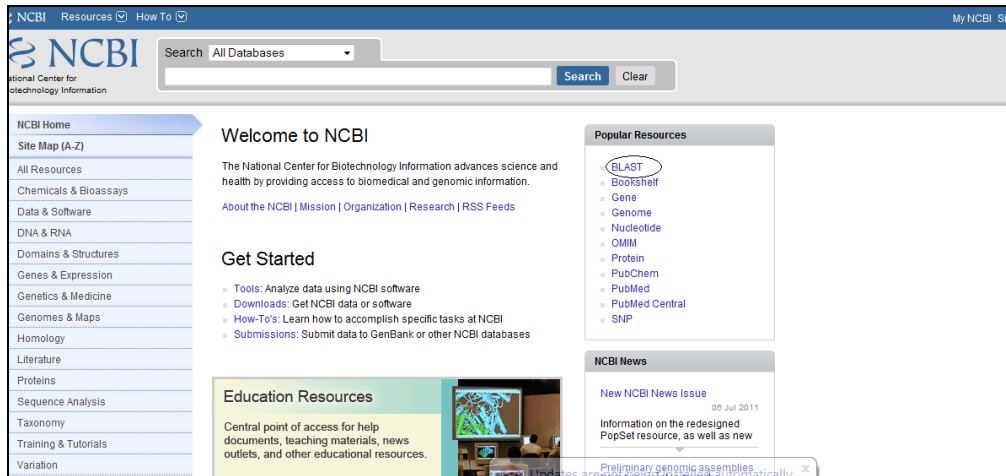


HANDOUT BLAST

Link: <http://www.ncbi.nlm.nih.gov/>

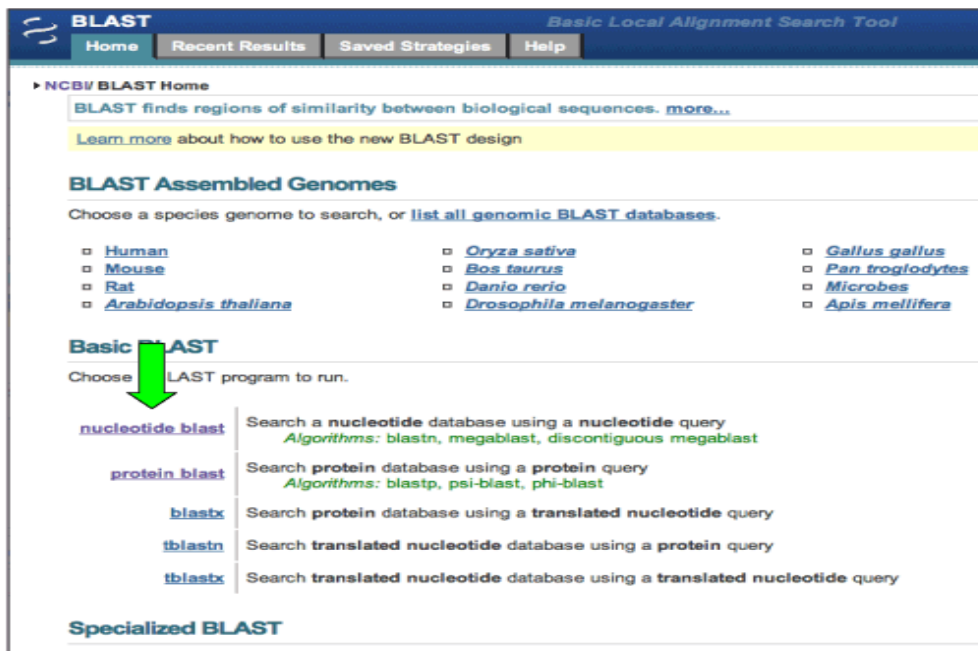
Step 1

Home page of NCBI- Select BLAST from Popular Resources



Step 2

Click on the nucleotide Blast or Protein Blast



Step 3

Paste your sequence in the tab “Enter Query Sequence”

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST/blastn suite: BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence Clear

Query subrange From To

3AATCGGAGAGTGTTGGTCACCTTAGCGCGGGGAACATCGAGCA
ATTCCAAGATGACCATTTTGACGACAAGCAGGTTTCAGGG
3AAGTTGTCGAGAAGCTCAGCGTAGCCGCCACTGGTGAGCCA
3TTCCGTCAGACCAGATCGACGAAAGGCTTAGAAAACATCACAA

Or, upload file Choose File no file selected

Job Title Enter a descriptive title for your BLAST search

Choose Search Set

Database ☐ Human genomic + transcript ☐ Mouse genomic + transcript ☒ Others (nr etc.):
Nucleotide collection (nr/nt)

Entrez Query Optional Enter an Entrez query to limit search

Program Selection

Optimize for ☐ Highly similar sequences (megablast) ☒ More dissimilar sequences (discontiguous megablast) ☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search database Test/gpipe/9606/allcontig

Show results in a new window

Algorithm parameters

For the BLAST activities:
1. Change the database to "Others"
2. Select "nucleotide collection" from the pull-down menu.
3. Change the program selection to optimize for more dissimilar sequences.

