BLAST Quick Start

blast-help@ncbi.nlm.nih.gov

Algorithm Basics
- Introduction
- Words & extensions

Search Basics
- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples
Basic Local Alignment Search Tool

Compare protein or nucleic acid sequences to protein or nucleic acid databases

- in NCBI databases
- in local databases (standalone BLAST)
- to a single protein or nucleotide sequence
  (BLAST 2 Sequences, or pairwise BLAST)

BLAST Web Searches, 2007

Total number of searches done

200,000/day
Basic Local Alignment Search Tool

- local alignments; isolated regions of similarity
- fast and sensitive
- breaks the query sequence into “words”
- word matches to database sequences are extended in both directions

Global vs Local Alignment

Seq 1
Seq 2
↓ Global alignment

Seq 1
Seq 2
↓ Local alignment
Global vs Local Alignment

Seq1: WHEREISWALTERNOW (16aa)
Seq2: HEWASHEREBUTNOWISHERE (21aa)

Global

<table>
<thead>
<tr>
<th>Seq1: 1</th>
<th>W--HEREISWALTERNOW 16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
</tr>
<tr>
<td>Seq2: 1</td>
<td>HEWASHEREBUTNOWISHERE 21</td>
</tr>
</tbody>
</table>

Local

<table>
<thead>
<tr>
<th>Seq1: 1</th>
<th>W--HERE 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
</tr>
<tr>
<td>Seq2: 3</td>
<td>WASHERE 9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Seq1: 1</th>
<th>W--HERE 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
</tr>
<tr>
<td>Seq2: 15</td>
<td>WISHERE 21</td>
</tr>
</tbody>
</table>

How BLAST Works

1. Make lookup table of “words” for query
2. Scan database for hits
3. Extend alignment both directions
Nucleotide Words

Make a lookup table based on the word size.

\[ \text{11-mer} \]
\[ \text{ATGCTGCTAGTGACGTA} \]
\[ \text{ATGCTGCTAGT} \]
\[ \text{TGCTGCTAGTC} \]
\[ \text{GCTGCTAGTCG} \]
\[ \ldots \]

Protein Words

\[ \text{AIEKCYTGCTLAQEADDTA} \]
\[ \text{AIE} \]
\[ \text{IEK} \]
\[ \text{LEK}, \text{IDK}, \text{IQK}, \text{IER}, \text{IDR}, \text{etc} \]
\[ \text{EKC} \]
\[ \text{KCY} \]
\[ \text{CYT} \]
\[ \ldots \]

Lookup table, including neighborhood words, is based on word size, score matrix, and threshold.
**Word Hits & Extensions**

Nucleotide: one exact match
ATGCTGCTAGTCGATGACGTAGCTA
⇔ GCTGCTAGTCG ⇔

Protein: two matches within 40 residues
AIEKCYTGCTLAQEADDTA
⇔ IDK ⇔
⇔ EAD ⇔

**BLAST**

**Search Basics**

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples
**BLAST Programs**

What is your goal?

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>blastn</td>
<td>nucleotide X nucleotide</td>
</tr>
<tr>
<td>blastp</td>
<td>protein X protein</td>
</tr>
</tbody>
</table>

6 frame, translated nucleotide searches

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>blastx</td>
<td>nucleotide X protein</td>
</tr>
<tr>
<td>tblastx</td>
<td>nucleotide X nucleotide</td>
</tr>
<tr>
<td>tblastn</td>
<td>protein X nucleotide</td>
</tr>
</tbody>
</table>

**More BLAST Programs**

<table>
<thead>
<tr>
<th>Program</th>
<th>Description</th>
<th>Word Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>MegaBLAST</td>
<td>- batch nucleotide queries</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>- very similar sequences</td>
<td></td>
</tr>
<tr>
<td>Discontiguous MegaBLAST</td>
<td>- batch nucleotide queries</td>
<td>11 matches</td>
</tr>
<tr>
<td></td>
<td>- divergent sequences</td>
<td>out of 18</td>
</tr>
</tbody>
</table>
**Search Basics**

