**NAME: OKEKE OTITOCHI MARYANN**

**MATRIC NO: 16/SCI01/030**

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

1. **Sensitivity**: This 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 and these correct hits are considered ***“true positives”*** in the database searching exercise.
2. **Selectivity**: Also called Specificity which is the ability to exclude incorrect hits. This incorrect hits are unrelated sequences mistakenly identified in database searching and are considered as ***“false positives”.***
3. **Speed**: This is the times it takes to get results from database searches.

1b). BLAST (Basic Local Alignment Search Tool) is the most common used sequence similarity search tool and it is a LOCAL alignment algorithm. It uses heuristic method to perform fast local alignment search and to find the highest scoring local alignment between the query sequence and the search set sequence.

BLAST programs search database in a special compressed format and uses “common words” for the initial database search.

2a). DAYHOFF PAM MATRIX is the replacement of a single amino acid in the primary structure of a protein with another single amino acid, which is accepted by the processes of natural selection. It is also a matrix where each column and row represent one of the twenty standards amino acids **while** BLOSUM MATRIX **(BLOcks SUbstitution Matrix)** matrix is asubstitution matrix used for sequence alignment of proteins. BLOSUM matrices are used to score alignments between evolutionary divergent protein sequences and they are based on local alignments.

2b). Heuristic Database Searching- This is a computational strategy to find an empirical or near optimal solution by using the rules of thumb. The Heuristic algorithms performs faster searches because they examine only a fraction of the possible alignments examined in regular dynamic processing.

3a)

1. 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.
2. It is also the biological homology between DNA, RNA, or protein sequences defined in the terms of shared ancestry in the evolutionary history of life.
3. Sequence Identity is the amount of characters which match exactly between two different sequences. Sequence identity is the percentage of aligned residues that are similar in physiochemical properties such as size, charge, and hydrophobicity.
4. Sequence Similarity is the measure of an empirical relationship between sequences. A common objective of sequence calculations is establishing the likelihood for sequence homology, the chance that sequences have evolved from a common ancestor.

3bi).

1. Dot Matrix
2. Dynamic Programming
3. Word Method

3bii). Pairwise Sequence Alignment- This method is used to find the best-matching local or global alignments of two query sequences. Pairwise alignments can only be used between two sequences at a time. It is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences (protein or nucleic acid).

4a. Global Alignment Algorithms start at the beginning of two sequences and add gaps to each until the end of one is reached while Local Alignment Algorithm finds the region (or regions) of highest similarity between two sequences and build the alignment outward from there.

In, Global Alignment Algorithms two sequences to be aligned are assumed to be generally similar over their entire length while Local Alignment does not assume that the two sequences in question have similarity over the entire length.

4b.

|  |  |  |
| --- | --- | --- |
| S/N | SEQUENCE HOMOLOGY | SEQUENCE SIMILARITY |
| 1 | Refers to shared ancestry | Refers to the likeness or % identity between 2 sequences |
| 2 | Two sequences are homologous if they are derived from common ancestral sequence | Means sharing a statistically significant number of bases or amino acids. |
| 3 | Homology usually implies similarity | Similarity does not imply homology |

|  |  |  |
| --- | --- | --- |
| S/N | SEQUENCE SIMILARITY | SEQUENCE IDENTITY |
| 1 | Measure of empirical relationship between sequences | Amount of characters which match exactly between two different sequences |
| 2 | In a protein sequence alignment, similarity refers to the percentage of aligned residues that have similar physicochemical characteristics and can be more readily substituted for each other. | In a protein sequence alignment, sequence identity refers to the percentage of matches of the same amino acid resides between two aligned sequences. |