

---

NCBI Webinars

## NCBI BLAST Services

July 2011

---

---

---

---

---

---

---

---

### Using BLAST

---

NCBI Webinars

- Basics of using NCBI BLAST
- Using the new Interface
  - Improved organism and filter options
- New Services
  - Primer BLAST
  - Align 2 Sequences Integration
  - COBALT – protein multiple alignment
  - Global Alignment tool

---

---

---

---

---

---

---

---

### Basic Local Alignment Search Tool

---

NCBI Webinars

- Widely used similarity search tool
- Heuristic approach based on Smith Waterman algorithm
- Finds best local alignments
- Provides statistical significance
- All combinations (DNA/Protein) query and database.
  - DNA vs DNA
  - DNA translation vs Protein
  - Protein vs Protein
  - Protein vs DNA translation
  - DNA translation vs DNA translation
- [www](#), standalone, and network client

---

---

---

---

---

---

---

---

## Local Alignment Statistics

High scores of local alignments between two random sequences follow the Extreme Value Distribution

**Expect Value**  
**E = number of database hits you expect to find by chance**

$E = Kmne^{-\lambda S}$  or  $E = mn2^{-S'}$

K = scale for search space  
 $\lambda$  = scale for scoring system  
 $S'$  = bitscore =  $(\lambda S - \ln K) / \ln 2$

(applies to ungapped alignments)

NCBI Webinars

---

---

---

---

---

---

---

---

---

---

## BLAST and BLAST-like programs

- **Traditional BLAST (formerly blastall)** nucleotide, protein, translations
  - **blastn** nucleotide query vs. nucleotide database
  - **blastp** protein query vs. protein database
  - **blastx** nucleotide query vs. protein database
  - **tblastn** protein query vs. translated nucleotide database
  - **tblastx** translated query vs. translated database
- **Megablast** nucleotide only
  - **Contiguous megablast**
    - Nearly identical sequences
  - **Discontiguous megablast**
    - Cross-species comparison
- **Position Specific BLAST Programs** protein only
  - **Position Specific Iterative BLAST (PSI-BLAST)**
    - Automatically generates a position specific score matrix (PSSM)
  - **Reverse PSI-BLAST (RPS-BLAST)**
    - Searches a database of PSI-BLAST PSSMs

NCBI Webinars

---

---

---

---

---

---

---

---

---

---

## Web Access: www.ncbi.nlm.nih.gov

NCBI Webinars

---

---

---

---

---

---

---

---

---

---



## Nucleotide Databases: Human and Mouse

Megablast, blastn service

- Human and mouse genomic and transcript now default
- Separate sections in output for mRNA and genomic
- Direct links to Map Viewer for genomic sequences

NCBI Webinars

---

---

---

---

---

---

---

---

## Nucleotide Databases: Traditional

NCBI Webinars

---

---

---

---

---

---

---

---

## Nucleotide Databases: Traditional

Databases are mostly non-overlapping

- |  |   |
|--|---|
| <ul style="list-style-type: none"> <li>• <b>nr (nt)</b> <ul style="list-style-type: none"> <li>- Traditional GenBank</li> <li>- NM_ and XM_ RefSeqs                             <ul style="list-style-type: none"> <li>• <a href="#">refseq_rna</a></li> </ul> </li> </ul> </li> <li>• <b>NCBI Genomes</b> <ul style="list-style-type: none"> <li>- NC_ RefSeqs</li> <li>- GenBank Chromosomes</li> </ul> </li> <li>• <b>dbest</b> <ul style="list-style-type: none"> <li>- EST Division                             <ul style="list-style-type: none"> <li>• <a href="#">non-human, non-mouse ests</a></li> </ul> </li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>• <b>htgs</b> <ul style="list-style-type: none"> <li>- HTG division</li> </ul> </li> <li>• <b>gss</b> <ul style="list-style-type: none"> <li>- GSS division</li> </ul> </li> <li>• <b>wgs</b> <ul style="list-style-type: none"> <li>- whole genome shotgun contigs</li> </ul> </li> <li>• <b>env_nt</b> <ul style="list-style-type: none"> <li>- environmental samples</li> </ul> </li> </ul> |
|--|---|

NCBI Webinars

---

---

---

---

---

---

---

---



## BLAST extensions and improvements

NCBI Webinars

- PrimerBlast – primer designer / specificity checker
- COBALT – Protein Multiple Alignment tool
- Integration / expansion of BLAST 2 Sequences
- Global Alignment (Needleman-Wunsch)

---

---

---

---

---

---

---

---

## Specialized BLAST Pages

NCBI Webinars

### Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- ▶ Make specific primers with [Primer-BLAST](#)
- ▶ Search [trace archives](#)
- ▶ Find [conserved domains](#) in your sequence (cds)
- ▶ Find sequences with similar [conserved domain architecture](#) (cdart)
- ▶ Search sequences that have [gene expression profiles](#) (GEO)
- ▶ Search [immunoglobulins](#) (IgBLAST)
- ▶ Search for [SNPs](#) (snp)
- ▶ Screen sequence for [vector contamination](#) (vecscreen)
- ▶ [Align](#) two (or more) sequences using BLAST (bl2seq)
- ▶ Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay
- ▶ Search SRA [transcript and genomic libraries](#)
- ▶ Constraint Based Protein [Multiple Alignment Tool](#)
- ▶ Needleman-Wunsch [Global Sequence Alignment Tool](#)
- ▶ Search [RefSeqGene](#)
- ▶ Search [WGS sequences](#) grouped by organism

---

---

---

---

---

---

---

---

## Hands-on Practice: Goals

NCBI Webinars

- Select the appropriate BLAST database and program to get the most relevant results
- Use taxonomic / organism limit on the BLAST database to obtain specific results
- Map a sequence onto an assembled genome using BLAST
- Tune BLAST parameters for specific kinds of searches
- Design PCR primers for a specific template and check specificity
- Directly compare two sets of protein sequences and generate a multiple-alignment using link to COBALT

---

---

---

---

---

---

---

---