K M G G P G C

Digital Note Book

Name:

Class : M.Sc. Zoology

Topic → BLAST

(Basic local alignment search tool)

Jyoti

- BLAST stands for Basic local Alignment Search tool.

- BLAST is a program which uses specific scoring matrices (like PAM or BLOSSUM) for performing sequence - similarity searches against a variety of sequences data bases, to give us high - scoring ungapped segments among related sequences.

- Complex - requires multiple steps and many parameters.

- The BLAST algorithm is fast, accurate & web - accessible.

- Is relatively faster than other sequences similarity search tools.

- Provides us with ability to perform analy-sis by different type of programs.

Types of BLAST Algorithm :-

Program	Input	Query search	Database
blastn	DNA	1 →	DNA
blastp	Protein	1 →	Protein
blastx	DNA	6 →	Protein
tblastn	Protein	6 →	DNA
tblastx	DNA	36 →	DNA

Types of BLAST Algorithm :-

blastn Compares a DNA query sequence against a DNA database, allowing for gaps

blastp Compares a protein query sequence against a protein database, allowing for gaps

blastx Compares a DNA query sequence translated into six reading frames against a protein database, allowing for gaps

tblastn	Compares a protein query sequence against a DNA database translated into six reading frames, allowing for gaps
tblastx	Compares a DNA query sequence translated into six reading frames against a DNA database translated into six reading frames. tblastx doesn't allow for gaps

Special Algorithm of BLAST :-

MEGABLAST	:- for comparison of large sets of long DNA sequences
RPS-BLAST	:- Conserved Domain Detection
BLAST 2 Sequences	:- for performing pair-wise alignments for 2 chosen sequences
Genomic BLAST	:- for alignments against select human, microbial or malarial genomes

PSI-BLAST :- Construct a multiple alignment from matches

PHI-BLAST :- Specify a pattern that hits must match

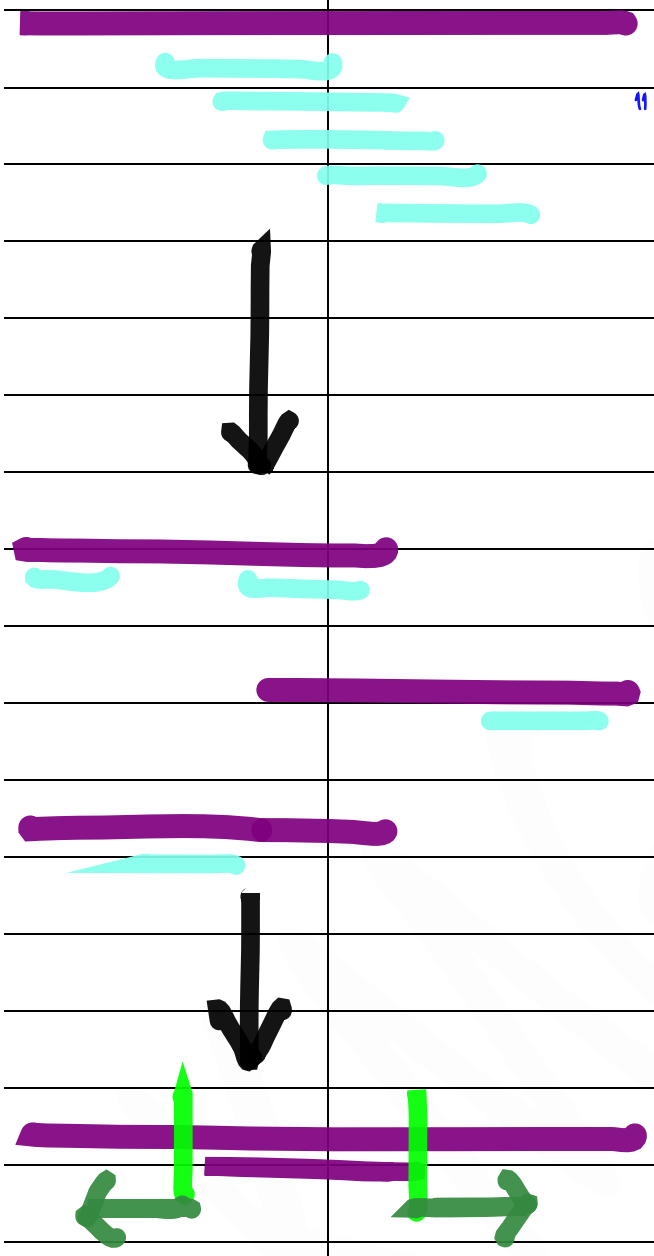
HOW BLAST WORKS ?

