BLAST AND FASTA

INTRODUCTION

The number of DNA and protein sequences in public databases is very large.
Searching a database involves aligning the query sequence to each sequence in the database, to find significant local alignment.
BLAST and FASTA are two similarity searching programs that identify homologous DNA sequences and proteins based on the excess sequence similarity.
They provide facilities for comparing DNA and proteins sequences with the existing DNA and protein databases.
They are two major heuristic algorithms for performing database searches.

FASTA and BLAST are the software tools used in bioinformatics. Both BLAST and FASTA use a heuristic word method for fast pairwise sequence alignment.
It works by finding short stretches of identical or nearly identical letters in two sequences. These short strings of characters are called words.
The basic assumption is that two related sequences must have at least one word in common.
By first identifying word matches, a longer alignment can be obtained by extending similarity regions from the words.
Once regions of high sequence similarity are found, adjacent high-scoring regions can be joined into a full alignment.
The main difference between BLAST and FASTA is that BLAST is mostly involved in finding of ungapped, locally optimal sequence alignments whereas FASTA is involved in finding similarities between less similar sequences.

BLAST

The BLAST program was developed by Stephen Altschul of NCBI in 1990 and has since become one of the most popular programs for sequence analysis.
BLAST uses heuristics to align a query sequence with all sequences in a database.
The objective is to find high-scoring ungapped segments among related sequences.
The existence of such segments above a given threshold indicates pairwise similarity beyond random chance, which helps to discriminate related sequences from unrelated sequences in a database.
BLAST is popular as a bioinformatics tool due to its ability to identify regions of local similarity between two sequences quickly. BLAST calculates an expectation value, which estimates the number of matches between two sequences. It uses the local alignment of sequences.

VARIANTS OF BLAST

BLAST-N: compares nucleotide sequence with nucleotide sequences
BLAST-P: compares protein sequences with protein sequences
BLAST-X: Compares nucleotide sequences against the protein sequences
tBLAST-N: compares the protein sequences against the six frame translations of nucleotide sequences
tBLAST-X: Compares the six frame translations of nucleotide sequence against the six frame translations of protein sequences.

FASTA

FASTA stands for fast-all" or "FastA".
It was the first database similarity search tool developed, preceding the development of BLAST.
FASTA is another sequence alignment tool which is used to search similarities between sequences of DNA and proteins.
FASTA uses a "hashing" strategy to find matches for a short stretch of identical residues with a length of k. The string of residues is known as ktuples or ktups, which are equivalent to words in BLAST, but are normally shorter than the words.
Typically, a ktup is composed of two residues for protein sequences and six residues for DNA sequences.
The query sequence is thus broken down into sequence patterns or words known as k-tuples and the target sequences are searched for these k-tuples in order to find the similarities between the two.
FASTA is a fine tool for similarity searches.
These methods are not guaranteed to find the optimal alignment or true homologs, but are 50–100 times faster than dynamic programming.