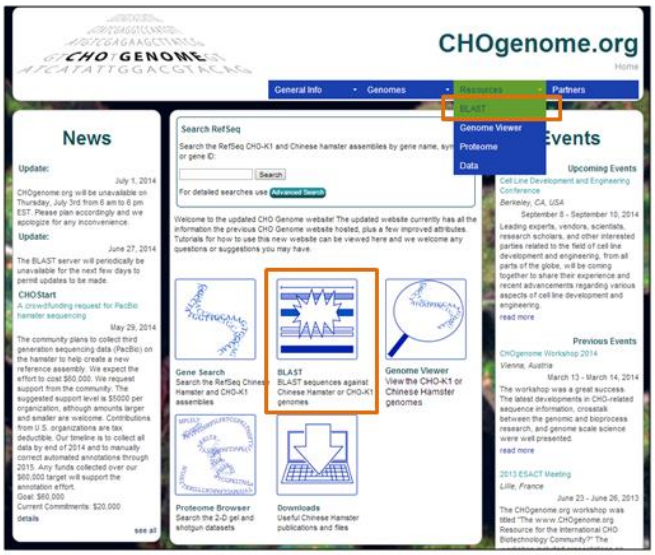
Tutorial 4 – BLAST Searching the CHO Genome

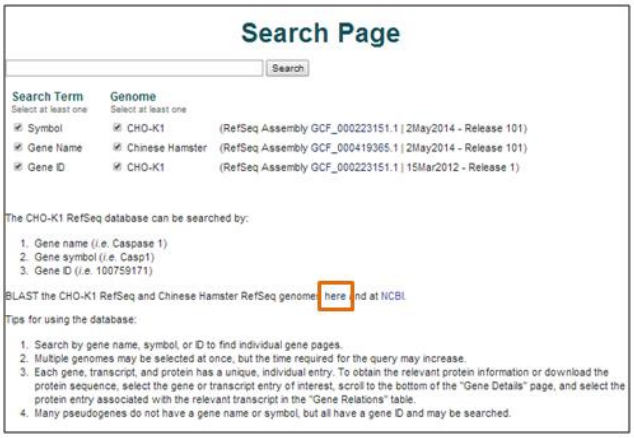
Accessing the CHO Genome BLAST Tool

The CHO BLAST server can be accessed by clicking on the BLAST button on the home page or by selecting “BLAST” from the menu bar under the Resources tab. This tab is available on all web pages within the CHO genome project. An additional link to the CHO genome BLAST web server is also provided on the CHO genome search pages, as well as a link to the NCBI BLAST web server.

