**Name-Surname :**

**No :**

**Email :** ============================================================================

Q1. Find the word or phrase from the list given in the box that best matches the description in the following statements. Use the numbers to the left of words in the answer. Each answer should be used only once.

01. genetic code

02. infographic

03. knowledge

04. genomics

05. data

06. molecular biology

07. signal

08. bandwidth

09. functional genomics

10. bioinformatics

An interdisciplinary field that develops methods and software tools for understanding biological data.

10…

The study of the genome, which is the complete set of the genetic material or DNA present in an organism.

4…

Field of molecular biology, which attempts to make use of the vast wealth of data given by genomic and transcriptomic projects to describe gene (and protein) functions and interactions.

…9

A combination of information, experience and insight that may benefit the individual or the organization.

3…

An information variable represented by physical quantity.

7…

Understanding the interactions between the various systems of a cell, including the interactions between DNA, RNA and protein biosynthesis and learning how these interactions are regulated.

6…

Graphical representation that combines one or more data visualizations with other non-data elements.

2…

Unprocessed facts and figures without any added interpretation or analysis.

…5

The amount of information that can be communicated per unit time.

8…

The set of rules by which information encoded within genetic material (DNA or mRNA sequences) is translated into proteins by living cells.

1…

--------------------------------------------------------------------------------------------------------------------------------------

Q2. Find the computational complexity of your name+surname.

My name is burak my surname is çulha. So length of my name+surname becomes burakçulha. |burakçulha| = 10. 10^1 = 10

Complexity of my name+surname = 10.

--------------------------------------------------------------------------------------------------------------------------------------

Q3. Assuming that we have 4 amino acid sequences in a data set;

1. How many possible **rooted trees** can be constructed?

1 3 15 105

..

..

.+.

..

1. In **sum of pairs** method, how many different pairwise alignments can be considered?

4 6 8 16

..

.+.

..

..

1. Suppose the length of each sequence is ***n* residues**, what will be the number of comparisons in Dynamic Programming?

*n*4 4*n* *n*8 4*n*4

+..

..

..

..

1. What will be the number of unique pairs in Dynamic programming space?

6 8 12 24

.+.

..

..

..

--------------------------------------------------------------------------------------------------------------------------------------

Q4. ACGTCAT and TCATGCA are two DNA sequences obtained from the same subject. Usually these pieces of sequences are assembled to form a single DNA sequence.

1. Suggest a suitable Dynamic Programming algorithm to assemble these two sequences.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | A | C | G | T | C | A | T |  |
|  | 0 | -1 | -2 | -3 | -4 | -5 | -6 | -7 |  |
| T | -1 | -1 | -2 | -3 | -2 | -3 | -4 | -5 |  |
| C | -2 | -2 | 0 | -1 | -2 | -1 | -2 | -3 |  |
| A | -3 | -1 | -1 | -1 | -2 | -2 | 0 | -1 |  |
| T | -4 | -2 | -2 | -2 | 0 | -1 | -1 | 1 |  |
| G | -5 | -3 | -3 | -1 | -1 | -1 | -2 | 0 |  |
| C | -6 | -4 | -2 | -2 | -2 | 0 | -1 | -1 |  |
| A | -7 | -5 | -3 | -3 | -3 | -1 | 1 | 0 |  |

Needleman-wunsch global alignment

1. Assuming a match score of 1, mismatch score of -1 and gap score of -1, fill in the following Dynamic Programming matrix according to your suggested algorithm.
2. Determine the optimal alignment

T C A T G C A \_  
| \* | \* \* \*  
A C G T \_ C A T

--------------------------------------------------------------------------------------------------------------------------------------

Q5. Considering the following multiple alignments of four sequences:

**A A C G C G T T G G G C G A T G G C A A C**

**B A C G C G T T G G G C G A C G G T A A T**

**C A C G C A T T G A A T G A T G A T A A T**

**D A C A C A T T G A G T G A T A A T A A T**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Seq.** | *A* | *B* | *C* | *D* |
| *A* | 0 | 3 | 7 | 8 |
| *B* |  | 0 | 6 | 7 |
| *C* |  |  | 0 | 3 |
| *D* |  |  |  | 0 |

1. Complete the following distance matrix for sequences *A*, *B*, *C*, and *D*.
2. Complete (determine **a**, **b**, **c**, **d**, and **e** in) the following un-rooted tree.

