

Tutorial 4 – BLAST Searching the CHO Genome

The CHO BLAST server can be accessed in two ways: from the CHO-K1 genome search page and by clicking on the BLAST button on the home page. A link to the NCBI BLAST web server is also provided on the CHO-K1 genome search page .

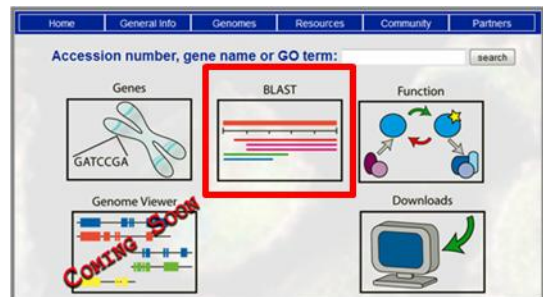
BLAST searching the CHO-K1 Genome

1) The CHO BLAST server can be accessed in two ways: from the CHO-K1 genome search page and by clicking on the BLAST button on the home page.

Click on the BLAST server link on the CHO-K1
genome search page

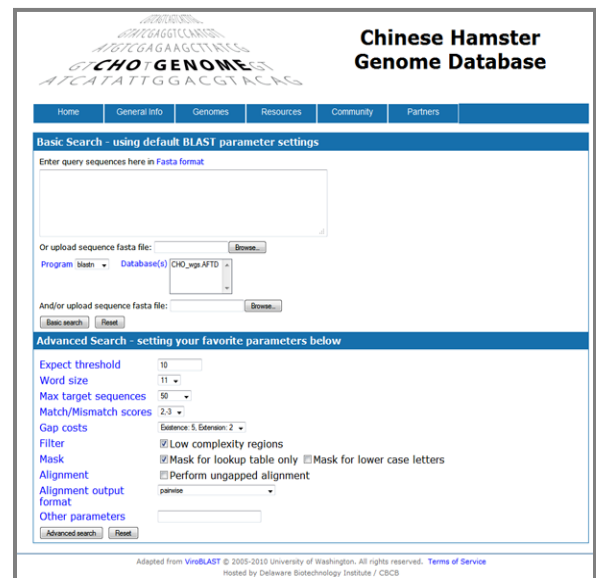


Select the BLAST icon on the
home page



2) The CHO BLAST page allows for basic and advanced
BLAST searches against the CHO-K1 genome database.

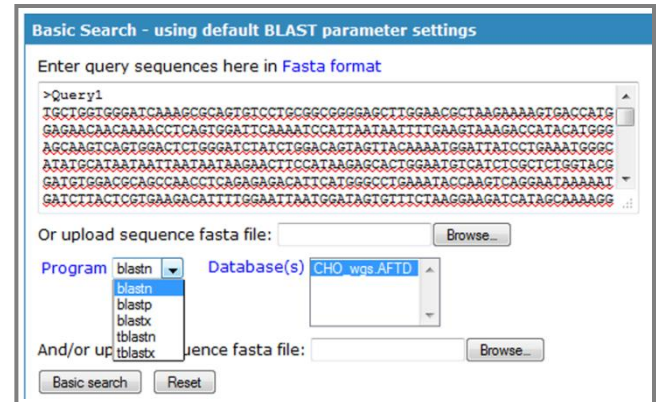
Clicking on the [Program](#) and [Database\(s\)](#) links will provide
a brief description of the programs and databases
currently available for BLAST searching.



3) Query sequences in FASTA format can be pasted into the search box at the top of the page or by uploading sequences in a FASTA file. Multiple query sequences may be entered at once.

The BLAST program to use and database to search are then selected from the currently available options.

To BLAST the CHO-K1 genome database, select the “blastn” program and the “CHO_wgs_AFTD” database.