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

**Nucleotide BLAST Databases**

- **nr (nt)**
  - Traditional GenBank Divisions
  - NM_ and XM_ RefSeqs
- **refseq_rna**
  - NM_, XM_, NR_
- **refseq_genomic**
  - NC_, NT_, NG_
- **est**
  - EST Division
- **htgs**
  - HTG division
- **dbsts**
  - STS Division
- **chromosome**
  - NC genomic records
- **gss**
  - GSS division
- **pat**
  - PAT Division
- **wgs**
  - wgs entries from traditional divisions
- **pdb**
  - Nucleotide sequences from structures
- **env_nt**
  - environmental samples
**New Nucleotide Databases**

NCBI BLAST

BLASTN programs search nucleotide databases using a nucleotide query.

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

Or, upload file

Job Title

Choose Search Set

Database

- Human genomic + transcript
- Mouse genomic + transcript
- Others (e.g.,)
- Human genomic plus transcript

Entrez Query

optional

Choose Search Set

Other Databases

- Nucleotide collection (nt)
- Reference mRNA sequences (refseq_rna)
- Reference genomic sequences (refseq_genomic)
- Expressed sequence tags (est)
- Non-human, non-mouse ESTs (est_others)
- Genomic survey sequences (gss)
- High-throughput genomic sequences (HTGS)
- Patent sequence (pat)
- Protein Data Bank (pdb)
- Human ALU repeat elements (alu_repeats)
- Sequence tagged sites (dts)
- Whole-genome shotgun reads (wgs)
- Environmental samples (env)

Human genomic plus transcript
Protein BLAST Databases

Protein
• nr
  traditional GenBank records
• refseq = NP_, XP_
• swissprot
• pdb
• pat
• env_nr

BLAST Databases: Genome-specific

Genomes
• Human, mouse, rat, chimp, cow, pig, dog, sheep, cat
• Chicken, puffer fish, zebrafish
• Fly, honey bee, other insects
• Microbes, environmental samples
• Plants, nematodes
• Fungi, protozoa, other eukaryotes
NCBI Minicourses

Genome-specific: Map Viewer

Search - Select Group or Organism - for

Vertebrates

Mammals

BLAST Bos taurus (cow) Build 3.1
BLAST Bos taurus (cow) Build 2.1
BLAST Canis familiaris (dog)
BLAST Felis catus (cat)
BLAST Homo sapiens (human) Build 36
BLAST Homo sapiens (Human) Build 35
BLAST Macaca mulatta (chimpanzee)
BLAST Mus musculus (mouse) Build 38
BLAST Mus musculus (mouse) Build 35

Other Vertebrates

BLAST Danio rerio (zebrafish)

Plants

BLAST Arabidopsis thaliana
BLAST Arabidopsis thaliana
BLAST Allium cepa (onion)
BLAST Arabidopsis thaliana (thale cress)
BLAST Arabidopsis thaliana
BLAST Beta vulgaris (beet)
BLAST Brassica juncea (Indian mustard)
BLAST Brassica napus (oilseed rape)
BLAST Brassica rapa (black mustard)
BLAST Brassica oleracea
BLAST Brassica rapa (field mustard)
BLAST Capsicum annuum
BLAST Eragrostis tef (teff)
BLAST Glycine max (soybean)
BLAST Hordeum vulgare (barley)
BLAST Lotus japonicus (lotus)
BLAST Manihot esculenta (cassava)

Save your searches

Basic BLAST

Search a nucleotide database using a nucleotide query
Search protein database using a protein query
Search protein database using a translated nucleotide query
Search translated nucleotide database using a protein query
Search translated nucleotide database using a translated nucleotide query

Specialized BLAST

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

BLAST 2.2.14 new release

BLAST release 2.2.14 offers a universal binary for Mac OS X and improved performance on some platforms.

