**NAME: ODOK-OGAR DIVINE ONYODUMA**

**MATRIC NO: 18/MHS05/009**

**DEPARTMENT: PHYSIOLOGY**

**COURSE:CSC 262**

**ASSIGNMENT**

**1(a)** List and explain any three(3) criterion of database searching.

**(b)** Explain Basic Local Alignment Search Tool(BLAST) as used in database similarity searching.

**2(a)** Briefly discuss the difference between Dayhoff PAM matrices and BLOSUM Matrices.

**(b)** Explain briefly Heuristic Database Searching.

**3(a)** Define the following;

* Sequence Homology
* Sequence Similarity
* Sequence Identity

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

**(ii)** Discuss briefly Pairwise Sequence Alignment.

**4(a)** Differentiate between Global Alignment and Local Alignment.

**(b)** Distinguish between the following;

* Sequence Homology and Sequence Similarity.
* Sequence Similarity and Sequence Identity.

**ANSWERS.**

1. **SENSITIVITY:** 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 database searching exercise.

* **SELECTIVITY:** this is also called **specificity**, 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:** refers to the time it takes to get results from database searches.

b) BLAST( Basic Local Alignment Search Tool) is a type of heuristic algorithm used to perform faster searches because they examine only a fraction of the possible alignments examined I regular dynamic programming.

1. Dayhoff PAM Matrices are based on scoring all amino acid 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.

b) Heuristic Database Searching: Heuristic search refers to a search strategy that attempts to optimize a problem by iteratively improving the solution based on a given heuristic function or a cost measure. A heuristic search method does not always guarantee to find an optimal or the best solution, but may instead find a good or acceptable solution within a reasonable amount of time and memory space.

1. Sequence Homology: sequence homology is an inference or a conclusion about a common ancestral relationship drawn from sequence similarity comparison when the two sequences share a high enough degree of similarity.

* Sequence Similarity: Sequence similarity is a measure of an empirical relationship between sequences. Its common objective is establishing the likelihood for sequence homology.
* Sequence Identity: Sequence identity is the amount of characters which match exactly between two different sequences.

b)(i) Dot Matrix Method

Dynamic Programming Method

Word method.

(ii) Pairwise Sequence Alignment: is the process of aligning two sequences and is the basis of database similarity searching and multiple sequence alignment.

•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 or/ evolutionary relationships between two biological sequence e.g say: (protein or nucleic acid)

• It is extremely useful in structural, functional, and evolutionary analyses of sequences. Pairwise sequence alignment provides inference for the relatedness of two sequences.

1. In global alignment, two sequences to be aligned are assumed to be generally similar over their entire length. Alignment is carried out from beginning to end of both sequences to find the best possible alignment across the entire length between the two sequences. You take entirety of both sequences into consideration when finding alignment.

Local alignment, on the other hand, 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. When you take small portion into account.

b)(i) Sequence Homology VS Sequence Similarity

Homology:

* Homology refers to shared ancestry
* Two sequences are homologous if they are derived from a common ancestral sequence.
* Homology usually implies similarity.

Similarity:

* Refers to the likeness or % identity between 2 sequences.
* Means sharing a statistically significant number of bases or amino acids.
* Does not imply homology.

(ii) Sequence Similarity VS Sequence Identity

Similarity:

* refers to the percentage of aligned residues that have similar physicochemical characteristics and can be more readily substituted for each other.
* is a measure of an empirical relationship between sequences.

Identity:

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