D. Shannon-entropy

67. What is 'hidden' in Hidden Markov Model?

- A. Paths
- B. States
- C. Transition probability
- D. Emission probability

68. What is the formula for affine gap penalty (W) where, γ is the gap-opening penalty, δ is the gap extension penalty, and k is the length of the gap.

- A. $W = \gamma + \delta \times (k - 1)$
- B. $W = \gamma k$
- C. $W = \gamma - \delta \times (k - 1)$
- D. $W = \gamma + \delta k$

69. Low learning rate in training neural network leads to __

- A. the rapid convergence of algorithm to the optimal solution.
- B. no effect in model training.
- C. the termination of the algorithm after few iterations.
- D. the slow convergence of algorithm to the optimal solution

70. In the given MSA, which sites are informative for constructing tree using maximum parsimony method?

	1	2	3	4	5	6	7	8	9	10
SeqA	A	G	G	T	A	A	C	T	G	G
SeqB	A	C	G	T	T	A	T	T	A	A
SeqC	A	T	A	T	T	G	T	C	T	A
SeqD	A	A	T	T	T	G	T	C	G	G

- A. 4,6,9
- B. 2,9,10
- C. 6,8,10
- D. 2,3,6