NAME: IKECHUKWU JOY NMESOMA

MATRIC NO: 18/MHS05/008

DEPARTMENT: PHYSIOLOGY

CSC Assignment

15/04/2020

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

Answer

1. Speed
2. Sensitivity
3. Specificity

SPEED: It is the time taken to get results from database searches

SENSITIVITY: It is the ability to find as many correct hits as possible. The correct hits are considered ‘true positives’.

SPECIFICITY: It is the ability to remove or exclude incorrect hits. The incorrect hits are known as ‘false positives’.

1b. The Basic Local Alignment Search Tool ( BLAST) is one of the two major heuristic algorithms for performing database searches. BLAST find regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

|  |  |
| --- | --- |
| PAM  Are used to score alignments between closely related protein sequence.  Based on global alignments.  Alignments have high similarity than BLOSUM alignments.  Mutations in global alignments are very significant | BLOSUM  Are used to score alignments between evolutionally divergent protein sequences.  Based on local alignments.  Alignments have low similarity than PAM alignments.  Based on highly conserved stretches of alignments. |

2a.

2b. Heuristic database searching: The heuristic algorithms performs faster searches because they check only a fraction of possible alignments. To speed of comparison, heuristic database must be used. There are currently two major heuristic database which are BLAST and FASTA.

3a. Define the following:

Sequence Homology: it is the biological homology between DNA, RNA and protein sequences, defined in terms of shared ancestry in the evolutionary history of life

Sequence similarity: it is a measure of an empirical relationship between sequence; its common objective is establishing the likelihood of sequence homology.

Sequence Identity: It is the amount of characters which match exactly between two different sequences.

3b. i) Three methods of alignment algorithm

1. Dot matrix method
2. Word method
3. Dynamic programming method

ii) Pair wise sequence alignment: This method is used to find the best matching local and global alignments of two query sequences. It can only be used between two sequences at a time. It can also be used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences.

4a.

|  |  |
| --- | --- |
| GLOBAL ALIGNMENT  An attempt is made to align the entire sequence (end to end alignment).  Suitable for aligning two closely related sequences.  Global alignments are usually done for comparing homologous genes like comparing two genes with same function. | LOCAL ALIGNMENT  Finds local regions with the highest level of similarity between two sequences.  Suitable for aligning more divergent sequences or distantly related sequences.  Used for finding out conserved patterns in DNA sequences or conserved domains or motifs in two proteins. |

4bi.

|  |  |
| --- | --- |
| SEQUENCE HOMOLOGY  Homology usually implies similarity  Homology refers to shared ancestry  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 significant number of bases or amino acids. |

4bii)

|  |  |
| --- | --- |
| SEQUENCE SIMILARITY  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.  Is a measure of an empirical relationships between sequences | SEQUENCE IDENTITY  Refers to the percentage of matches of the same amino acid residues between two aligned sequences.  Is the amount of characters which match exactly between two sequences. |