BLAST works in following two steps :-

1. BLAST first searches for short regions of a given length (W) called "words" (or substrings) that score at least "T" when compared to the query sequence that align with sequences in the database **(target sequenes)** using a substitution matrix.

2. For every pair of sequences (query & target) that have a word or words in common, BLAST extends the alignment in both directions to find alignments that score greater (are more similar) than a certain scoring pairs or HSPs ; the maximal scoring HSPs are called MSPs.

PICTORIAL REPRESENTATION :-



"Words" (Subsequences of the query seq.)

Query words are compared to the database (target sequences) and exact matches identified

For each word match, alignment is extended in both directions to find alignments that score greater than some threshold (maximal segment pairs, or MSPs)

There are various questions which a BLAST can handle which commonly arises

Now as to see whether this protein has any similarity between other organism we perform a BLAST to understand it's importance.

Input/Output of BLAST

- **Input :-**

  - Query sequence Q
  - Database of sequences DB
  - Minimal score S

- **Output :-**

  - Sequences from **DB** (seq), such that Q and **S** seq have score > **S**

1. Word searching method :-

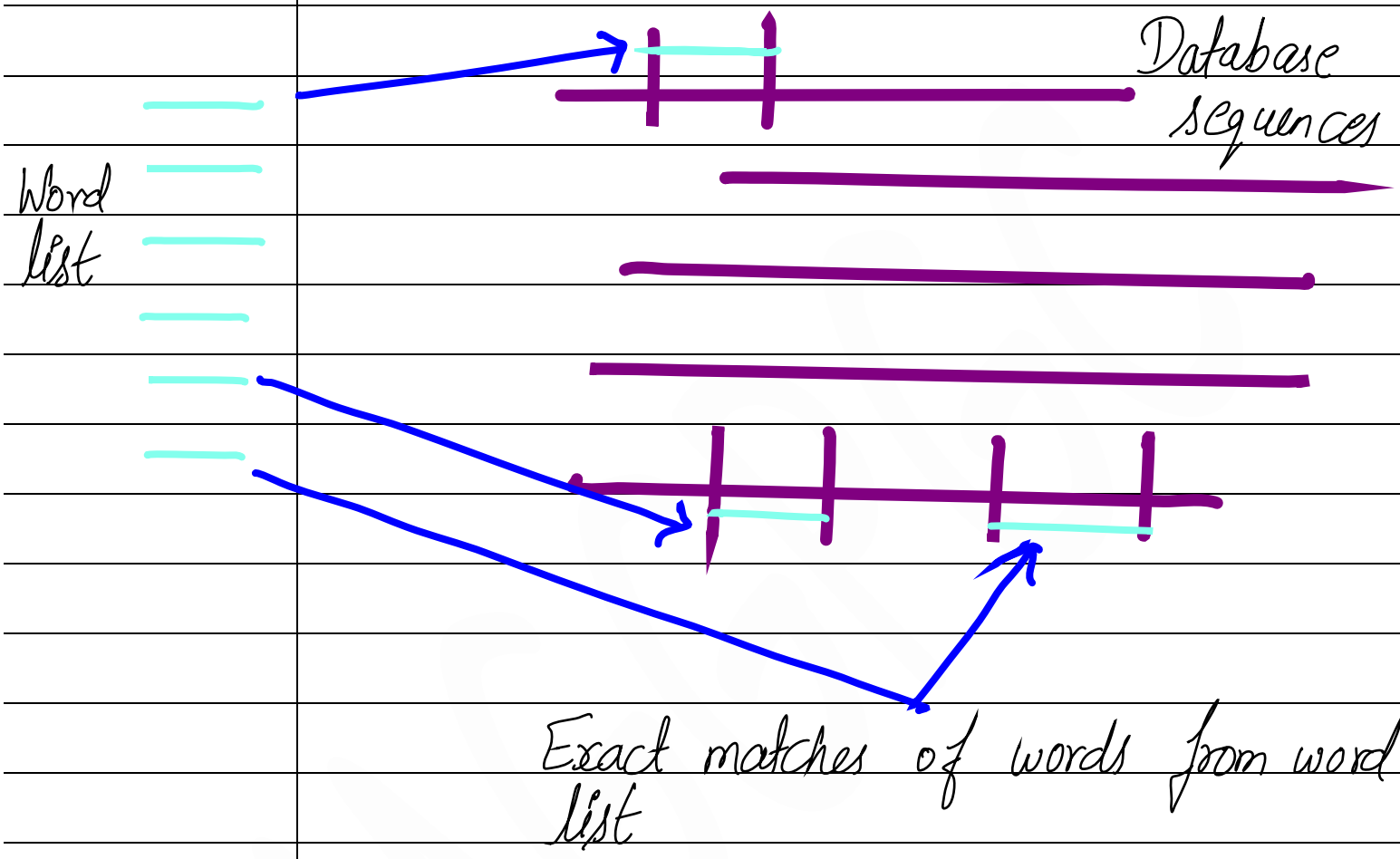
1. For the query, find the list of high scoring words of length

