Name: Ajibade Opeyemi Melody

Matric no: 18/MHS01/047

Department: Anatomy

Course: CSC 262

**Assignment**

1. Sensitivity

ii. Selectivity

iii. Speed.

1. 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. These correct hits are considered “true positives” in the database searching exercise.
2. Selectivity: also called specificity which refers to the ability to exclude incorrect hits. These incorrect hits are unrelated sequences mistakenly identified in database searching exercise.
3. Speed: which is the time it takes to get results from database searches.
4. BLAST ( Basic Local Alignment Search Tool) is an algorithm and program for comparing biological sequence information, such as the amino acid sequences of proteins or the nucleotides of DNA and/or RNA sequences. A BLAST search enables a researcher to compare a subject protein or nucleotide sequence (called a query) with a library or database of sequences that resemble the query sequence above a certain threshold. It is better for proteins than for nucleotides. BLAST programs search databases in a special compressed format. To use your own private database with BLAST, you need to format it in BLAST format.
5. Day Hoff PAM Matrices are used to store alignment between closely related protein sequences and are based on global alignment while BLOSUM Matrices are used to score alignments between evolutionarily divergent protein sequences and are based on local alignments.

b. Heuristic database searching works by finding short stretches of identical letters in two sequences. There are two major heuristic algorithms for performing database searches:

I. Text based security of annotations and related data.

II. Sequence based security.

1. I. Sequence homology is the biological homology between DNA, RNA or protein sequences, defined in terms of shared ancestry in the evolutionary history of life.
   1. Sequence similarity is a measure of an empirical relationship between sequences.
   2. Sequence identity is the amount of characters which match exactly between two different sequences.

b. Dot matric methods

Dynamic programming.

Word methods.

ii. Pairwise Sequence Alignment: Pairwise sequence alignment methods are used to find the best-matching piecewise (local or global) alignments of two query sequences. Pairwise alignments can only be used between two sequences at a time, but they are efficient to calculate and are often used for methods that do not require extreme precision (such as searching a database for sequences with high similarity to a query). The three primary methods of producing pairwise alignments are dot-matrix methods, dynamic programming, and word methods; however, multiple sequence alignment techniques can also align pairs of sequences. Although each method has its individual strengths and weaknesses, all three pairwise methods have difficulty with highly repetitive sequences of low information content - especially where the number of repetitions differ in the two sequences to be aligned. One way of quantifying the utility of a given pairwise alignment is the 'maximum unique match' (MUM), or the longest subsequence that occurs in both query sequences. Longer MUM sequences typically reflect closer relatedness.

1. The very basic difference between a local and a global alignments is that in a local alignment, you try to match the query with a substring (a portion) of your subject (reference) whereas in a global alignment with the subject and therefore you may end up with a lot of gaps in alignment of the sizes of query and subject are dissimilar.

Bi. Sequence Homology is a statement common evolutionarily ancestry of two sequences and can only be true or false. We can rarely be certain about this, it is therefore usually a hypothesis that may be more or less probable while Sequence similarity is the degree of likeness between two sequences usually expressed as a percentage of similar or identical residues over a given length of the alignment and can usually be easily calculated.

ii. The key difference between sequence similarity and sequence identity is that sequence similarity is the resemblance between two sequences in comparison while Sequence identity is the number of characters that match exactly between two different sequences.