Paste your sequence here

Step 4

Observe the Result page: wait few seconds to view the result

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST/blastn suite/Formatting Results - 1S2NU91W016 [\[Formatting options\]](#)

Job Title: [gi|44955887|ref|NM_203378.1| Homo sapiens...](#)

Request ID	1S2NU91W016
Status	Searching
Submitted at	Tue Jul 12 04:50:02 2011
Current time	Tue Jul 12 04:50:28 2011
Time since submission	00:00:26

This page will be automatically updated in 1 seconds

Step 5

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

► NCB/ BLAST/ blastn/ Formatting Results - 9AH5C86C015 [\[Reformat these Results\]](#)

Job Title: Nucleotide sequence (2110 letters)

BLASTN 2.2.17 (Jun-24-2007)

RID: 9AH5C86C015

Database: All GenBank+EMBL+DBJ+POB sequences (but no EST, STS, GSS, environmental samples or phase 0, 1 or 2 HTGS sequences)
 5,453,285 sequences; 21,092,363,288 total letters

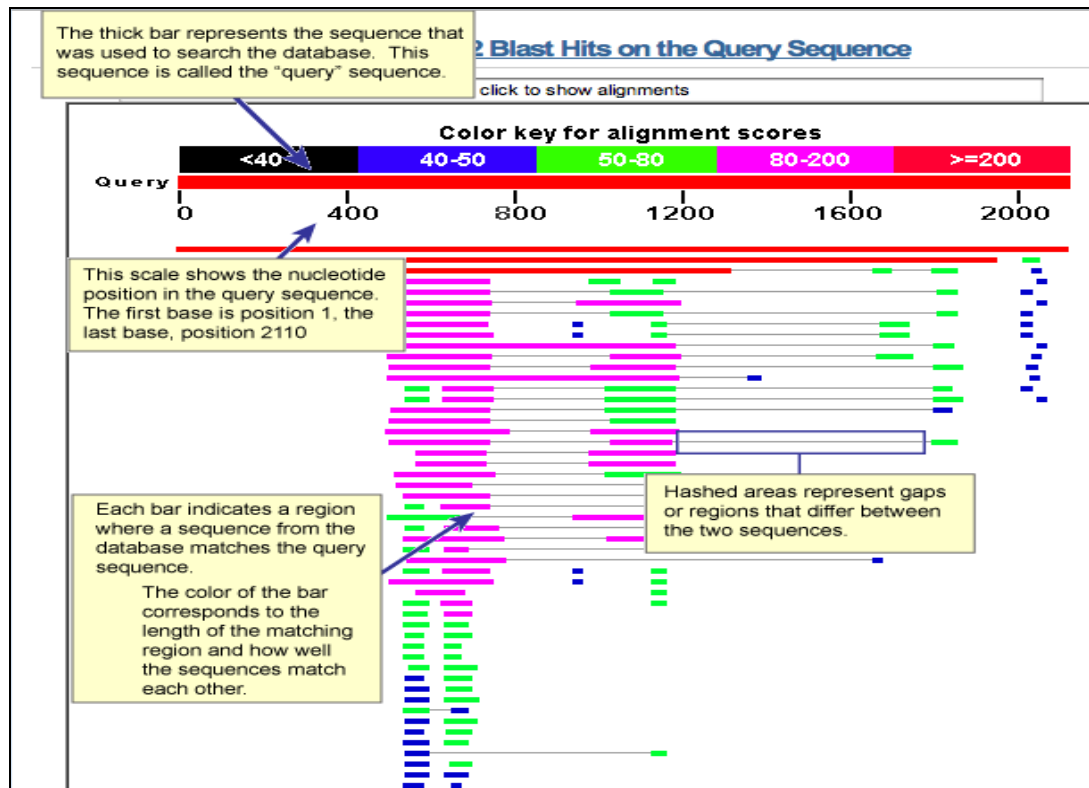
If you have any problems or questions, please refer to the [BLAST FAQs](#) or [Taxonomy reports](#)

Query= Length=2110

Annotations:

- The number of sequences in the database.
- The number of letters in the database.
- This is the length of your query sequence.