Select the BLAST icon on the home page or from the resources tab

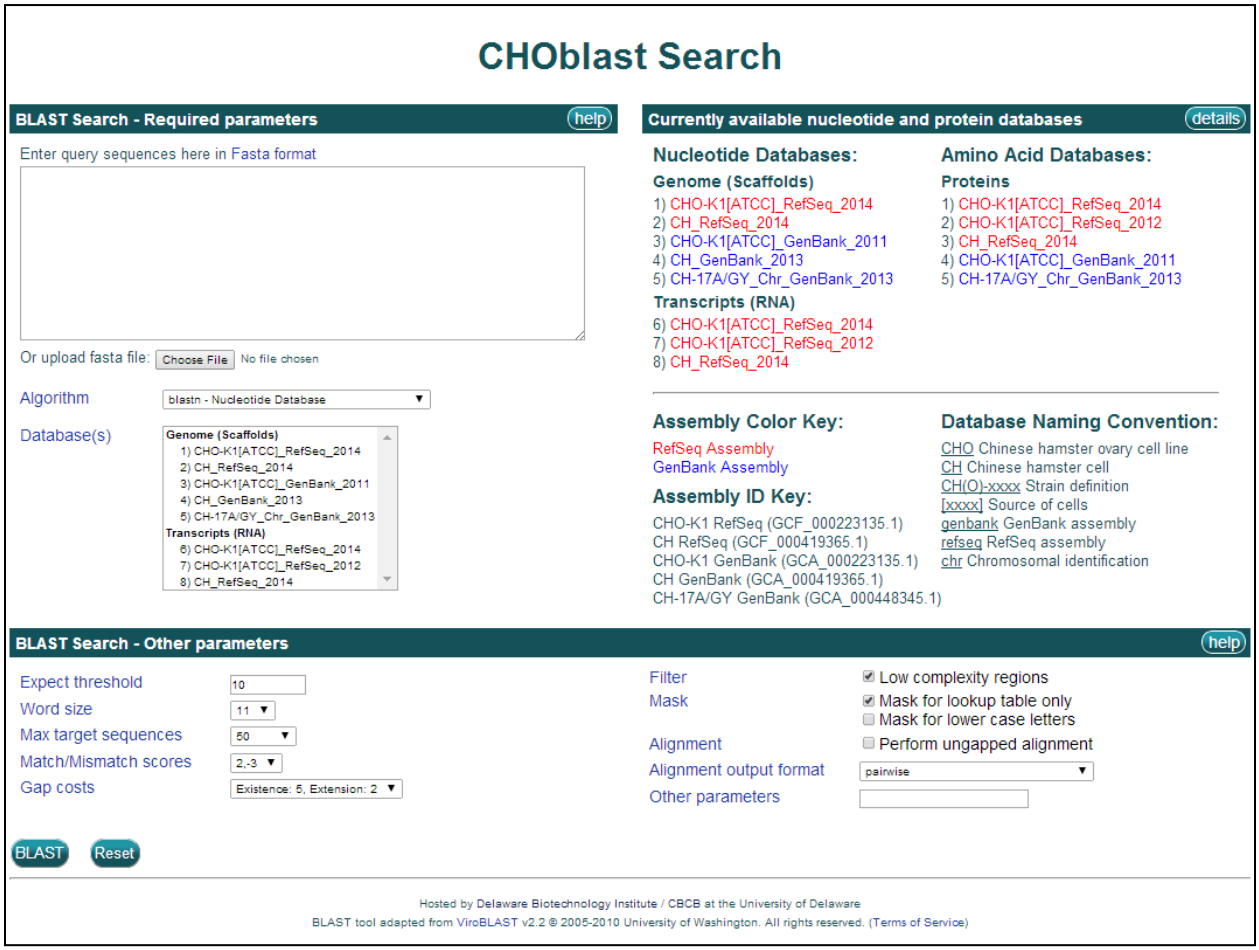


Click on the BLAST server link on the CHO genome search pages



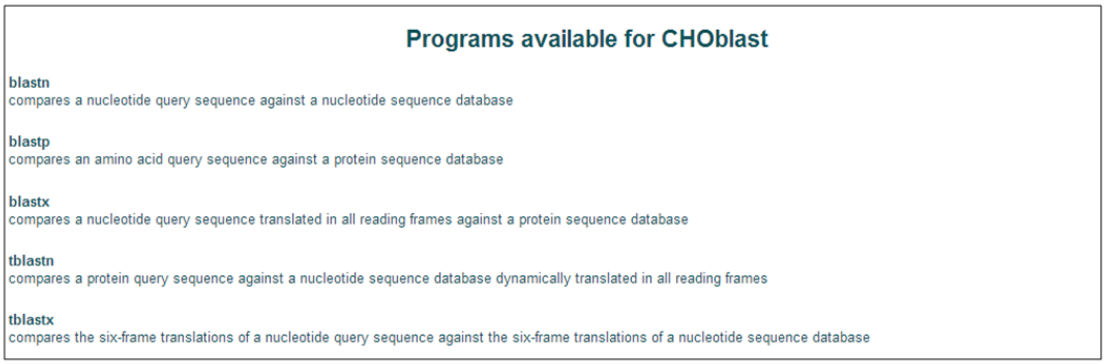
Using the CHO Genome BLAST Tool

1) The CHO BLAST page allows for BLAST searches against the CHO and Chinese hamster (CH) genome databases.