Basic Search - using default BLAST parameter settings

Enter query sequences here in Fasta format

>Query1
TGCTGGTGGGATCAAGCCGACGTGCTGCGCGGGGAGCTTGGACGCTAAGAAAAGTGACCATG
GAGAACACAAACCTCAGTGGATCAAAATCCATTAAATATTTGAGTAAGACCATACATGGG
AGCAAGTCAGTGGACTCAGGATCTATCTGGACAGTAGTTACAAATGGATTATCTGAAATGGG
ATATGCAATATTAATAAAGAACTTCCATAGAGCAGCTGGAATGTCATCTGCTCTGGTACG
GATCTGGACGCGCAACCTCAGAGAGACATTCATGGGCTGAAATACCAAGTCAGGAATAAAAT
GATCTTACTCGTGAAGACATTTGGAAATTAATGATAGTGTCTTAAAGGAGATCATAGCAAAAGG

Or upload sequence fasta file:

Program: **blastn** Database(s): **CHO_wgs_AFTD**

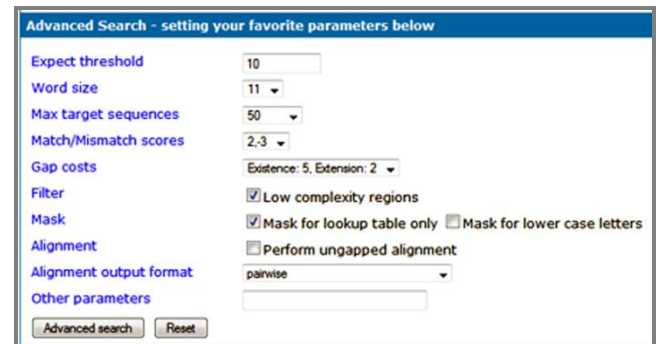
And/or upload sequence fasta file:

To perform a basic BLAST search, click the “Basic search” button once all information is entered.

4) In the Advanced Search sections, the BLAST parameters can be varied to perform an advanced BLAST search.

Clicking on the highlighted blue terms (such as [Expect threshold](#)) will provide a brief description of each the advanced search parameters that can be varied.

To perform an advanced BLAST search, click the “Advanced search” button once all information is entered.



Advanced Search - setting your favorite parameters below

Expect threshold 10

Word size 11

Max target sequences 50

Match/Mismatch scores 2:3

Gap costs Existence: 5, Extension: 2

Filter ☒ Low complexity regions

Mask ☒ Mask for lookup table only ☐ Mask for lower case letters

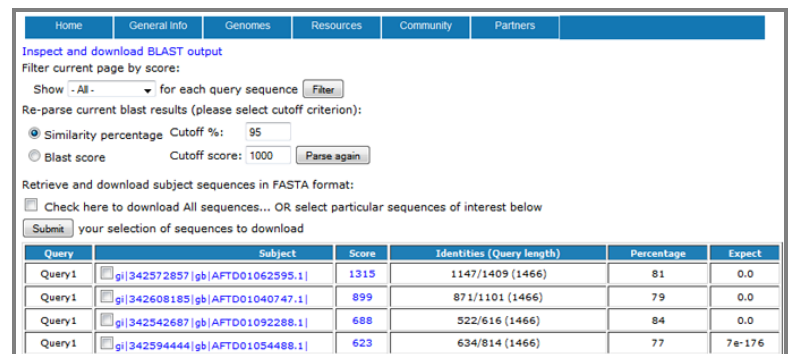
Alignment ☐ Perform ungapped alignment

Alignment output format pairwise

Other parameters

5) The results of the BLAST alignment are summarized in a table with the query sequence name, the subject sequence name, the bit score, the identity percentage, and the *E*-value.

The results can be filtered by score (showing only top 1, 5, or 10 alignments), by similarity percentage or by BLAST score.



Home General Info Genomes Resources Community Partners

Inspect and download BLAST output

Filter current page by score:

Show: All for each query sequence

Re-parse current blast results (please select cutoff criterion):

☒ Similarity percentage Cutoff %: 95

☐ Blast score Cutoff score: 1000

Retrieve and download subject sequences in FASTA format:

☐ Check here to download All sequences... OR select particular sequences of interest below

your selection of sequences to download

Query	Subject	Score	Identities (Query length)	Percentage	Expect
Query1	gi 342572857 gb AFTD01062595.1	1315	1147/1409 (1466)	81	0.0
Query1	gi 342608185 gb AFTD01040747.1	899	871/1101 (1466)	79	0.0
Query1	gi 342542687 gb AFTD01092288.1	688	522/616 (1466)	84	0.0
Query1	gi 342594444 gb AFTD01054488.1	623	634/814 (1466)	77	7e-176

After entering the filter parameter, click either the “Filter” of the “Parse again” buttons to refresh the results table.

