ASIBOR OSEMHANHU IMMANUELLA

BIOINFORMATICS

18/mhs01/087

ANSWERS:

1ai) 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.

ii) Selectivity: 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”.

iii) Speed: This is the time taken to get results from database searches. Depending on the database, speed sometimes can be a primary concern.

1b) BLAST is a set of algorithms that attempt to find a short fragment of a query sequence that aligns perfectly with a fragment of a subject sequence found in a database. BLAST looks for short sequence in the query that matches short sequences found in the database.

2a) PAM matrices are used to score alignments between closely related protein sequences while BLOSUM matrices are used to score alignments between evolutionary divergent protein sequences. PAM is based on global alignments while BLOSUM is based on local alignment.

b) Heuristic database searching works by finding short stretches of identical letters in two sequences. There are two major heuristic algorithms for performing database searches.

3ai) 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.

ii) Sequence similarity is a measure of an empirical relationship between sequences. It’s common objective is establishing the likelihood for sequence homology.

iii) Sequence identity is the amount of characters which match exactly between two different sequences.

3bi) Dot Matrix method

Dynamic programming method

Word method

3bii) pairwise sequence alignment is the fundamental component of many bioinformatics applications. It is the process of aligning two sequences and is the basis of database similarity searching and multiple sequence alignment. It is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences.

4a) Global alignment algorithms starts at the beginning of two sequences and add gaps to each until the end of one is reached while local alignments algorithms finds the region of highest similarity between two sequences and build the alignments outward from there.

4bi)

|  |  |
| --- | --- |
| Sequence homology | Sequence similarity |
| Refers to shared ancestry | Refers to the likeness of % identity between two sequences |
| Two sequences are homologous if they are derived from a common ancestral sequence | Similarity means sharing statistically significant number of bases or amino acids |
| Usually implies similarities | Does not imply homology |

4bii) Sequence similarity is a measure of an empirical relationship between sequences while sequence identity is the amount of characters which match exactly between two different sequences.