The nucleotide and amino acid databases hosted on the Chinese hamster genome database are listed to the right of the Basic Search panel. The nucleotide databases are divided into Genome (scaffold) and Transcript (RNA) databases, while the amino acid databases consist only of protein databases. The organism or cell line of origin is listed first, followed by the type of assembly (RefSeq or GenBank), and finally the year of release. The keys for the abbreviations and naming conventions are listed below these database lists.

2) Additional details regarding the multiple BLAST programs and databases are available. Clicking on the **Algorithm** link provides a brief description of the BLAST programs.



Programs available for CHOblast

blastn
compares a nucleotide query sequence against a nucleotide sequence database

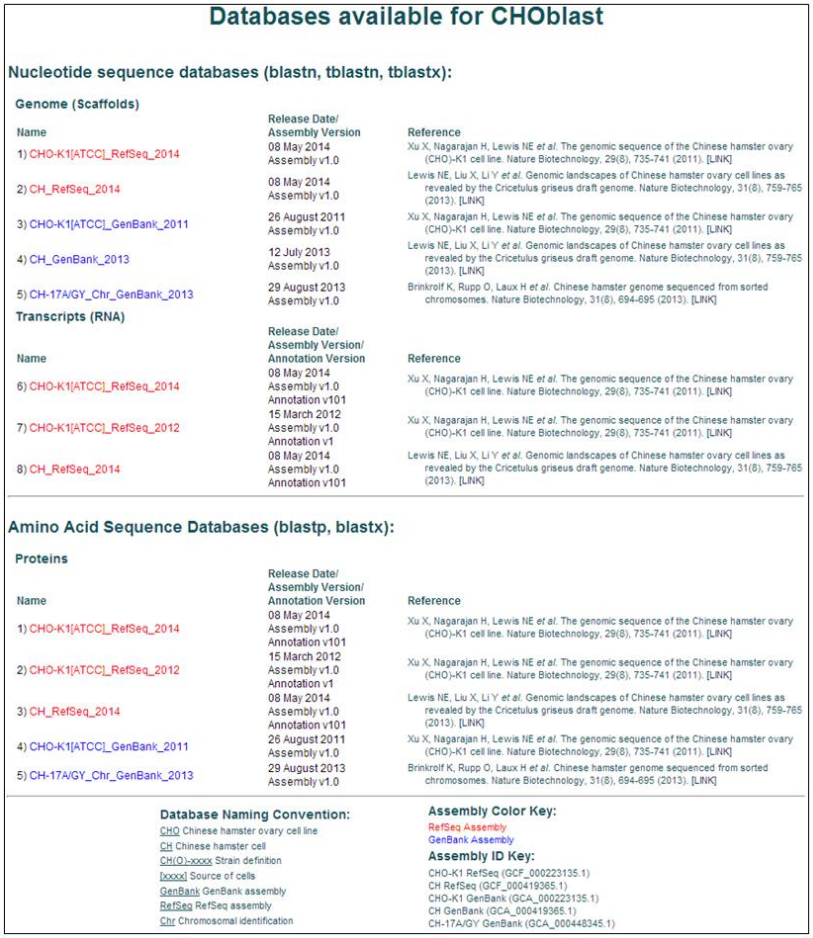
blastp
compares an amino acid query sequence against a protein sequence database

blastx
compares a nucleotide query sequence translated in all reading frames against a protein sequence database

tblastn
compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames

tblastx
compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database

Clicking on the **Database(s)** link or the **details** button in the "Currently available nucleotide and protein databases" section title bar brings up a webpage with a brief description of the databases currently available for BLAST searching, including the name, version, date, and a link to the original publication article.



