

Question 1: In

bioinformatics research, sequence alignment is a method of arranging DNA, RNA, or protein sequences.

The method of arranging columns to identify similar regions is one of the most basic tasks, and subsequent work on these region sequences...

Analysis of energy, structure, or evolutionary relationships. Methods for aligning nucleotide or amino acid bases have been developed, with

dot-matrix methods being a rapid and simple approach. Dot-matrix methods can be used to assess the similarity between two sequences.

Using a dot plot method, taking DNA sequences as an example, one sequence is arranged on the X-axis of the plot, and another sequence is arranged on the Y-axis.

One sequence is used to compare two sequences. When the bases of the two sequences match at the same positions in the diagram, the corresponding positions are plotted.

Set a point. The starting points of the two sequences on the two axes must be at the same point. If the sequence on the X-axis corresponds to the sequence on the Y-axis...

If the bases are the same, a point is plotted in this grid, as shown in the example below. Sequence 1 and Sequence CTTGCC are listed on the X-axis.

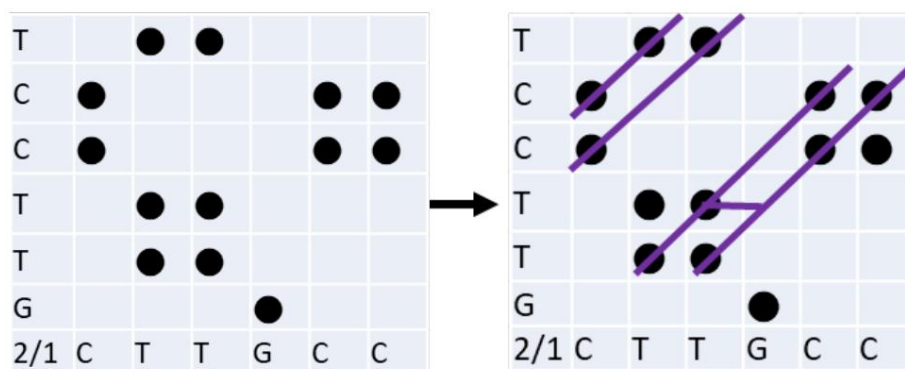
2. The GTTCCT sequence is plotted on the Y-axis, and a point is drawn at the corresponding base position on both the X and Y axes, forming the following structure.

fruit.

Once the points are drawn, connect neighboring points with 45-degree lines. The proximity of similar sequences will determine the diagonal and display.

The proximity of the connecting lines in a direct relationship is influenced by the differences between the two sequences, such as shifts.

The position is displayed by adding spaces when arranging the sequence.



The methods for permuting two sequences can be summarized into the following 5 possibilities:

1 CTTGCC- 1 ---CTTGCC 1----CTTGCC 1 CTTGCC 1 CTTGCC-
 II II I I II II
 2 GTT-CCT 2 GTTCCT--- 2 GTTCCT---- 2 GTTCCT 2 GTT-CCT

The optimal arrangement is determined using a calculation method, as shown in the following formula:

$$D = s + wg$$

's' represents the number of base substitutions, 'g' represents the total number of spaces, and 'w' is the space length weighted score. Therefore, the following four permutations...

The calculations for the formulas are as follows:

1 CTTGCC- 1 ---CTTGCC 1----CTTGCC 1 CTTGCC 1 CTTGCC-
 II II I I II II
 2 GTT-CCT 2 GTTCCT--- 2 GTTCCT---- 2 GTTCCT 2 GTT-CCT
 $D = 1 + 1 \times 2 = 2$ $D = 1 + 3 \times 6 = 19$ $D = 0 + 4 \times 8 = 32$ $D = 3 + 0 \times 0 = 3$ $D = 1 + 1 \times 2 = 3$

Based on the results above, the optimal permutation of these two sequences is as follows:

1 CTTGCC-
 II II

2 GTT-CCT because of calculation

D The value is 2, which is the minimum value.

Given two sequences:

Sequence A: ATTCGGGCA

and Sequence B: GAATCGGGC,

align and arrange these two sequences to answer the following questions:

Question 1:

Using the raster chart method, draw points in the following image: (5 points)

C
G
G
G
C
T
A
A
G

B/A ATTCGGGCA

Question

2: Using the raster diagram from Question 1, connect neighboring points with 45-degree lines to draw all possible sequence arrangements: (5 points)

C
G
G
G
C
T
A
A
G

B/A ATTCGGGCA

Question 3

Using question 2, draw up the possible permutations of the sequences, as described in the question above, and organize the possible permutations of the two sequences.

Write down five possibilities: (10 points)

Question 4

Using the $D = s + wg$ The calculation method involves using this formula to calculate the possible permutations of the two sequences derived from Problem 3.

algorithm, s represents the number of base substitutions, g represents the total number of spaces, and w is the weighted score based on space length, for each possible permutation...

Each example needs to clearly show the calculation formula and the calculation process: (10 points)

Question 5

Based on the results above, write down the optimal permutation of these two sequences and the reasons why. (4 points)