C

D

A

B

e=4

c=1

d=2

a=2

b=1

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **-** | **G** | **A** | **T** | **T** | **A** | **C** | **A** | **T** | **A** |
| **-** | 0 | -6 | -8 | -10 | -12 | -14 | -16 | -18 | -20 | -22 |
| **G** | -6 | 4 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 |
| **C** | -8 | -2 | 2 | -4 | -6 | -8 | -4 | -10 | -12 | -14 |
| **C** | -10 | -4 | -4 | 0 | -6 | -8 | -4 | -6 | -12 | -14 |
| **A** | -12 | -6 | 0 | -6 | -2 | -2 | -8 | 0 | -6 | -8 |
| **A** | -14 | -8 | -2 | -2 | -8 | 2 | -4 | -4 | -2 | -2 |
| **G** | -16 | -10 | -8 | -4 | -4 | -4 | 0 | -6 | -6 | -4 |
| **T** | -18 | -12 | -10 | -4 | 0 | -6 | -6 | -2 | -2 | -8 |
| **A** | -20 | -14 | -8 | -10 | -6 | 4 | -2 | -2 | -4 | 2 |
| **G** | -22 | -16 | -14 | -10 | -8 | -2 | 2 | -4 | -4 | -4 |
| **G** | -24 | -18 | -16 | -14 | -10 | -4 | -4 | 0 | -6 | -6 |

--------------------------------------------------------------------------------------------------------------------------------------

Q6. Following dynamic programming matrix was calculated by using the Needleman-Wunsch algorithm.

To construct this matrix, a match score ***m***, a mismatch score ***n***, the gap-opening penalty ***g*** and the gap-extension penalty ***r*** were used.

Determine the values of ***m***, ***n***, ***g***, and ***r*** and explain how you found them.

***Reminder***: Affine gap penalty is used for matrix initialization and calculation of the cell values. Affine gap penalty in this matrix is given as ***wx*=** where ***wx*** : total gap penalty; ***g***: gap open penalty; ***r***: gap extend penalty; ***k***: gap length

m=4 ,n =-2, g=-4, r=-2. Gap opening penalty can be found by looking first row. Formula = g+rk . for S0,1 = -4+ 1\*-2 = -6. We can see that every match increases by 4 so m=4. r=-2 is found when calculating S0,1 = -4+ 1\*-2 = -6. Mismatch score is n=-2

--------------------------------------------------------------------------------------------------------------------------------------

Q7. Human uses five senses (visual, smell, taste, touch, hearing) to acquire information. Visualization is the most widely used of a general class of perception technologies. Therefore, data visualization techniques are an important part of Bioinformatics.

In this respect, what is **dual-coding theory**? Explain.

Dual-coding theory is a theory of cognition according to which humans process and represent verbal and non-verbal information in separate, related systems. For example, the brain uses a different kind of representation for the word "tree" than it does for the image of a tree .

--------------------------------------------------------------------------------------------------------------------------------------

Q8. Determine whether the following statements are correct or not by placing F(alse) or T(rue) in paranthesis.

a. For database searching the commonly used matrix is PAM. ( F )

b. Global alignment aligns the whole sequence, while local alignment finds the best

subsequence that aligns. ( T )

c. Sequences are homologous if they are derived from a common ancestral sequence. ( T )

d. Perl is a typeless language that does not force the programmer to distinguish between

the types of data stored in a variable. ( T )

e. Moving average filtering is useful in impulsive noise removal in image processing. ( T )

--------------------------------------------------------------------------------------------------------------------------------------

Q9. Which of the following concepts are not used in bioinformatics?

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ..  statistics | ..  physics | .+.  pedagogy | ..  linguistics | ..  mathematics | .+.  statics | ..  chemistry |

--------------------------------------------------------------------------------------------------------------------------------------

Q10. What does the following Perl code do? Explain

open IN, “dna “;

open OUT, ">dna\_new ";

while ($line = <IN>) { print OUT $line; }

This code reads from “dna “ file and copies each line

to “dna\_new”.