Tip of the Day

Checking PCR Primers

A frequent use of nucleotide-nucleotide BLAST is to check the specificity of oligonucleotides for hybridization in PCR. The goal is usually to make sure that the primers will give a unique product from the target genome or DNA and rule out Bacterial BLAST.
NCBI Minicourses

BLAST

Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

Setting up a BLAST Search

Identifier or sequence

Title

Select database

NP_032294:homeobox B5 [Mus musculus]
Conserved Domain search run for protein queries
**Default Output**

**Graphic Overview**

Distribution of 104 Blast Hits on the Query Sequence

![Color key for alignment scores](image)

**Summary Table**

<table>
<thead>
<tr>
<th>Distance tree of results</th>
<th>Score (Bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
<td>NP_033204.11</td>
<td>homeobox B</td>
</tr>
<tr>
<td>ref</td>
<td>XP_075165.11</td>
<td>PREDICTED: similar to homeobox protein Box-B...</td>
</tr>
<tr>
<td>ref</td>
<td>NP_046176.11</td>
<td>PREDICTED: similar to Homebox protein Box-B...</td>
</tr>
<tr>
<td>ref</td>
<td>NP_002195.11</td>
<td>homeobox B5 [Homo sapiens] ref</td>
</tr>
<tr>
<td>ref</td>
<td>XP_00167149.11</td>
<td>PREDICTED: similar to homeobox protein [Homo sapiens]</td>
</tr>
<tr>
<td>db</td>
<td>BAB28059.1</td>
<td>unnamed protein product [Mus musculus]</td>
</tr>
<tr>
<td>ref</td>
<td>NP_00120428.1</td>
<td>homeobox B5 [Gallus gallus] ref</td>
</tr>
<tr>
<td>db</td>
<td>BA005558.11</td>
<td>Hoxb-5 [Gallus gallus]</td>
</tr>
<tr>
<td>emb</td>
<td>CA902225.1</td>
<td>homeodomain protein [Danio rerio] ref</td>
</tr>
<tr>
<td>ref</td>
<td>NP_071176.11</td>
<td>homeobox B5 [Danio rerio] ref</td>
</tr>
<tr>
<td>db</td>
<td>BA005561.11</td>
<td>Hoxb-5 [Felis catus sylvestris]</td>
</tr>
<tr>
<td>ref</td>
<td>XP_001349297.11</td>
<td>PREDICTED: similar to homeobox protein B...</td>
</tr>
<tr>
<td>ref</td>
<td>NP_071412.11</td>
<td>homeobox B5 [Danio rerio] ref</td>
</tr>
<tr>
<td>gb</td>
<td>ABQ10594.1</td>
<td>homeobox protein [Danio rerio]</td>
</tr>
<tr>
<td>gb</td>
<td>ABD74565.11</td>
<td>homeodomain protein [Megalobrama amblycephala]</td>
</tr>
</tbody>
</table>
BLAST Output – tree view

This tree was produced using BLAST pairwise alignments. more...

Alignment View - pairwise

Score = 827 bits (567), Expect = 5e-55, Method: Composition-based stats.

Query 1  NYSYFVMSFSYRPFQGVLYNYGSSSS--SLSYVDP...NMNTHSGYVHVDIVGSLYTVRS  59
  Subject 1 NYSYFVMSFSYRPFQGVLYNYGSSS...SLSYVDP...NMNTHSGYVHVDIVGSLYTVRS  60

Query 2  54445GAVGESSFSAARHPSATEFQPQATCASLSFLSTI...CMTMHTGSDGTS...SSESD  122
  Subject 2 54445GAVGESSFSAARHPSATEFQPQATCASLSFLSTI...CMTMHTGSDGTS...SSESD  119

Query 3  QG+T+  S+KIEDEASAGK  HEA+ + +  Q + + +  G  2  177
  Subject 3  QG+T+  S+KIEDEASAGK  HEA+ + +  Q + + +  G  2  177
New Output View

<table>
<thead>
<tr>
<th>Entrez ID</th>
<th>Accession</th>
<th>Description</th>
<th>expect</th>
<th>bit score</th>
<th>total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NM_001046</td>
<td>Homo sapiens</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>NM_001047</td>
<td>Homo sapiens</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

NCBI Minicourses

Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

...con't

BLAST
Where do Scores Come From?

