JAMES, Emem Grace

18/MHS01/192

CSC 262

ANSWERS

1a. 3 criteria for database searching are namely;

* Sensitivity

This refers to then ability to find as many correct hits as possible. It is measured by the extent of inclusion of correctly identified sequence membered of the same family.

* Selectivity or Specificity

This refers to the ability to exclude incorrect hits. These incorrect hits are unrelated sequences mistakenly identified in database searching and are considered ‘false positives’.

* Speed

This is the time it takes to get results from data base searches

 **b.** Basic Local Alignment Search Tool (BLAST) is a heuristic algorithm which can perform faster searches because it can examine only a fraction of the possible alignments examined in regular dynamic programming.

2bs. Heuristic Database Searching can be said to be searching a large database using the dynamic programming methods, although accurate and reliable, is too slow and impractical when computational resources are limited. Thus, speed of searching became an important issue. To speed up the comparison, heuristic methods have to be used.

3a. Define the following terms

* 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.
* Sequence alignment a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences.
* Sequence similarity refers to the percentage of aligned residues that have similar physicochemical characteristics and can be more readily substituted for each other.

b) i.3 methods of alignment algorithm

* Dot Matrix Method is the most basic sequence alignment method is the dot matrix method, also known as the dot plot method. It is a graphical way of comparing two sequences in a two dimensional matrix. In a dot matrix, two sequences to be compared are written in the horizontal and vertical axes of the matrix.
* Dynamic programming is a method that determines optimal alignment by matching two sequences for all possible pairs of characters between the two sequences.
* The word method is used in fast database similarity searching .

 ii. Pairwise sequence alignment is the process of aligning two sequences and is the basis of database similarity searching and multiple sequence alignment. It is the fundamental component of many bioinformatics applications. Pairwise sequence alignment is used to identify regions of similarity that may indicate functional, structural and or/ evolutionary relationships between two biological sequence e.g say: (protein or nucleic acid)

• It is extremely useful in structural, functional, and evolutionary analyses of sequences. Pairwise sequence alignment provides inference for the relatedness of two sequences.

4a. In global alignment, two sequences to be aligned are assumed to be generally similar over their entire length. Alignment is carried out from beginning to end of both sequences to find the best possible alignment across the entire length between the two sequences. You take entirety of both sequences into consideration when finding alignment while,

 Local alignment on the other hand, does not assume that the two sequences in question have similarity over the entire length. It only finds local regions with the highest level of similarity between the two sequences and aligns these regions without regard for the alignment of the rest of the sequence regions. When you take small portion into account.

 b. Distinguish the following

* 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. On the other hand, Sequence similarity is a direct result of observation from the sequence alignment.
* In a protein sequence alignment, *sequence identity* refers to the percentage of matches of the same amino acid residues between two aligned sequences. Similarity refers to the percentage of aligned residues that have similar physicochemical characteristics and can be more readily substituted for each other.