

## **MY Assignment**

**Name: Adesina Benita tomisin**

**Department: human anatomy**

**College: medicine and health sciences**

**Matric number: 18/mhs03/001**

**Course : csc 262 - bioinformatics 2**

**Level: 200**

### **Answer**

1a. Criterion of database searching

**Sensitivity:** This 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.

**Selectivity:** This refers to the ability to exclude incorrect hits.

**Speed:** This is the time it takes to get results from database searches.

1b. In bioinformatics BLAST is an algorithm and program used for comparing primary biological sequence information, such as the amino acid sequence of proteins or nucleotides of DNA and RNA sequences. A blast search enables a researcher to compare a subject protein or nucleotide sequence with a database of sequences and identify sequences that resemble the query sequence. When working with genes, BLAST can locate common genes in two related species, and can be used to map annotations from one organism to another.

2a.

### **DIFFERENCE BETWEEN PAM AND BLOSUM MATRICES**

#### **PAM matrices**

#### **BLOSUM matrices**

Based on global alignment	Based on local alignment
Alignment has high similarities	Alignment has a low similarities
Used to score alignment between closely related protein sequences	Used to score alignments between evolutionarily divergent protein sequences

Higher numbers in the PAM matrix naming denotes great evolutionary distance.	Higher numbers in the BLOSUM matrix naming denotes higher sequences similarity and smaller evolutionary distance.
--	---

2b. Heuristic database searching is a technique or methods designed for solving problems more quickly when classic methods are too slow. The heuristic algorithms performs faster searches because they examine only a fraction of the possible alignments examined in regular dynamic programming.

3a

**i. Sequence homology:** This 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.

**ii. Sequence similarity:** This is the percentage of aligned residues that are similar in physiochemical properties such as size, charge, and hydrophobicity.

**iii. Sequence identity:** This refers to the percentage of matches of the same amino acid residues between two aligned sequences.

3bi

The dot matrix method

The dynamic programming method

The word method

3bii.

Pairwise sequence alignment is the process of aligning two sequence and is the basis of database similarity searching and multiple sequence alignment. It is a fundamental component of many bioinformatics applications. It is used to identify regions of similarity that may indicate functional, structural and evolutionary relationships between two biological sequences.

4a.

<b>GLOBAL ALIGNMENTS</b>	<b>LOCAL ALIGNMENTS</b>
Two sequences to be aligned are assumed to be generally similar over their entire length.	Does not assume that two sequences in question have similarity over the entire length.
Alignment is carried out at the beginning to the end of both sequence	It only finds local regions with the highest level of similarity between the two sequences and aligns these regions without regard for the rest of the sequence

4bi. **SEQUENCE SIMILARITY**

**SEQUENCE HOMOLOGY**

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

ii. **SEQUENCE SIMILARITY**

**SEQUENCE IDENTITY**

It refers to the percentage of aligned residues that have similar physicochemical characteristics and can do 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.