Query sequence of length L

Maximum of **L-W+1** words (typically W=3 for proteins)

For each word from the query sequence find the list of words that will score atleast **T** when scored using a pair-score matrix (e.g. PAM 25)

2. Compare the word list to the database & identify exact matches



2. Identification of exact word :-

MEAAVKEEISVEDEAVDK

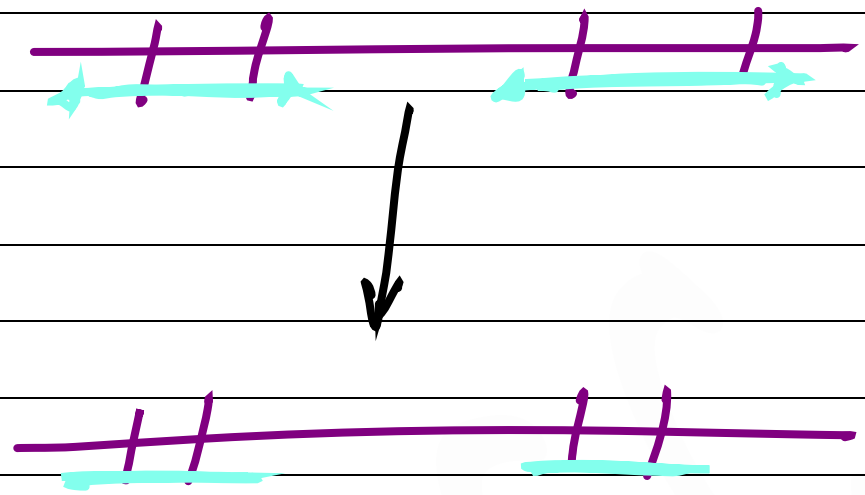
MEA
EAA
AAV
AVK
VKE
KEE
EEI
EIS
ISV

Break query into words

Break data base sequences into word

3. Maximum segment pair alignment : Extend hits one base at a time

For each word match, extend the alignment in both directions to find alignments that score greater than a threshold of value S



Maximal segment Pairs (MSPs)

Uses of BLAST :-

- **Identifying species :-**

With the use of BLAST, identify a 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 with in the sequence of interest.

DNA Mapping :-

Comparison :-