Name: Ugorji Nkemjika Jane

Matric No: 18/mhs05/012

Department: Physiology

College: Medicine And Health Sciences

Course Code: CSC 262

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

1. Sensitivity: is 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
2. Selectivity: also called specificity which refers to the inability to exclude incorrect hits. These incorrect are unrelated hits are unrelated sequences mistakenly identified in database searching and are considered “false positives”
3. Speed: 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. Ideally one wants to have the greatest sensitivity, selectivity, and in speed in database searches.

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

It is an algorithm and program for comparing primary biological sequence information, such as the amino acid sequences of proteins or the nucleotides of DNA and RNA sequences. A BLAST search engine enables a researcher to compare a subject protein or nucleotide sequence and identify library sequences that resemble the query sequence above a certain threshold.

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

# PAM Matrices

PAM Matrices are used to score alignments between closely related protein sequences , they are based on global alignments. Alignments have high similarity than BLOSUM alignments. Mutations in global alignments are very significant. Higher numbers in the PAM matrix naming denotes greater evolutionary distance. Example:PAM 250 is used for more distant sequences than PAM 120

# BLOSUM Matrices

BLOSUM Matrices are used to score alignments between evolutionary divergent protein sequences, they are on local alignments. Alignments have low similarity than PAM alignments. Their mutation is based on highly conserved stretches of alignments. Higher numbers in the BLOSUM matrix naming denotes higher sequence similarity and smaller evolutionary distance. For example: BLOSUM 80is used for closely related sequences than BLOSUM 62.

2b: Explain briefly Heuristic Database searching

The Heuristic Database is used instead of the Smith-Waterman algorithm due to its speed, the perform faster searches because they examine only a fraction of the possible alignments examined in regular dynamic programming.

3a: Define the following

1. Sequence: is a way of arranging the sequences of DNA and RNA or protein to identify regions of similarity that maybe a consequence of functional, structural, or evolutionary relationships between sequences.
2. Homology: is an inference or a conclusion about a common ancestral relationship drawn from sequence similarity comparisons when the two sequences share a high enough degree of similarity.
3. Sequence similarity: is a measure of an empirical relationship between sequences.
4. Sequence identity: is the amount of characters which match exactly between two different sequences.

3b i. Give any three methods of alignment algorithm

* The dynamic programming method
* Dot matrix method
* Word method

ii. Discuss briefly pairwise sequence alignment

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

4a: Differentiate between global alignment and local alignment

# Global alignment

**▪** Starts at beginning of two squares and add gaps to each until the end of one is reached

▪ Two sequences to be aligned are assumed to be generally similar over their entire length.

# Local alignment

▪ Finds the region or regions of highest similarity between two sequences and build the alignment outward from there

▪ Does not assume that two sequences in question have similarity over the entire length

4b: Distinguish between the following

i. Sequence homology and sequence similarity

# Sequence Homology

▪ Homology usually implies similarity

▪ Homology refers to common evolutionary ancestry of two sequences

▪ Two sequences are homologous if they are derived from a common ancestral sequence

# Sequence Similarity

**▪** Similarity does not imply homology

▪ similarity refers to the likeness or percentage identity between 2 sequences

▪ Similarity means sharing a statistically

ii. Sequence similarity and sequence identity

# Sequence Similarity

**▪** Similarity does not imply homology

▪ similarity refers to the likeness or percentage identity between 2 sequences

▪ Similarity means sharing a statistically

# Sequence Identity

**•** Refers to the percentage of matches of the same amino acid residues between two aligned sequences