

PETER MMENYENE USEN

18/MHS01/331

PHYSIOLOGY

CSC 262

19/04/2020

1a) Three criterion of database searching.

a) Sensitivity: This is the ability to find as many correct hits as possible. These correct hits are considered 'true positives' in the database searching exercise. It is measured by the extent of including of correctly identified sequence members of the same family.

b) Selectivity: This is also known as specificity. It refers to the ability to excuse incorrect hits. These incorrect hits are unrelated sequences mistakenly identified in database searching and are considered 'false positives'.

c) Speed: Speed refers to how long it takes to get results from database searches. Speed can sometimes be a primary concern depending on the size of the database.

1b) Basic Local Alignment Search Tool (BLAST)

BLAST is an algorithm and program used for comparing primary biological sequence information, such as the amino acid sequences of proteins or the nucleotides of DNA or RNA sequences. Different types of BLASTs are available according to the query sequences and the target databases. It is one of the most widely used bioinformatics programs for sequence searching. It addresses a fundamental problem in bioinformatics research.

2a) i) Dayhoff PAM matrices: Are used to score alignments between closely related protein sequences.

BLOSUM matrices: Used to score alignments between evolutionary divergent protein.

sequences.

ii) Dayhoff PAM matrices: Based on global alignments.

BLOSUM matrices: Based on local alignments.

iii) Dayhoff PAM matrices: Alignments have high similarity.

BLOSUM matrices: Alignments have low similarity.

iv) Dayhoff PAM matrices: Mutations in global alignments are very significant.

BLOSUM matrices: Mutations are based on highly conserved stretches of alignments.

2b) Heuristic Database Searching: This is the use of heuristic methods for database searching. Heuristic algorithms perform faster searches because they examine only a fraction of possible alignments examined in regular dynamic programming. There are two major heuristic algorithms - BLAST and FASTA.

3a) Sequence homology: This is an inference or a conclusion about a common ancestral relationship. It is when two sequences are descended from a common evolutionary origin.

b) Sequence similarity: This is a measure of an empirical relationship between sequences. It establishes the likelihood for sequence homology.

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

3b) Methods of alignment algorithm.

i) Dot matrix method.

ii) Word method.

iii) Dynamic programming method.

3bii) Pairwise Sequence Alignment: This is the process of aligning two sequences and is the basis of database similarity searching and multiple sequence alignment.

4a) In global alignment, two sequences to be aligned are assumed to be generally similar over

their entire length, while in local alignment, it does not assume the two sequences have similarity over the length.

b) Sequence homology is when two sequences are descended from a common evolutionary origin while sequence similarity is the percentage of aligned residues that are similar in physiochemical properties.

4bii) Sequence similarity is the resemblance between two sequences in comparison while sequence identity is the number of characters that match exactly between two different sequences.