**NAME:** OGUNSEMOWO AYOOLA ENIOLUWADUROTI.

**COURSE:** CSC 262.

**COLLEGE:** MEDICINE AND HEALTH SCIENCES.

**DEPERTMENT:** ANATOMY.

**MATRIC NUMBER:** 18/MHS01/254.

ASSIGNMENT.

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

1b. Explain briefly basic local alignment search tool (BLAST) as used in database similarity searching.

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

2b. Explain briefly Heuristic Database Searching

3a. Define the following (i) Sequence Homology (ii) Sequence Similarity

(iii) Sequence Identity

3b. (i) Give any three (3) 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

ANSWERS.

1a. sensitivity; which refers to the ability to find as many correct hits as possible. It is measured by the extent of inclusion of correctly identified sequence members of the same family. These correct hits are considered “true positives” in the database searching exercise.

Selectivity; It is also called specificity, which

refers to the ability to exclude incorrect hits. These incorrect hits

are unrelated sequences mistakenly identified in database

searching and are considered “false positives.”

Speed; It is the time it takes to get results from database searches. Depending on the size of the database, speed sometimes can be a

primary concern.

1b. Basic local alignment search tool does not assume that the two sequences in question have similarity over the entire length. It only finds local regions with the highest level of similarity between the two sequences and aligns these regions without regard for the alignment of the rest of the sequence regions.

2a.

|  |  |
| --- | --- |
| Dayoff PAM Matrix. | BLOSUM Matrix. |
| It is usually used to score alignments between closely related protein sequences. | It is usually used to score alignments between evolutionarily divergent protein sequences. |
| It is based on global alignments. | It is based on local alignments. |
| Mutations in global alignments are very significant. | It is based on highly conserved stretches of alignment. |
| Higher numbers in the PAM matrix naming denotes greater evolutionary distance. | Higher number in the BLOSUM matrix naming denotes higher sequence similarity and smaller evolutionary distance. |

2b. The heuristic algorithms perform faster searches because

they examine only a fraction of the possible alignments

examined in regular dynamic programming. Currently, there are two major heuristic algorithms for performing database searches: BLAST and FASTA.

3a. Sequence homology is when two sequences are descended from a common evolutionary origin.

Sequence similarity is a measure of an empirical relationship between sequences.

Sequence identity is the amount of characters which match exactly between two different sequences.

3b. i The dot matrix method; It is a graphical way of comparing two sequences in a two dimensional matrix.

The dynamic programming method; It determines optimal alignment by matching two sequences for all possible pairs of characters between the two sequences.

The word method; It is used in fast database similarity searching.

ii. Pairwise sequence alignment is the fundamental

component of many bioinformatics applications. Pairwise sequence alignment is used to identify regions of similarity that may indicate functional, structural and evolutionary relationships between two biological sequence. It is also extremely useful in structural, functional, and evolutionary analyses of sequences.

4a.

|  |  |
| --- | --- |
| Global alignment. | Local alignment. |
| It starts at the beginning of two sequences and add gaps to each until the end of one is reached. | It finds the region/regions of highest similarity between two sequences and build the alignment outward from there. |
| The two sequences to be aligned are assumed to be generally similar over their entire length. | It does not assume that the two sequences in question have similarity over the entire length. |

4b. i

|  |  |
| --- | --- |
| Sequence homology. | Sequence similarity. |
| It refers to shared ancestry. | It refers to the likeness between two sequences. |
| Two sequences are homologous if they are derived from a common ancestral sequence. | Similarity means sharing a statistically significant number of bases. |

ii.

|  |  |
| --- | --- |
| Sequence similarity. | Sequence identity. |
| It refers to the percentage of aligned residues that have similar physicochemical characteristics and can be more readily substituted for each other. | It refers to the percentage of matches of the same amino acid residues between two aligned sequences. |
| It is a measure of an empirical relationship between sequences. | It is the amount of characters which match exactly between two different sequences. |