Step 6



Step 7

Scroll down to view the details of the result obtained

A description of the sequence

The total score includes scores from non-contiguous portions of the subject sequence that match the query.

The Max % ident corresponds to the match to a subject sequence with the highest percentage of identical bases.

Legend for links to other resources: U UniGene E EMBL G Gene S Structure M Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
X15893.1	Tarantula mRNA for hemocyanin subunit a	4057	4057	100%	0.0	100%	
AF547807.1	Nephila inaurata madagascariensis mRNA for hemocyanin	662	662	79%	0.0	73%	
AF547811.1	Nephila inaurata madagascariensis mRNA for hemocyanin	202	319	43%	5e-48	88%	
AF547809.1	Nephila inaurata madagascariensis mRNA for hemocyanin	185	241	14%	7e-43	83%	
AF302908.1	Cupiennius salei mRNA for hemocyanin subunit 5' (hc-5')	175	296	19%	6e-40	84%	
AF277452.1	Eurytelma californicum mRNA for hemocyanin subunit g	171	278	22%	8e-39	79%	

The Accession number is linked to the GenBank record.

A score that indicates how well the sequences match. For nucleotide sequences, this is approximately equal to twice the length of the matching region.

The Max score is linked to data that show where the sequences match.

The query coverage corresponds to the fraction of the query sequence that matches a subject sequence.

Step 8

Learn more about the E value

The E value is equal to the number of sequences that you would expect to find in a database composed entirely of random sequences.

Two important parameters that influence the E value are:

- The number of sequences in the database (database size).
- The length of the query sequence.

There is a greater chance of finding a match in a larger database. And the chance of finding a match for a short sequence is greater than the chance of finding a match to a longer sequence.

E
Value

0.0
1e-21
7e-20
4e-18
2e-14
1e-09
4e-09
1e-08
1e-08
1e-08
0.014
0.014
0.014
0.014
0.014
0.22
0.22
0.86
0.86
0.86
0.86
3.4
3.4

In this example, the E value equals

$$1 \times 10^{-21}$$

The letter "e" is used to show that -21 is the exponent. You would "expect" to find very few random sequences in this database that match the query sequence this well.

This sequence has an E value of 3.4. A database of random sequences would be likely to contain 3.4 sequences that matched the query sequence equally well.

Step 9

>emb|X16893.1|ECHEMSUA Tarantula mRNA for hemocyanin subunit a
Length = 2110
Score = 4183 bits (2110), Expect = 0.0
Identities = 2110/2110 (100%)
Strand = Plus / Plus

This is the score assigned by BLAST. In general, the higher the score, the better the match between the query sequence and the sequence found in the database.

In this example, 100% of the nucleotides in a 2110 base stretch of the query sequence are identical to a 2110 nucleotide region in the sequence obtained from GenBank.

```

Query: 1  gaatoggagaggtgttggtcacttaggc
          |||
Sbjct: 1  ggcgggggaacatcgagcaattccaagatgaccatt 60
          |||
Query: 121 ggtgagccagtttoctgcagaccagatogacgaaaggottagaacatcacacottaggt 180
          |||
Sbjct: 121 ggtgagccagtttoctgcagaccagatogacgaaaggottagaacatcacacottaggt 180
          |||
Query: 181 cccaatgaattcttctctgtgtttacccagaccacttggaacaagccaagagaggtotac 240
          |||
Sbjct: 181 cccaatgaattcttctctgtgtttacccagaccacttggaacaagccaagagaggtotac 240
          |||
Query: 241 gaagttttctgccaatgctgctaaacttcgatgacttcgtcagcttgccaaagcaagcgoga 300
          |||
Sbjct: 241 gaagttttctgccaatgctgctaaacttcgatgacttcgtcagcttgccaaagcaagcgoga 300
          |||
Query: 301 agcttcatgaactccactctgtttgcocttctctgcagaagttgcocttctcatogggaa 360
          |||
Sbjct: 301 agcttcatgaactccactctgtttgcocttctctgcagaagttgcocttctcatogggaa 360
          |||
Query: 361 gactgcgagggcgctcatctgacccctccagaaagttttcgctgacagattcatccccc 420
          |||
Sbjct: 361 gactgcgagggcgctcatctgacccctccagaaagttttcgctgacagattcatccccc 420
  
```

Step 10

View the detailed description of the BLAST result page

Sequences producing significant alignments:

	Score (bits)	E Value
gi 9266 emb X16893.1 ECHEMSUA Tarantula mRNA for hemocyanin...	4183	0.0
gi 28569681 emb AJ547807.1 NIN547807 Nephila inaurata madag...	202	2e-48
gi 15027023 emb AJ307908.1 CSA307908 Cupiennius salei mRNA ...	121	6e-24