### Scoring Matrices

**Nucleotide search: identity matrix**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>T</td>
<td>-3</td>
<td>1</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>-3</td>
<td>1</td>
<td>-3</td>
</tr>
<tr>
<td>G</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>1</td>
</tr>
</tbody>
</table>

**CAGTTAGCAAGCTTGGTCA**

|||||
raw score = 19 - 9* = 10*

**CAGTTAGCAAGCTTGGTCA**

* Ignores gap costs
Scoring Matrices

Protein search: substitution matrix

- Percent Accepted Mutation (PAM)
- Blocks Substitution Matrix (BLOSUM)

PAM Matrices

- global alignments of closely related proteins
- PAM1 calculated from sequences with no more than 1% divergence
- Other PAM matrices are extrapolated from PAM1
- Examples: PAM30 and PAM70
BLOSUM Matrices

- local alignments
- all matrices based on observed alignments; not extrapolated
- BLOSUM 62 calculated from sequences with no more than 62% identity
- Examples: BLOSUM45, BLOSUM62 and BLOSUM80
- BLOSUM62: very good for detecting weak protein similarities
- BLOSUM62 is the default matrix in BLAST
**Substitution Matrices**

**A Provisional Table**

<table>
<thead>
<tr>
<th>Query Length</th>
<th>Substitution Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;35</td>
<td>PAM30</td>
</tr>
<tr>
<td>35-50</td>
<td>PAM70</td>
</tr>
<tr>
<td>50-85</td>
<td>BLOSUM80</td>
</tr>
<tr>
<td>&gt;85</td>
<td>BLOSUM62</td>
</tr>
</tbody>
</table>

**BLOSUM 80**

**BLOSUM 62**

**BLOSUM 45**

**PAM 1**

**PAM 120**

**PAM 250**

Less divergent ← ----------------→ More divergent

**Where do Expect Values Come From?**

**Sequences producing significant alignments:**

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query coverage</th>
<th>E value</th>
<th>Max ident</th>
<th>Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcripts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM_001250.1</td>
<td>Homo sapiens CDC20 cell division</td>
<td>2876</td>
<td>2876</td>
<td>95%</td>
<td>0.0</td>
<td>97%</td>
<td>U E O M</td>
</tr>
<tr>
<td>XM_933013.2</td>
<td>PREDICTED: Homo sapiens chr 1</td>
<td>40.1</td>
<td>40.1</td>
<td>1%</td>
<td>8.2</td>
<td>100%</td>
<td>O N</td>
</tr>
<tr>
<td>XM_944935.2</td>
<td>PREDICTED: Homo sapiens chr 1</td>
<td>40.1</td>
<td>40.1</td>
<td>1%</td>
<td>8.2</td>
<td>100%</td>
<td>O N</td>
</tr>
<tr>
<td>Genomic sequences [show first]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NW_921351</td>
<td>Homo sapiens chromosome 1</td>
<td>428</td>
<td>3010</td>
<td>95%</td>
<td>9e-117</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>NT_022072.1</td>
<td>Homo sapiens chromosome 1</td>
<td>428</td>
<td>2002</td>
<td>95%</td>
<td>9e-117</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>NT_023932.17</td>
<td>Homo sapiens chromosome 9</td>
<td>2629</td>
<td>2673</td>
<td>94%</td>
<td>0.0</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>NW_921807.1</td>
<td>Homo sapiens chromosome 9</td>
<td>2801</td>
<td>2845</td>
<td>94%</td>
<td>0.0</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>NW_921807.1</td>
<td>Homo sapiens chromosome 3</td>
<td>44.1</td>
<td>44.1</td>
<td>1%</td>
<td>0.53</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
**Expect Value**

$E = \text{number of database hits you expect to find by chance, } \geq S$

$$E = K m n e^{-\lambda S} \quad \text{or} \quad E = m n 2^{-\lambda S'}$$

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

$E$ is dependent on $m \times n$ (search space)

More info: The Statistics of Sequence Similarity Scores

---

Short query =

<table>
<thead>
<tr>
<th>low score</th>
<th>high Expect</th>
</tr>
</thead>
</table>

Score = 30.2 bits (15), Expect = 19
Identities = 15/15 (100%), Gaps = 0/15 (0%)
Strand = Plus/Plus

