NAME: OKAM EDNA ACHI CHINMA

MATRIC NUMBER: 19/MHS03/015

DEPARTMENT: HUMAN ANATOMY

1ai. Selectivity/Specificity- this refers to the ability to exclude incorrect hits. This incorrect hits are unrelated sequences mistakenly identified in database searching and are considered false positives.

ii. 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.

iii. Speed- which 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, and speed in database searches.

b. Basic local alignment search tool does not assume that the two sequences in question have similarity over the entire length. It only finds regions with the highest level of similarity between the two sequences that aligns these regions without regard for the alignments of the rest of the sequence regions.

2a.

|  |  |
| --- | --- |
| Dayoff PAM Matrix | BLOSUM Matrix |
| Mutations in global alignments are very significant. | It is based on highly conserved stretches of alignment |
| it is usually used to score alignments between closely related protein sequences. | it is usually used to score alignments between evolutionally divergent protein sequences. |
| Higher numbers in the PAM matrix naming denotes greater evolutionary distance. | Higher number in the BLOSUM matrix naming denotes higher sequence similarity and smaller evolutionary distance |
| It is based on global alignments. | It is based on local alignments. |

2b. The heuristic algorithms perform faster searches because they examine only a fraction of the possible alignments examined in regular dynamic programming. Currently there are two major heuristic algorithms for performing database searches, BLAST and FASTA.

3a. -Sequence homology is when two sequence are descended from a common evolutionary origin.

-Sequence similarity is a measure of an empirical relationship between sequences.

-sequence identity is the amount of characters which match exactly between two different sequences,

3bi. The dot matrix method: it is a graphical way comparing two sequence in a two-dimensional matrix.

-The dynamic programming method- it determines optimal alignment by matching two sequences for all possible pairs of characters between the two sequences,

-The word method- It is used in fast database similarity searching.

ii. Pairwise sequence alignment is the fundamental component of many bioinformatics applications. Pairwise sequence alignment is used to identify regions of similarity that may indicate functional, structural and evolutionary relationships between two biological sequence. It is also extremely useful in structural, functional, and evolutionary analyses of sequences.

4a.

|  |  |
| --- | --- |
| Global Alignment | Local Alignment |
| It starts at the beginning of two sequences and add gaps to each until the end of one is reached. | It finds the region(s) of highest similarity between two sequences and build the alignment outward from there. |
| The two sequences to be aligned are assumed to be generally similar over their entire length. | It does not assume that the two sequences in the question have similarity over the entire length. |

4bi.

|  |  |
| --- | --- |
| Sequence Homology | Sequence Similarity |
| It refers to shared ancestry | It refers to the likeness between two sequences. |
| Two sequences are homologous if they are derived from a common ancestral sequence. | Similarity means sharing a statistically significant number of bases. |

ii.

|  |  |
| --- | --- |
| Sequence Similarity | Sequence Identity |
| It refers to the percentage of aligned residues that have similar physicochemical characteristics and can be 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. |