**NAME:** ADEBIYI ITUNUNOLUWA ISAAC

**MATRIC NO**: 18/MHS01/013

**DEPARTMENT:** ANATOMY

**COURSE:** BIO-INFORMATICS

**1a.**

* Sensitivity
* Selectivity
* Speed

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. These correct hits are  
considered “true positives” in the database searching exercise.

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.”

Speed: which is the time it takes to get results  
from database searches

**1b.**

BLAST (basic local alignment search tool)[[2]](https://en.wikipedia.org/wiki/BLAST_(biotechnology)#cite_note-KarlinObit.NYT-2) is an algorithm and program for comparing primary biological sequence information, such as the amino-acid sequences of [proteins](https://en.wikipedia.org/wiki/Protein) or the [nucleotides](https://en.wikipedia.org/wiki/Nucleotide) of [DNA](https://en.wikipedia.org/wiki/DNA_sequence) and/or [RNA](https://en.wikipedia.org/wiki/RNA) sequences. A BLAST search enables a researcher to compare a subject protein or nucleotide sequence (called a query) with a library or [database](https://en.wikipedia.org/wiki/Database) of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

**2a.**

The PAM matrices are based on scoring all amino acid positions in related sequences, whereas the BLOSUM matrices are based on substitutions and conserved positions in blocks, which represent the most-alike common regions in related sequences. PAM matrices are used to score alignments between closely related protein sequences. BLOSUM matrices are used to score alignments between evolutionarily divergent protein sequences.

**2b.**

Heuristic Database Searching: which is a computational strategy to find an empirical or near optimal solution by using rules of thumb. Essentially, this type of algorithms take shortcuts by reducing the search space according to some criteria. It is often used because of the need for obtaining results within a realistic time frame without significantly sacrificing the accuracy of the computational output.

**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. It is the 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 similarity: is a measure of an empirical relationship between sequences. Its common objective is establishing the likelihood for sequence homology i.e chance that sequences has evolved from a common ancestor. It is the direct result of observation from the sequence  
alignment.

Sequence identity: is the amount of characters which match exactly between two different sequences. A similarity score is therefore aimed to approximate the evolutionary distance between a  
pair of nucleotide or protein sequences.

**3b**

1. Dot matrix method; Dynamic programming method; word method
2. Pairwise Sequence Alignment is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences (protein or nucleic acid).

**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. 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.

**4b**.

**i.** Sequence homology 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. While Sequence similarity: is a measure of an empirical relationship between sequences. Its common objective is establishing the likelihood for sequence homology i.e chance that sequences has evolved from a common ancestor.

**ii.** 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.