

NAME: The Chumaka Atina

MATRICO: 18/MH501/171

DEPARTMENT: Anatomy

COURSE: CSC 262

Assignment

1. List and discuss any three (3) criterion of database searching.
 - a. Explain briefly basic local Alignment Search Tool as used in database similarity searching.
2. Briefly discuss the differences between Dayhoff PAM Matrices and BLOSUM matrices.
 - a. Explain briefly Heuristic Database Searching
3. Define the following.
 - i) Sequence Homology, ii) Sequence Similarity, iii) Sequence Identity.
 - a. Give any 3 methods of Alignment Algorithm.
 - i) Discuss briefly Pairwise Sequence Alignment.
4. Differentiate Between Global Alignment and Local Alignment.
 - a. Distinguish between the following
 - Sequence Homology and Sequence Similarity
 - Sequence Similarity and Sequence Identity

ANSWER

1. Sensitivity
 - Selectivity
 - Speed

Sensitivity

This refers to the ability to find as many correct hit as possible. It is measured by the extent of inclusion of correctly identified sequence members of the same family.

- Selectivity

This refers to the ability to exclude incorrect hits.

- Speed

This refers to the time it takes to get results from database searches.

b. Basic Local Alignment Search Tool (BLAST) is an algorithm and program for comparing primary biological sequence information, such as amino-acid sequences of proteins or the nucleotides of DNA or RNA sequences. It finds regions with the highest level of similarity between the two sequences and align these regions without regard for the alignment of the rest of the sequence regions.

2 Dayhoff PAM matrices

PAM matrices are used to score alignments between closely related protein sequences.

- Based on global alignments
- Higher numbers in the PAM matrices naming denotes greater evolutionary distance

BLOSUM Matrices

Blosum matrices are used to score alignments between evolutionarily divergent protein sequences.

- Based on local alignments
- Higher numbers in the Blosum matrix naming denotes higher sequence similarity and smaller evolutionary distance.

b. Heuristic Database Searching is a technique designed for solving a problem more quickly when classic methods are too slow, or for finding an appropriate approximate solution when classic methods fail to find any exact solution. There are two major heuristic algorithms for performing database searches - BLAST and FASTA

3. Sequence Homology: - This is the biological homology between DNA, RNA or protein sequences, defined in terms of shared ancestry in the evolutionary history of life.

18/MH501/191

- Sequence Similarity :- This is a measure of an empirical relationship between sequences. Its common objective is establishing the likelihood for sequence homology.

- Sequence Identity :- This is the amount of characters which match exactly between two different sequences.

b. Dot matrix method

- Dynamic programming method

- Word method.

i) Pairwise Sequence Alignment :- This is the process of aligning two sequences and is the basis of database similarity searching and multiple sequence alignment. It is used to identify regions of similarity that may indicate functional, structural and evolutionary relationships between two biological sequences.

A. Global Alignment

In global alignment, two sequences to be aligned are assumed to be generally similar over their entire length.

- It is made to align the entire sequence (end to end alignment)

Local Alignment

In local alignment, does not assume that the two sequences in question have similarity over the entire length.

It finds local regions with the highest level of similarity between the two sequences

b. Sequence Homology

- Homology is a qualitative statement

- It usually implies similarity

- It refers to shared ancestry

Sequence Similarity

Similarity can be quantified using percentages.

It does not imply homology

Similarity refers to the likeness or % identity between 2 sequences

ii) Sequence Similarity

- For protein sequences, it refers to the percentage of aligned residues that have similar physicochemical characteristics

Sequence Identity

It refers to the percentage of matches of the same amino acid residues between two aligned sequences