

Name: Ginika Obiadi

Department: Anatomy

Matric No: 18/MHS03/009

Course Code: CSC 262

Course Title: Bio Informatics

## ASSIGNMENT

1. a. List and discuss any three criterion of database searching  
b. Explain briefly Basic Local Alignment Search Tool (BLAST) as used in database similarity searching.
2. a. Briefly discuss the differences between Dayhoff PAM Matrices and BLOSUM Matrices.  
b. Explain briefly Heuristics Database Searching
3. a. Define the following
  - i. Sequence Homology
  - ii. Sequence Similarity
  - iii. Sequence Identityb. Give any three methods of Alignment Algorithm, Discuss briefly Pairwise Sequence Alignment.
4. a. Differentiate between Global Alignment and Local Alignment  
b. Distinguish between the following:
  - i. Sequence Homology and Sequence Similarity
  - ii. Sequence Similarity and Sequence Identity

1. List and discuss any three criteria of database searching:

- a. Mathematical Criteria
- b. Like Criteria
- c. Logical Operations

a. Mathematical Criteria

>	More Than or After	> <b>100</b> for example
<	Less Than or Before	< <b>100</b> for example
>=	More Than or Equal To	>= <b>100</b> for example
<=	Less Than or Equal To	<= <b>100</b> for example

b. Like Criteria

Between AND	<b>Between 1980 And 2000</b> for example
Or	<b>January Or April</b> for example
Not	Used to exclude data – <b>Not 2000 for example</b>

c. Logical Operators

Like **	Contains the words..... – <b>Like *kite*</b> for example
Like a*	Fields <b>starting with a</b> for example
Like *a	Fields <b>ending with a</b> for example

b. Explain briefly Basic Local Alignment Search Tool (BLAST) as used in database similarity searching:

**Identifying Species:** With the use of BLAST, you can possibly correct and identify a species or find homologous species. This can be useful, for example, when you are working with a DNA sequence from an unknown species.

**Locating Domains:** When working with a protein sequences you can input it into BLAST, to locate known domains within the sequence of interest.

**Establishing Phylogeny:** Using the results received through BLAST you can create a phlogenetic tree using the BLAST web-page. Phylogenies based on BLAST alone are less reliable than other purpose-built computational phlogenetic methods, so should only be relied upon for “first pass” phlogenetic analyses.

**DNA Mapping:** When working with a known species, and looking to sequence a gene at an unknown location, BLAST can compare the chromosomal position of the sequence of interest, to relevant sequences in the database. NCBI has a “Magic-BLAST” tool built around BLAST for this purpose.

**Comparison:** When working with genes, BLAST can locate common genes in two related species and can be used to map annotations from one organism to another.

2 .a) Briefly discuss the differences between Dayhoff PAM Matrices and BLOSUM Matrices:

PAM	BLOSUM
PAM matrices are used to score alignments between closely related protein sequences.	BLOSUM matrices are used to score alignments between evolutionarily divergent protein sequences.
Based on global alignments	Based on logical alignments
Alignments have high similarity than BLOSUM alignments	Alignments have low similarity than PAM alignments
Mutations in global alignments are very significant	Based on highly conserved stretches of alignments
Higher numbers in the PAM matrix naming denotes greater evolutionary distance	Higher numbers in the BLOSUM matrix naming denotes higher sequence similarity and smaller evolutionary distance
Example: PAM 250 is used for more distant sequences than PAM 120	Example: BLOSUM 80 is used for closely related sequences than BLOSUM 62

b. Explain briefly Heuristics Database Searching:

Is a function that ranks alternatives in search algorithms at each branching step based on available information to decide which branch to follow. For example, it may be approximate the exact solution. Technique designed for solving a problem more quickly when classic methods are too slow. They may be used in situations where there are no known algorithms.

3 .a.) Define the following:

- i. Sequence Homology- Is the biological homology between DNA, RNA, or protein sequences, defined in terms of shared ancestry.
- ii. Sequence Similarity- Is the amount of characters which match exactly between two different sequences.
- iii. Sequence Identity- Is the amount of characters which match exactly between sequences.

b.

- i. Give any three (3) methods of Alignment Algorithm:
  - I. Dot-matrix methods
  - II. Dynamic programming
  - III. Words methods
- ii. Discuss briefly Pairwise Sequence Alignment: Is used to identify regions of similarity that may indicate functional, structural, and evolutionary relationships between two biological sequences. (Protein and nucleic acid).

#### 4. Differentiate between Global Alignment and Local Alignment

Global Alignment	Local Alignment
Is when you take the entirety of both sequences into consideration of when finding alignments.	You can only take a small portion into account.

b. Distinguish between the following:

i. Sequence Homology and Sequence Similarity

Sequence Homology	Sequence Similarity
Is the biological homology between protein sequences	Is a measure of an empirical relationship between sequences.

ii. Sequence Similarity and Sequence Identity

Sequence Similarity	Sequence Identity
Is a measure of an empirical relationship between sequences	Is the amount of characters which match exactly between sequences.