Databases available for CHOblast

Nucleotide sequence databases (blastn, tblastn, tblastx):

Genome (Scaffolds)

Name	Release Date/ Assembly Version	Reference
1) CHO-K1[ATCC]_RefSeq_2014	08 May 2014 Assembly v1.0	Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735-741 (2011). [LINK]
2) CH_RefSeq_2014	08 May 2014 Assembly v1.0	Lewis NE, Liu X, Li Y *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nature Biotechnology, 31(8), 759-765 (2013). [LINK]
3) CHO-K1[ATCC]_GenBank_2011	26 August 2011 Assembly v1.0	Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735-741 (2011). [LINK]
4) CH_GenBank_2013	12 July 2013 Assembly v1.0	Lewis NE, Liu X, Li Y *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nature Biotechnology, 31(8), 759-765 (2013). [LINK]
5) CH-17A/GY_Chr_GenBank_2013	29 August 2013 Assembly v1.0	Brinkrolf K, Rupp O, Laux H *et al.* Chinese hamster genome sequenced from sorted chromosomes. Nature Biotechnology, 31(8), 694-695 (2013). [LINK]

Transcripts (RNA)

Name	Release Date/ Assembly Version/ Annotation Version	Reference
6) CHO-K1[ATCC]_RefSeq_2014	08 May 2014 Assembly v1.0 Annotation v101	Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735-741 (2011). [LINK]
7) CHO-K1[ATCC]_RefSeq_2012	15 March 2012 Assembly v1.0 Annotation v1	Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735-741 (2011). [LINK]
8) CH_RefSeq_2014	08 May 2014 Assembly v1.0 Annotation v101	Lewis NE, Liu X, Li Y *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nature Biotechnology, 31(8), 759-765 (2013). [LINK]

Amino Acid Sequence Databases (blastp, blastx):

Proteins

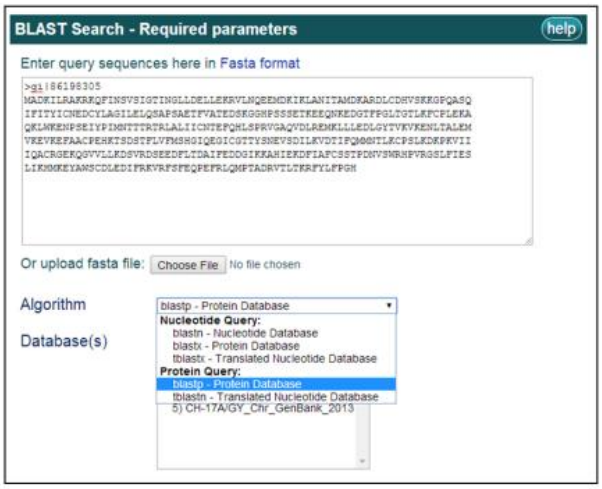
Name	Release Date/ Assembly Version/ Annotation Version	Reference
1) CHO-K1[ATCC]_RefSeq_2014	08 May 2014 Assembly v1.0 Annotation v101	Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735-741 (2011). [LINK]
2) CHO-K1[ATCC]_RefSeq_2012	15 March 2012 Assembly v1.0 Annotation v1	Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735-741 (2011). [LINK]
3) CH_RefSeq_2014	08 May 2014 Assembly v1.0 Annotation v101	Lewis NE, Liu X, Li Y *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nature Biotechnology, 31(8), 759-765 (2013). [LINK]
4) CHO-K1[ATCC]_GenBank_2011	26 August 2011 Assembly v1.0	Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735-741 (2011). [LINK]
5) CH-17A/GY_Chr_GenBank_2013	29 August 2013 Assembly v1.0	Brinkrolf K, Rupp O, Laux H *et al.* Chinese hamster genome sequenced from sorted chromosomes. Nature Biotechnology, 31(8), 694-695 (2013). [LINK]