Query 4

```
GTAGCAAGCTTGCA
```

Sbjct 40347983

```
GTAGCAAGCTTGCA
```

raw score = 19 - 9 = 10

More info: The Statistics of Sequence Similarity Scores
Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

Initial Options

Limit results

- Descriptions: 100
- Graphical overview: 100
- Alignments: 100

Enter query:

Organism: *Use common name, binomial, taxid, or group name. Only 20 top taxa will be shown*

Expect min: [ ]

Expect Max: [ ]

- biomol mrna[properties]
- all[Filter] NOT mammalia[organism]
- chimpanzee[organism]
- viridiplantae[organism]
Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

Default Output

BLAST 2.2.16 (Mar 25, 2007)

Reference:
ENST0000000057014 (ENST0000000057014)

Format:

Alignment View:
- Aligner:
- HTML:
- Advanced View:

Display:
- Graphical Overview:
- List:
- Sequence Retrieval:
- PSI-BLAST:

Limit results:
- Description:
- Graphical overview:
- Alignments:

Options:
- Type sequence name, functional, label, organism name. Only the top hits will be shown.
- Enter organism name or species searches will be suggested.

Format:
- PSI-BLAST with inclusion threshold:

Job Title: NP_032294, human box E5 (Mus musculus)

Show results in a new window
BLAST Output: Alignments & Filter

NCBI Minicourses

BLAST Output: Alignments & Filter

NCBI Minicourses

Alignment View Options

Show: Alignment △ as HTML △ □ Advance
Alignment View: Parwise △
Display: Parwise △ Parwise with dots for identities △ Query-anchored with dots for identities △ Query-anchored with letters for identities △ Fast query-anchored with dots for identities △ Fast query-anchored with letters for identities △ Hit Table △
Limit results: △

low complexity sequence filtered
Alignment View: Flat Query-Anchored with Dots for Identities

<table>
<thead>
<tr>
<th>Query</th>
<th>MSSYFVNSF-8GRYN-GPDyqllnygqsgslsgsy-ROG---AAMH----TG8Y---G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1708354                                                  46</td>
</tr>
<tr>
<td></td>
<td>74265740                                                 46</td>
</tr>
<tr>
<td></td>
<td>84573655                                                 46</td>
</tr>
<tr>
<td></td>
<td>1139533                                                  46</td>
</tr>
<tr>
<td></td>
<td>12848720                                                 7</td>
</tr>
<tr>
<td></td>
<td>76644466                                                 46</td>
</tr>
<tr>
<td></td>
<td>57225883                                                 46</td>
</tr>
<tr>
<td></td>
<td>62556079                                                 46</td>
</tr>
<tr>
<td></td>
<td>22526089                                                 46</td>
</tr>
<tr>
<td></td>
<td>28629649                                                 7</td>
</tr>
<tr>
<td></td>
<td>54267140                                                 47</td>
</tr>
<tr>
<td></td>
<td>62837                                                   47</td>
</tr>
<tr>
<td></td>
<td>37949656                                                 46</td>
</tr>
<tr>
<td></td>
<td>22316369                                                 46</td>
</tr>
<tr>
<td></td>
<td>3922074                                                  46</td>
</tr>
<tr>
<td></td>
<td>74265740                                                 46</td>
</tr>
<tr>
<td></td>
<td>34894445                                                 46</td>
</tr>
<tr>
<td></td>
<td>30256652                                                 46</td>
</tr>
<tr>
<td></td>
<td>3827933                                                  46</td>
</tr>
<tr>
<td></td>
<td>12679958                                                 46</td>
</tr>
</tbody>
</table>

