


Name-Surname : 

No 


Email : 

=====


Q01. An interdisciplinary field that develops methods and software tools for understanding biological data and combines computer science, statistics, mathematics, and engineering:

- a. Genetics
-  b. Bioinformatics
- c. Proteomics
- d. Genomics


Q02. The term used to refer something performed on computer or computer simulation:

- a. Dry lab
- b. Wet lab
- c. Web lab
-  d. In silico


Q03. Bioinformatics does not use techniques and concepts from ...

- a. chemistry
- b. physics
- c. statics
-  d. linguistics

Q04. Application of bioinformatics include

- a. data storage and management
- b. drug designing
- c. understand relationship between organisms
-  d. all of the above

Q05. Meta data is a set of data that describes and gives information about other data. Meta data for finding or understanding a resource is known as

-  a. Descriptive metadata
- b. Administrative metadata
- c. Structural metadata
- d. Markup languages

Q06. Molecular data (DNA or protein sequences) are

-  a. discrete quantitative data

Introduction to Bioinformatics – Final Exam

- b. continuous quantitative data
 - c. nominal categorical data
 - d. ordinal categorical data
-

Q07. Human genome contains about ...

- a. 2 billion base pairs
 - ☒ b. 3 billion base pairs
 - c. 4 billion base pairs
 - d. 5 billion base pairs
-

Q08. Which of the following is a nucleotide sequence data base?

- ☒ a. EMBL
 - b. SWISS PROT
 - c. PROSITE
 - d. TREMBL
-

Q09. Phylogenetic relationship can be shown by

- a. Data search tool
 - b. Gene Bank
 - c. Data retrieving tool
 - ☒ d. Dendrogram
-

Q10. Step-by-step procedure for solving problems in computer science is called

- a. flowchart
 - b. sequential design
 - ☒ c. algorithm
 - d. procedure
-

Q11. A **scalar** variable in Perl can hold ...

Introduction to Bioinformatics – Final Exam

- a. a number
 - b. a word
 - c. a sentence
 - ☒ d. all of the above
-

Q12. Which of the following Perl statements is equivalent to the Perl statement given as

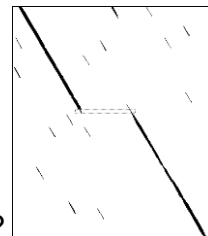
`if ($string !~ m/xyz/) ...`

- ☒ a. `if (!$string =~ m/xyz /) ...`
 - b. `if ($string =~ m/!xyz /) ...`
 - c. `if ($string =~ m!/xyz /) ...`
 - d. `if ($!string =~ m/xyz /) ...`
-

Q13. A data base is a collection of one or more related ...

- a. rows of data
 - b. columns of data
 - ☒ c. files of data
 - d. tables of data
-

Q14. The given dotplot indicates which of the following conditions?



- ☒ a. An indel (insertion or deletion)
 - b. A tandem duplication
 - c. An inversion
 - d. Joining sequences
-

Q15. Two sequences are given as ACGT and TCGG. Which of the following dot plots is the correct one when a sliding window size of 3 and a similarity cut-off of 2 nucleotides are used.

- a.

Introduction to Bioinformatics – Final Exam

-	A	C	G	T
T				●
C		●		
G			●	
G			●	

b.

-	A	C	G	T
T				
C		●		
G			●	
G				



-	A	C	G	T
T				●
C		●		
G			●	
G				

d.

-	A	C	G	T
T				
C		●		
G			●	
G			●	

Q16. A research group has performed a number of micro-array experiments and they would like to perform a cluster analysis on the data. The experiment consists of four arrays. On each array, there are two genes. Results of experiments are given in the following table.

Introduction to Bioinformatics – Final Exam

	Gene 1 (x)	Gene 2 (y)
Array1 (A1)	5	-2
Array2 (A2)	2	3
Array3 (A3)	-4	2
Array4 (A4)	-1	-4

What will be the **Euclidean squared distance** between Array 1 and Array 4.