To inspect all the BLAST pair-wise alignments, click on [Inspect and download BLAST output](#).

To view the pair-wise alignment for a specific alignment, click on the value in the Score column for any alignment (such as [1315](#)).

To view the GenBank entry for each subject sequence, click on the sequence name in the Subject column (such as [gi|342572857|gb|AFTD01062595.1|](#)).

Inspect and download BLAST output

Filter current page by score:

Show: All ▾ for each query sequence Filter

Re-parse current blast results (please select cutoff criterion):

☒ Similarity percentage Cutoff %: 95
☐ Blast score Cutoff score: 1000

Retrieve and download subject sequences in FASTA format:

☐ Check here to download All sequences... OR select particular sequences of interest below

your selection of sequences to download

Query	Subject	Score	Identities (Query length)	Percentage	Expect
Query1	<input type="checkbox"/> gi 342572857 gb AFDT01062595.1	1315	1147/1409 (1466)	81	0.0
Query1	<input type="checkbox"/> gi 342608185 gb AFDT01040747.1	899	871/1101 (1466)	79	0.0
Query1	<input type="checkbox"/> gi 342542687 gb AFDT01093288.1	688	522/616 (1466)	84	0.0
Query1	<input type="checkbox"/> gi 342594444 gb AFDT01054489.1	623	634/814 (1466)	77	7e-176

```
> Query1 on gi|342572857|gb|AFDT01062595.1 Cricetus griseus scaffold1353_40, whole genome shotgun sequence
Length=74348

Score = 1315 bits (1458), Expect = 0.0
Identities = 1147/1409 (81%), Gaps = 46/1409 (3%)
Strand=Plus/Minus

Query 47      CGCTAAGAAAAGTGACCATGGGAACAACAAACCCTCAGTGGATTCAAAATCCATTAAATA   106
              || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 47864    CGTTAAGAAAAGCGATCATGGA-----CAGAACACCTCAGTAGCTCCAAATCCATTAAAA   47790
```

Cricetus griseus scaffold1353_40, whole genome shotgun sequence

GenBank: AFDT01062595.1

[FASTA](#) [Graphics](#)

[Go! :](#) ☒

LOCUS AFDT01062595 74348 bp DNA linear ROD 03-AUG-2011

DEFINITION Cricetus griseus scaffold1353_40, whole genome shotgun sequence.

ACCESSION AFDT01062595 [AFDT01000000](#)

VERSION AFDT01062595.1 GI:342572857

DBLINK Project: [69991](#)

KEYWORDS WGS.

SOURCE Cricetus griseus (Chinese hamster)

ORGANISM [Cricetus griseus](#)

6) To download the subject sequences, click the “Check here to download all sequences” box or select individual subject sequences from the table below and then hit the “Submit” button. This will download a text file of the selected sequences in FASTA format in a new window.

If the BLAST search was against a genomic database (such as the CHO WGS contigs), the subject sequences will be the WGS genomic contig sequences.

Home
General Info
Genomes
Resources
Community
Partners

Inspect and download BLAST output
Filter current page by score:
Show: All for each query sequence
Re-parse current blast results (please select cutoff criterion):
☒ Similarity percentage Cutoff %: 95
☐ Blast score Cutoff score: 1000

Retrieval and download options in FASTA format:
☐ Check here to download All sequences... OR select particular sequences of interest below
 your selection of sequences to download