Alignment View: Hit Table

* BLASTP 2.2.15 (Nov-27-2005)*
* Query: gi|6602551|ref|0_02294|1.1| home box B5 [Mus musculus]*
* Database: or*  
* Fields: query id, subject ids, % identity, % positives, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score*  
* 514 hits found*  
* gi|6602551|ref|0_02294|1.1| gi|18709564|ref|0_09791|1005_HOM_GSGS|gi|6602551|ref|0_02294|1.1|*  
* gi|6602551|ref|0_02294|1.1| gi|74227470|dbj|HAA21695.1|gi|73885452|gb|AAT06687.1|gi|72|*  
* gi|6602551|ref|0_02294|1.1| gi|84573655|dbj|1BAAT5826.1|gi|73966032|gb|X641176.2|*  
* gi|6602551|ref|0_02294|1.1| gi|1118933|db|AAS25558.3|gi|4904569|ref|0_645228.1|gi|406|*  
* gi|6602551|ref|0_02294|1.1| gi|10547510|dbj|1BAAD5955.1|99.57|106.50|240|1|*  
* gi|6602551|ref|0_02294|1.1| gi|10547510|dbj|1BAAD5955.1|99.57|106.50|240|1|*  
* gi|6602551|ref|0_02294|1.1| gi|57225883|gb|J0186161.1|gi|17077845|ref|0_0010261|gi|21|*  
* gi|6602551|ref|0_02294|1.1| gi|62556079|dbj|1BAAD9836.1|77.68|84.35|240|53|*  
* gi|6602551|ref|0_02294|1.1| gi|62556079|dbj|1BAAD9836.1|77.68|84.35|240|53|*  
* gi|6602551|ref|0_02294|1.1| gi|62556079|dbj|1BAAD9836.1|77.68|84.35|240|53|*  
* gi|6602551|ref|0_02294|1.1| gi|62556079|dbj|1BAAD9836.1|77.68|84.35|240|53|*  
* gi|6602551|ref|0_02294|1.1| gi|62556079|dbj|1BAAD9836.1|77.68|84.35|240|53|*  

# Fields: query id, subject ids, % identity, % positives, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score
Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

Examples

- Protein searches more sensitive than nucleotide searches
  - redundancy of the genetic code
- blastn, megablast best when searching within same organism
**BLASTN vs BLASTP**

**BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.14 [Apr-09-2006]**

<table>
<thead>
<tr>
<th>Matrix</th>
<th>BLOSUM62</th>
<th>gap open</th>
<th>11</th>
<th>gap extension</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_dropoff</td>
<td>50</td>
<td>expect: 10,000</td>
<td>wordsize</td>
<td>3</td>
<td>Filter</td>
</tr>
<tr>
<td>Masking character option</td>
<td>X for protein, x for nucleotides</td>
<td>Masking color option</td>
<td>Black</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Bitscore and expect value are calculated based on the size of the nr database.

Score = 49.8 bits (114), Expect = 7e-04
Identities = 50/181 (27%), Positives = 80/181 (44%), Gaps = 27/181 (14%)

**BLAST is a shortcut . . .**

An alignment BLAST cannot make:

```
1  GAATATATGAAGACCAAGATTGCAGTCCTGCTGGCCTGAACCAGCTATTTGCTGTTG
|| || || || |  || || ||   ||  |  ||| |||||| | | || | ||| |
1  GAGTGTACGATGAGCCCGAGTGTAGCAGTGAAGATCTGGACCACGGTGTACTCGTTGTCG
61  GTTACGGAACCGAGAATGGTAAAGACTACTGGATCATTAAGAACTCCTGGGGAGCCAGTT
| || ||     ||  ||| ||  | |||||| || | ||||||  |||||  |     |
61  GCTATGGTGTTAAGGGTGGGAAGAAGTACTGGCTCGTCAAGAACAGCTGGGCTGAATCCT
121  GGGGTGAACAAGGTTATTTCAGGCTTGCTCGTGGTAAAAAC
|||| || ||||| ||  ||    |  | ||||  || |||
121  GGGGAGACCAAGGCTACATCCTTATGTCCCGTGACAACAAC
```

Reason:
no contiguous exact match of 7 bp.
An Alignment BLAST Can Make

**Score** = 290 bits (741), **Expect** = 7e-77
**Identities** = 147/331 (44%), **Positives** = 206/331 (61%), **Gaps** = 8/331 (2%)
**Frame** = +3

---

**BLASTN vs TBLASTX**

*Anopheles gambiae* mitochondrion, complete genome.
**ACCESSION** NC_002084 **GI**:5834911

*Homo sapiens* mitochondrion, complete genome.
**ACCESSION** NC_001807 **GI**:17981852

---
Hands-On