- a. 20
- b. 30
- ☒ c. 40
- d. 72

Q17. You want to calculate compositional complexity of your name. Which formula you use to calculate compositional complexity?

a.

$$K = \frac{1}{L} \log_4 \left(\frac{L!}{\prod_{i=1}^4 n_i!} \right)$$

b.

$$K = \frac{1}{L} \log_{20} \left(\frac{L!}{\prod_{i=1}^{20} n_i!} \right)$$

c.

$$K = \frac{1}{L} \log_{26} \left(\frac{L!}{\prod_{i=1}^{26} n_i!} \right)$$

☒

$$K = \frac{1}{L} \log_{29} \left(\frac{L!}{\prod_{i=1}^{29} n_i!} \right)$$

Q18. Considering the following global dynamic programming matrix, which of the following statements about the alignment significance of sequences **F M D T** and **F K D T** are correct?

Introduction to Bioinformatics – Final Exam

	-	F	M	D	T
-	0	-2	-4	-6	-8
F	-2	9	7	5	3
K	-4	7	9	7	5
D	-6	5	7	13	11
T	-8	3	11	11	16

- Alignment score > Significance cut-off. This is a significant alignment.
- Alignment score < Significance cut-off. This is not a significant alignment.
- Alignment score = Significance cut-off. This is neither significant nor insignificant alignment.
- All are incorrect.

Q19. Considering the following global dynamic programming matrix, determine the match score, the mismatch score, and the gap penalty.

	-	G	A	C
-	0	-2	-4	-6
G	-2	1	-1	-3
G	-4	-1	0	-2
G	-6	-3	-2	-1
C	-8	-5	-4	-1

- match score = +1 mismatch score = 0 gap penalty = -2
- match score = +1 mismatch score = -1 gap penalty = -2
- match score = +2 mismatch score = -1 gap penalty = -2
- match score = +1 mismatch score = -1 gap penalty = -1


Q20. Considering the following global dynamic programming matrix, which of the following pairwise global alignments is the optimum alignment?

	-	G	A	C
-	0	-2	-4	-6
G	-2	1	-1	-3
G	-4	-1	0	-2
G	-6	-3	-2	-1
C	-8	-5	-4	-1


- a. - G A C
 G G G C
- b. G A - C
 G G G C
- c. G - A C
 G G G C

 All

Q21. The **sum of pairs** method scores all possible combinations of pairs of residues in a column of a multiple sequence alignment. Assuming that we have 6 amino acid sequences in a data set, how many possible unique pairs will be formed?

- a. 6
- b. 10
-  15
- d. 21

Q22. Graphical representation of quantifiable data, usually by means of well-known chart, graph or map types is known as ...

- a. information visualization
-  data visualization
- c. computer visualization
- d. natural visualization

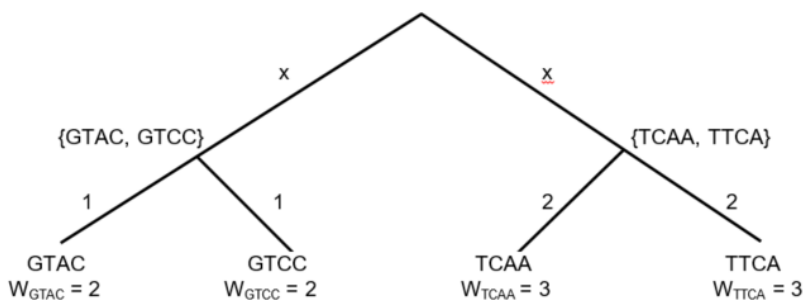
Q23. ----- is used to convey overall impression of relationship between two variables.

- ☒ a. Scatter plot
 - b. Line graph
 - c. Star plot
 - d. Bar graph
-

Q24. A DNA sequence can be stored in the memory as a character array. However, when the memory size limitation is considered, saving a DNA sequence as character array is not desirable. Instead, a simple binary coding can be used for compressing the DNA sequence. In this case, how much memory saving can be achieved?

- a. 25%
 - b. 35%
 - c. 50%
 - ☒ d. 75%
-

Q25. In the following phylogenetic tree, the distance between cluster “{GTAC, GTCC}” and cluster “{TCAA, TTCA}” can be calculated as **2x**. **x** is the distance between the root of the tree and either the cluster “{GTAC, GTCC}” or cluster “{TCAA, TTCA}”. Determine the value of **x** for the given node weights?



- a. $x = 1.5$
- b. $x = 2$
- ☒ c. $x = 2.5$
- d. $x = 3$