Database Naming Convention:
CHO Chinese hamster ovary cell line
CH Chinese hamster cell
CH(O)-xxxx Strain definition
[xxxx] Source of cells
GenBank GenBank assembly
RefSeq RefSeq assembly
Chr Chromosomal identification

Assembly Color Key:
RefSeq Assembly
GenBank Assembly

Assembly ID Key:
CHO-K1 RefSeq (GCF_000223135.1)
CH RefSeq (GCF_000419365.1)
CHO-K1 GenBank (GCA_000223135.1)
CH GenBank (GCA_000419365.1)
CH-17A/GY GenBank (GCA_000448345.1)

3) Query sequences in FASTA format can be pasted into the search box at the top of the page or uploaded as a FASTA file. Multiple query sequences may be entered for each search.

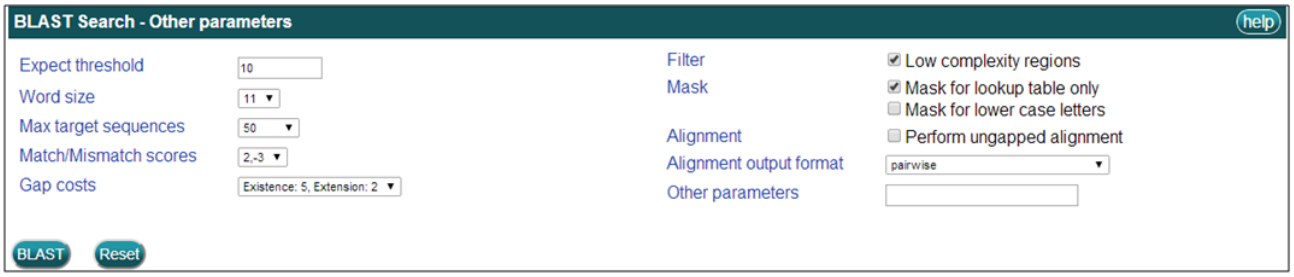


The BLAST program and database are then selected from the currently available options.

For example, to BLAST the most recent CH genome protein database, select the **blastp** program and the **CH_RefSeq_2014** database from the "Proteins" list.

To perform a basic BLAST search, click the **BLAST** button after all the above information is entered and selected. If you wish to perform a more advanced search, do not hit basic search yet and proceed to instruction #4.

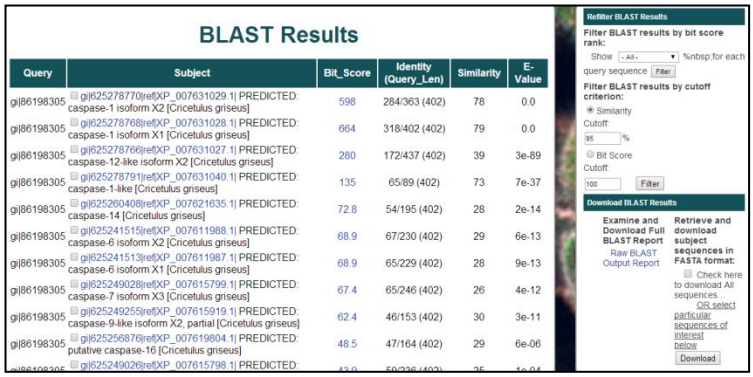
4) In the BLAST Search – Other parameters section, the default BLAST parameters can be varied to perform an altered, more advanced BLAST search.



Clicking on the highlighted blue terms (such as Expect threshold, etc.) will provide a brief description of each advanced search parameter that can be varied.

To perform an advanced BLAST search, click the **BLAST** button once all the required information is entered and the advanced parameters are altered.

