NAME: OGUNNIKE IYANUOLUWA. O

MATRIC NO: 18/MHS05/011

COURSE: CSC 262

DEPARTMENT: PHYSIOLOGY

COLLEGE: MHS

QUESTION

1a) List and discuss any three criterion of database searching.

1b) Explain briefly Basic Local Alignment Search Tool (BLAST) as usual in database similarity searching.

2a) Briefly discuss the differences between Dayhoff PAM Matrices and BLOSUM Matrices.

2b) Explain briefly Heuristic Database Searching.

3a) Define the following;

1. Sequence Homology
2. Sequence Similarity
3. Sequence Identity.

3b) i. Give any three methods of Alignment Algorithm.

ii. Discuss briefly Pairwise Sequence Alignment

4a) Differentiate between Global Alignment and Local Alignment.

4b) Distinguish between the following;

1. Sequence Homology and Sequence Similarity.
2. Sequence Similarity and Sequence Identity.

ANSWER

1a) (i) Sensitivity

(ii) Selectivity

(iii) Speed

1. Sensitivity

The ability required to find as many correct hits as possible. These correct hits are considered “true positives” in database searching exercise.

1. Selectivity

This criterion is also called “specificity”, which refers to the ability to exclude incorrect hits. The incorrect hits are considered “false positives” which are unrelated sequences mistakenly identified in database searching.

1. Speed

The time it takes to get result from database searches. Depending on the size of the database, speed can be a primary concern.

1b) Basic Local Alignment Search Tool (BLAST) is a major heuristic Algorithm for performing database searches. It perform faster searches because they examine only a fraction of the possible alignments examined in regular dynamic programming.

2a) Dayhoff PAM Matrices are based on scoring all amino positions in related sequences.

BLOSUM Matrices are based on substitutions and conserved positions in blocks, which represent the most alike common regions in related sequences.

2b) Heuristic Database Searching means searching a large database using the heuristic algorithms such as BLAST and FASTA. These algorithms perform faster searches because they examine only a fraction of the possible alignments examined in a regular dynamic programming.

3a) (i) Sequence Homology

This is an inference or conclusion about a common ancestral relationship drawn from sequence similarity comparison when the two sequences share a high degree of similarity.

(ii)Sequence Similarity

The percentage of aligned residues that are similar in physiochemical properties such as charge, hyophobicity and size. It is a direct result of observation from the sequence alignment .

(iii)Sequence Identity

Sequence Identity is the amount of characters which match exactly between two different sequences. For protein sequences, it refers to the percentage of matches of the same amino and residues between two aligned sequences.

3bi)

* Dynamic programming method
* Word method
* Dot matrix method

3bii) Pairwise Sequence Alignment is the process of aligning two sequences and it is the basis of database similarity searching and multiple sequence alignment. It is the basic component of many bioinformatics applications. It is also used to identify regions of similarity that may indicate functional, structural relationships between two biological sequence. It can only be used between two sequences at a time.

e.g Sequence A : C A T - T C A - C

│ │ │ │ │

Sequence B : C - T C G C A G C

4a) Global Alignment Algorithms start at the beginning of two sequences and add gaps to each until the end of one is reached. While, Local Alignment Algorithm finds the region of highest similarity between two sequences and build the alignment outward from there.

4bi)

|  |  |
| --- | --- |
| Sequence Homology | Sequence Similarity |
| It refers to shared ancestry. | It refers to the likeness of percentage identity between 2 sequences. |
| Sequence homology usually implies similarity. | Sequence similarity does not imply homology. |
| Two sequences are homologous if they are derived from a common ancestral sequence. | Sequence similarity means sharing a statistically significant number of bases or amino acids. |

4bii)

|  |  |
| --- | --- |
| Sequence Similarity | Sequence Identity |
| In a protein sequence alignment, it refers to the percentage of aligned residues that have similar physiochemical features. | In a protein sequence alignment, it refers to the percentage of match of the same amino acid residues between two aligned residues. |