Look at more alignments

>gi|28569681|emb|AJ547807.1|NIN547807 Nephila inaurata madagascariensis mRNA
(gene)
Length = 2089
Score = 202 bits (102), Expect = 2e-48
Identities = 153/170 (90%)
Strand = Plus / Plus

These numbers show the nucleotide positions where the alignments begin and end.

```

Query: 1027 cacaactggggacatgttatgatggcttacattcatgactcgtgatggcagattcaggaa 1086
Sbjct: 1013 cacaactggggacacgtaatgatggcctacatccatgactcgtgatggcagattcaggaa 1072

Query: 1087 acaccaggtgtcatgactgacacagccacaagtccttagggatccaatctctacagatac 1146
Sbjct: 1073 actccaggtgtcatgaccgatactgctacaagtccttagggatccaatctctacagatac 1132

Query: 1147 cacagattcatcgacaacggttttccaagaatacaagaaaactctgccagt 1196
Sbjct: 1133 cacagattcatcgataacggttttccaggaatacaagaaaactcttccagt 1182
  
```

blastn did not detect matching sequences in the dashed regions

Query: 467 ctggagatgagagcgaca 527
Sbjct: 453 ctggagatgagagcgatg 527

Query: 527 aatacaaaactgcctact

Nephila sequence: 0 453 700 1013 1182

Page shows the alignment with the best matching sequences

Step 11

Click the Accession numbers for detailed view

Legend for links to other resources: [U](#) UniGene [F](#) GDB [G](#) Gene [S](#) Structure [M](#) Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
X16893.1	Tarantula mRNA for hemocyanin subunit a	4057	4057	100%	0.0	100%	
AF010717.1	Nephila inaurata madagascariensis mRNA for hemocyanin	662	662	79%	0.0	73%	
AF010718.1	Nephila inaurata madagascariensis mRNA for hemocyanin	202	319	43%	5e-48	88%	
AF010719.1	Nephila inaurata madagascariensis mRNA for hemocyanin	185	241	14%	7e-43	83%	
AF010720.1	Cupiennius salei mRNA for hemocyanin subunit 5' (hc-5')	175	298	19%	6e-40	84%	
AF010721.1	Eurytelma californicum mRNA for hemocyanin subunit g	171	278	22%	8e-39	79%	

Look at the GenBank record.

Step 12

Search [Nucleotide](#) [Protein](#) [Genome](#) [Structure](#) for _____ Go Clear

Accession number [Limits](#) [Preview/Index](#) [History](#) [Clipboard](#)

Display [GenBank](#) [Save](#) [Text](#) [Add to Clipboard](#)

1 : X16893.1 Tarantula mRNA for...[gi:9266]

LOCUS ECHMSUA 2110 bp mRNA INV 12-SEP-1993

DEFINITION Tarantula mRNA for hemocyanin subunit a.

ACCESSION X16893

VERSION X16893.1 GI:9266

KEYWORDS hemocyanin; hemocyanin subunit a.

SOURCE Aphonopelma sp.

ORGANISM [Aphonopelma sp.](#) [Link to taxonomy database](#)

REFERENCE 1 (bases 1 to 2110)

AUTHORS Voit, R.

TITLE Direct Submission

JOURNAL Submitted (12-OCT-1989) Voit R., Zoologisches Institut, Universitaet Muenchen, Luisenstrasse 14, D-8000 Muenchen 2, FRG

REFERENCE 2 (bases 1 to 2110)

AUTHORS Voit, R. and Feldmaier-Fuchs, G.

TITLE Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs encoding the tarantula hemocyanin subunits a and e

JOURNAL J. Biol. Chem. 265 (32), 19447-19452 (1990)

MEDLINE 91060944

COMMENT Data kindly reviewed (26-MAR-1990) by Voit R.

FEATURES

source

1..2110

/organism="Aphonopelma sp."

/db_xref="taxon:29932"

/clone="lambda-K1"

/clone_lib="lambda-gt10"

52..1947

/note="hemocyanin subunit a (AA 1-631)"

/codon_start=1

in_id="CAA34771.1"

ef="GI:9267"

lation="MTILHDKQVQALKLFEKLSYAATGEPVPADQIDERLRNITTLGP

FYEDHLEQAKRYVEVECHAANEDNEVSLAKQARSFHNSTLEAFESAVALIHR

CDS

CDS is short for "coding sequence". The region of DNA between nucleotides 52 and 1947 is predicted to code for the hemocyanin subunit a protein.

The source of the sequence

The amino acid sequence for the predicted protein

Step 13

Click on the journal medline number to view the Actual function of the protein.