**DATE; 6TH OF APRILL 2020**

**NAME; ABIMBOLA OLAMIDE REBECCA**

**MATRIC NUMBER; 18/MHS02/004**

**DEPARTMENT; HUMAN ANATOMY**

**COURSE; BIOINFORMANCE** **(CSC)**

1b .Explain briefly basic local alignment search tools (BLASTS) as used in database similarity searching.

Blast is a pair wise local alignment search tools that is designed to operate more quickly than exact methods, but without a quarantine of finding the best possible alignment.

At present blast is the preferred tool for searching large sequence database such as genbank.

Searching a database with a query sequence can be considered to be a problem in local alignment searches.

PSI-BLAST

Psi-blast allows users to construct and perform a blast search with a custom, position-specific, scoring matrix which can help find distant evolutionary relationship .PHI blast functionality is also available to restrict result using patterns.

2a .differences between PAM AND BLOSUM

PAM

1. Matrices are found base on an explicit evolutionary model (i.e. replacements are counted on the branches of a phylogenetic tree), whereas the blosum matrices are based on an implicit model of evolution

2. The Pam matrices are based on mutations observed throughout a global alignment, this include both highly conserved and highly mutable region. The blosum matrices are based only on highly conserved regions in series of alignments forbidden to contain gaps.

3 The method used to count the replacement is different unlike the pam matrix, the blosum procedure uses groups of sequences within which not all mutations are counted the same

4 Higher numbers in the PAM matrix naming scheme denote larger evolutionary distance, while larger numbers in the blosum matrix naming scheme denote higher sequences similarity and therefore smaller evolutionary distance.

Example; PAM150 used for more distant sequences than PAM100; blosum62 is used for closer sequences than blosum50

2b Heuristic alignment motivation

Too slow for large database with high query traffic

Heuristic methods do fast approximation to dynamic programming

FASTA (PEARSON&LIPMAN, 1988)

BLAST (ALTSCHULET, 1990)

Consider the task of searching swiss-prot against a query sequence

Say our query sequences is 362 amino-acid long

Swiss-prot release 38 contain 29,085,265 amino acids

Finding local alignment via dynamic programming would entail matrix operation

Many servers handled thousands of such queries a day

NCBI>50,000

3a.sequence homology is the biological homology between DNA&RNA or protein sequences, defined in terms of shared ancestry in the evolutionary history of life .Two segment of DNA can have shared ancestry because of three phenomena, either a speciation event(orthodox) or a duplication event (paralogs)or else a horizontal(or lateral) gene transfer event (xenologs)

Sequences similarity is a measure of an empirical relationship between common objectives of sequences. A common objective of sequences similarly calculation is establishing the like hood for sequences homology, the chance that sequences have evolved from common ancestor’s similarity score is therefore aimed to approximate the evolutionary distance between pair of nucleotide or protein sequences. Many implementations for measuring sequences similarity exist where a general aim is to infer structural or functional characteristic of an annotated molecular sequence.

Sequences identity is the amount of character which matches exactly between two different sequences. Hereby gaps are not counted and the measurement is relational to the shorter of the two sequences .this has the effect that sequences identity is not transitive i.e. if sequences A=B and B=C then A is not necessary equal C in terms of of the identity distances measure

**3b three methods of alignment algorithm**

1. Pairwise and multiple alignment

2. Progressive alignment and insertion

3.Progressive alignment algorithm for sequences with insertion

Pair wise sequences alignment is used to identify region of similarity that may indicate functional, structural and evolutionary relationship between two biological sequences (protein or nucleic acid)

**4a**

|  |  |
| --- | --- |
| Global sequences alignment | Local sequences alignment |
| 1.in global alignment an attempt is made to align the entire sequences (end to end alignment) | Find local regions with the highest level of similarity between the two sequences |
| 2.A global alignment contains all letters from both query and target sequences | A local alignment aligns a substring of the query sequences of the target sequences |
| 3.if two sequences have approximately the same length and a quite similar, they are suitable for global alignment | Any two sequences can be locally aligned as local alignment finds stretches of sequences with high level of matches without considering the alignment of rest of the sequences region |
| 4.sultable for aligning closely related sequences | Suitable for aligning more divergent sequences or distantly related sequences |
| 5.Global alignment are usually done for complain homologues gene like comparing two genes with same function (in human vs. mouse) or comparing two protein with similar function | Used for finding out conserved pattern in DNA sequences or conserved domain or motifs in two protein |
| 6.A general global alignment technique is the Needleman-wunsch algorithm  Example of global alignment tools  -Emboss needle  -Needleman –wunsh global align nucleotide sequences (specialized Blast) | A general local alignment method is smith-waterman algorithm,  Example of local alignment tools  -Blast  -Emboss water  -Lalign |

**4b sequences homology and sequences similarity**

Sequences homology is the biological homology between DNA, RNA or protein sequences defined in term of shared ancestry in the evolutionary history of life

Sequences similarity is the method of searching sequences database by using alignment to a query sequences

Sequences similarity and sequences identity

Sequences similarity is the method of searching sequences database by using alignment to a query sequences

Sequences identity is the amount of characters which match exactly between sequences. Hereby gaps are not counted and the measurement is relational to the shorter of the two sequences.