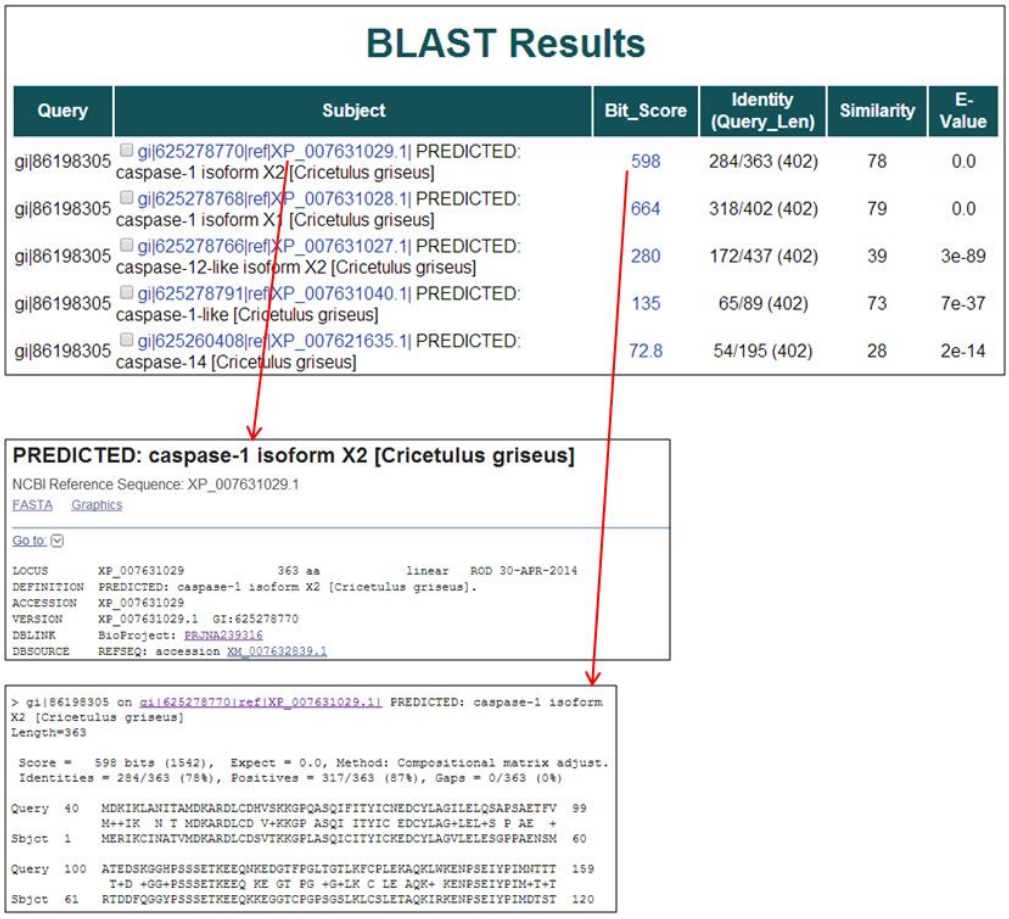
5) The results of the BLAST alignment are summarized in a table with the query sequence name, the subject sequence name, the bit score, the identity length, the identity percentage, and the *E*-value.



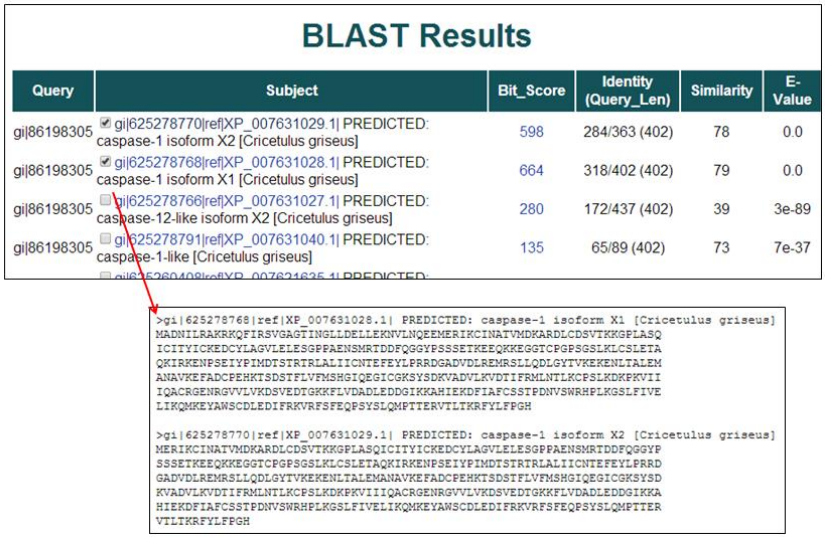
The results can be filtered by score (showing only the top 1, 5, or 10 alignments), by Similarity Cutoff Percentage, or by BLAST Bit Score.