Query	Subject	Score	Identical (Query length)	Percentage	Expect
Query1	gi 342572857 gb AFPTD01062955.1	1315	1147/1409 (1466)	81	0.0
Query1	gi 342608185 gb AFPTD010407.1	899	817/1101 (1466)	79	0.0
Query1	gi 342542687 gb AFPTD01092288.1	688	632/616 (1466)	84	0.0
Query1	gi 342594444 gb AFPTD01054488.1	623	528/814 (1466)	77	7e-176

```

>gi|342542687|gb|AFPTD01092288.1| Cricetulus griseus scaffold2357_15,
whole genome shotgun sequence
TGTGTGTGTAATGTTGTAATCTGTGAGTATGAATATTTTCATAACAGTTAGAAAGGAGATCAAGAAAGGTTA
ACATTAGACATCATGAGGAAAGCAGGATATAGGCCAAAGGACAAAGAGTCATGGAAGAGGATATAACAT
ATCTTTTCATATATGCCCAAGGTGACCACTCTCCAGGGAATGGTGCTATCCACAGTGGGCTGGACCTCC
CACATCAATTCAACAACAGACACTCCCTACAGACATGGGAGACAACTCTGATCCAGCAATTCTGTGAT
CTGAGAATCCCTTCTCAGGAGACCTTAGGCTGTGTCAAAACAACATGTAAGTGACTAGGACAGTCCCC
TTCAATGCACACTCCGATGACCAAGAGTTTCTTTTATTGTTAATGATGACAAAATACACGACAAAAGCA

```

6) Multiple query sequences can be BLAST searched at once. The results from all query sequences will be displayed in a single results table. To inspect all the BLAST pair-wise alignments, click on [Inspect and download BLAST output](#). Links provided at the top of the page can be used to quickly navigate to the alignment results for an individual query sequence.

Basic Search - using default BLAST parameter settings

Enter query sequences here in [Fasta format](#)

>Query1
TGCTGGTGGGATCAAGGCGCAGTGTCTGCGGCGGGAGCTTGGACGCTAAGAAAAGTGACCATG
GAGAACACAAAACCTCAGTGGATCAGAAATCAATTAATTAATTTGAAGTAAAGACCATACATGGG
AGCAAGTCAGTGGACTCTGGGATCTATCTGACAGTAGTACAAAATGGAT

>Query2
ACAAATTAATCACTCCCTTGCTACTAGACATCCCGGAGCTCCTTTTTCTATATTTTATGTT

Or upload sequence fasta file:

Program **blastn** Database(s) **CHO_wgs.AFTD**



Home	General Info	Genomes	Resources	Community	Partners
Inspect and download BLAST output					
Filter current page by score:					
Show All for each query sequence <input type="button" value="Filter"/>					
Re-parse current blast results (please select cutoff criterion):					
<input checked="" type="radio"/> Similarity percentage Cutoff %: <input type="text" value="95"/>					
<input type="radio"/> Blast score Cutoff score: <input type="text" value="1000"/> <input type="button" value="Parse again"/>					
Retrieve and download subject sequences in FASTA format:					
<input type="checkbox"/> Check here to download All sequences... OR select particular sequences of interest below					
<input type="button" value="Submit"/> your selection of sequences to download					
Query	Subject	Score	Identities (Query length)	Percentage	Expect
Query1	gi 342372857 gb AFTD01062595.1 	1315	1147/1409 (1466)	81	0.0
Query1	gi 342608185 gb AFTD01040747.1 	899	871/1101 (1466)	79	0.0
Query1	gi 342542687 gb AFTD01092288.1 	668	522/616 (1466)	84	0.0
Query1	gi 342594444 gb AFTD01054488.1 	623	634/814 (1466)	77	7e-176



BLAST Results		
 Query1 Query2 Query3 		
Database: CHO_wgs.AFTD		
265,786 sequences; 2,318,115,958 total letters		
Query= Query1		
Length=183		
Sequences producing significant alignments:		
	Score	E
	(Bits)	Value
gi 342542687 gb AFTD01092288.1 Cricetulus griseus scaffold2357...	196	3e-48
gi 342608185 gb AFTD01040747.1 Cricetulus griseus scaffold669...	170	1e-40
gi 342572857 gb AFTD01062595.1 Cricetulus griseus scaffold1353...	150	1e-34

BLAST searching the CHO-K1 Genome at NCBI

1) A link to the NCBI BLAST web server is also provided on the CHO-K1 genome search page. To BLAST the CHO-K1 genome using the NCBI BLAST web server, select the “*Cricetulus griseus* WGS” database under the “Choose Search Set” menu.

