

What is BLAST?

- Basic BLAST search
 - What is BLAST?
 - The framework of BLAST
 - Different BLAST programs
 - BLAST databases you can search
 - Where can I run BLAST?

What is BLAST?

- BLAST stands for
 - Basic Local Alignment Search Tool
- Why BLAST is popular?
 - Good balance of sensitivity and speed
 - Reliable
 - Flexible
- Produce local alignments: short significant stretches of similarity, irrespective of where they are in the sequence

BLAST Programs

The most common BLAST search include five programs:

| Program | Database (Subject) | Query |
|---------|--------------------|---------------|
| BLASTN | Nucleotide | Nucleotide |
| BLASTP | Protein | Protein |
| BLASTX | Protein | Nt. → Protein |
| TBLASTN | Nt. → Protein | Protein |
| TBLASTX | Nt. → Protein | Nt. → Protein |

BLASTN

- BLASTN
 - The query is a nucleotide sequence
 - The database is a nucleotide database
 - No conversion is done on the query or database
- DNA :: DNA homology
 - Mapping oligos to a genome
 - Annotating genomic DNA with transcriptome data from ESTs and RNA-Seq
 - Annotating untranslated regions

BLASTP

- BLASTP
 - The query is an amino acid sequence
 - The database is an amino acid database
 - No conversion is done on the query or database
- Protein :: Protein homology
 - Protein function exploration
 - Novel gene → make parameters more sensitive

BLASTX

- BLASTX
 - The query is a nucleotide sequence
 - The database is an amino acid database
 - All six reading frames are translated on the query and used to search the database
- Coding nucleotide seq :: Protein homology
 - Gene finding in genomic DNA
 - Annotating ESTs and transcripts assembled from RNA-Seq data

TBLASTN

- TBLASTN
 - The query is an amino sequence
 - The database is a nucleotide database
 - All six frames are translated in the database and searched with the protein sequence
- Protein :: Coding nucleotide DB homology
 - Mapping a protein to a genome
 - Mining ESTs and RNA-Seq data for protein similarities

TBLASTX

- TBLASTX
 - The query is a nucleotide sequence
 - The database is a nucleotide database
 - All six frames are translated on the query and on the database
- Coding :: Coding homology
 - Searching distantly-related species
 - Sensitive but expensive

BLAST output

1. List of sequences with scores
 - Raw score
 - Higher is better
 - Depends on aligned length
 - Expect Value (E-value)
 - Smaller is better
 - Independent of length and database size
2. List of alignments

The Databases (1)

- GenBank NR (protein and nucleotide versions)
 - Non-redundant large databases (compile and remove duplicates)
 - Anyone can submit, you can call your sequence anything
 - Low quality; names can be meaningless
- Transcriptome Shotgun Assembly (TSA) Database
 - Transcripts assembled from overlapping ESTs and RNA-Seq reads
 - Most of the sequences have no annotations

The Databases (2)

- UniProt/Swiss-Prot
 - Curated from literature
 - **REAL** proteins; **REAL** functions; small;
- Genomic databases
 - Human, Mouse, Drosophila, Arabidopsis, etc.
 - NCBI, species-specific web pages

Where Can I run BLAST?

1. NCBI BLAST web service
 - <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
2. EBI BLAST web service
 - <https://www.ebi.ac.uk/Tools/sss/ncbiblast/>
3. FlyBase BLAST
 - <http://flybase.org/blast/>
 - Drosophila and other insects