After entering the filter parameter, click either the "Filter" or the "Parse again" buttons to refresh the results table.

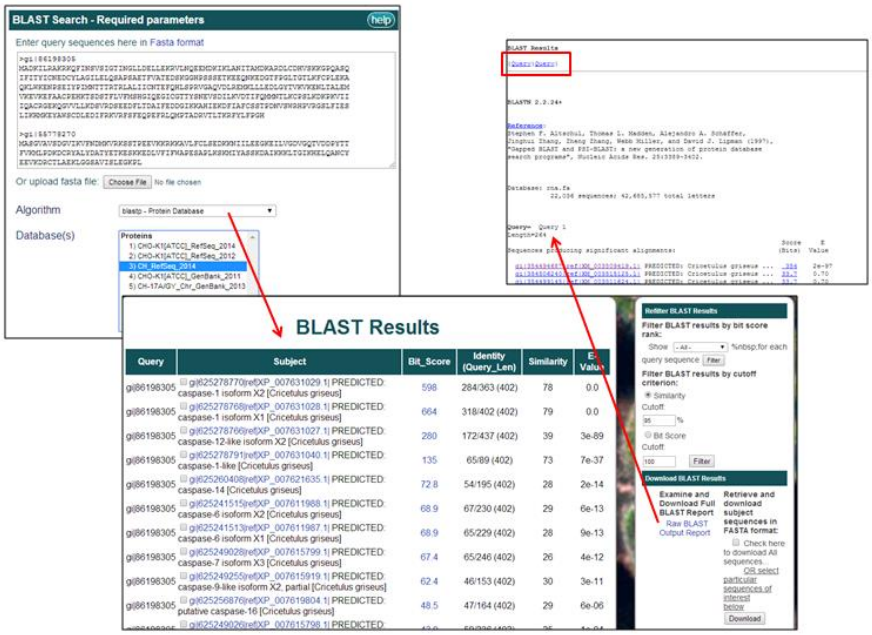
To view the RefSeq/GenBank entry for each subject sequence, click on the sequence name in the Subject column (*i.e.* gi|625278770|ref|XP_007631029.1|).
To view the pair-wise alignment for a specific alignment, click on the value in the Score column for any alignment (*i.e.* 598).



6) To download the resulting BLAST sequences, click the “Check here to download all sequences” box or select individual subject boxes within the table and then hit the “Download” button. A text file of the selected BLAST sequence(s) in the FASTA format will open in a new window.



7) Multiple query sequences can be BLAST searched at one time. The results from all query sequences will be displayed in a single results table. To inspect all the BLAST pair-wise alignments, click on Raw BLAST Output Report, the link located under the “Download BLAST Results” heading. Links provided at the top of the page, in the red outlined box, can be used to quickly navigate between the alignment results for each, individual query sequence.



BLAST searching the CHO-K1 Genome at NCBI

A link to the NCBI BLAST web server is also provided on the CHO-K1 genome search pages. To BLAST the CHO genome using the NCBI BLAST web server, enter the required BLAST information and select the "*Cricetulus griseus* WGS" database under the "Choose Search Set" menu.