

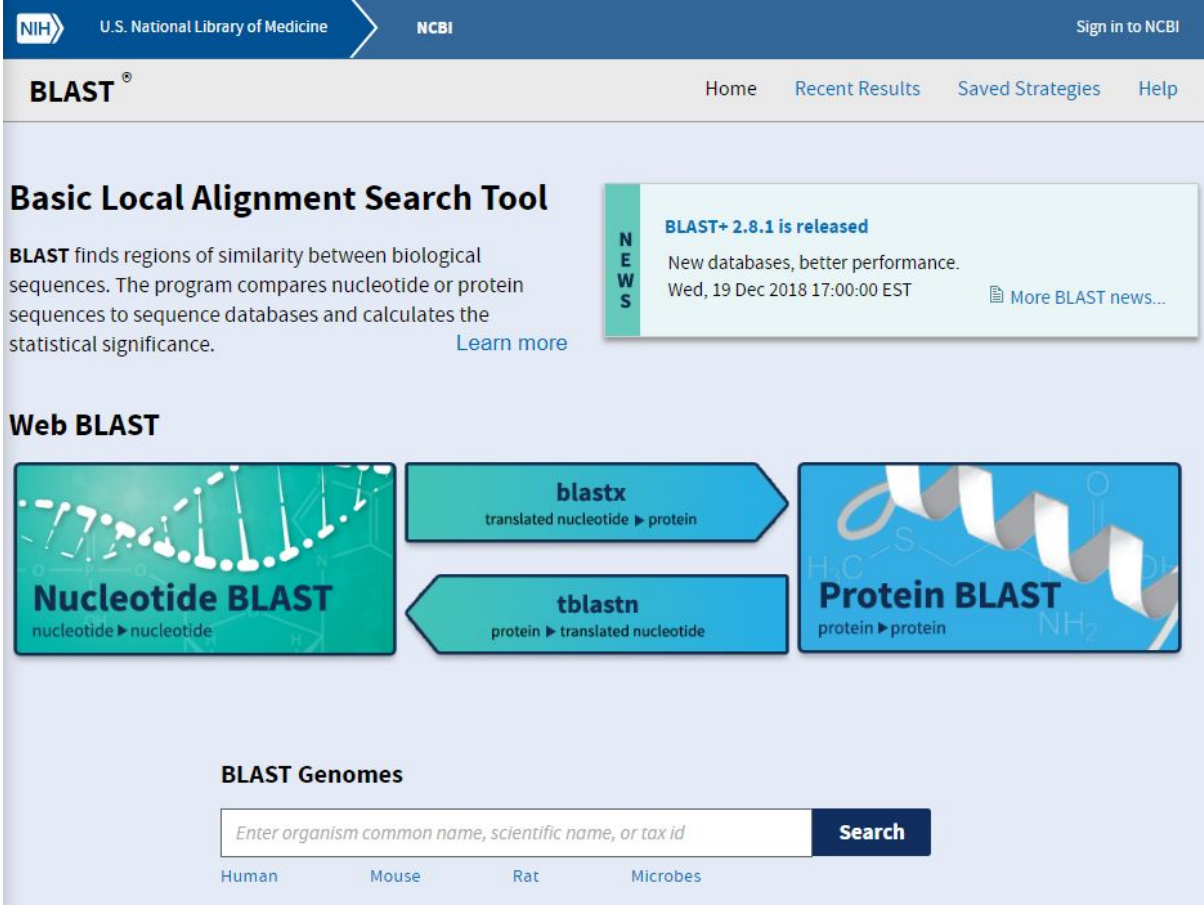
BLAST Guidebook

BLAST, which stands for Basic Local Alignment Search Tool, is a tool that looks for regions of similarity between nucleotide or protein sequences. It allows for researchers to look for other people or databases that have used same or similar sequences in their work.

Access BLAST at

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome

The home page looks like this:

The image is a screenshot of the BLAST (Basic Local Alignment Search Tool) home page. At the top, there is a navigation bar with the NIH logo, "U.S. National Library of Medicine", "NCBI", and a "Sign in to NCBI" link. Below this, the "BLAST" logo is on the left, and navigation links for "Home", "Recent Results", "Saved Strategies", and "Help" are on the right. The main content area features a large heading "Basic Local Alignment Search Tool" with a brief description of BLAST's function and a "Learn more" link. To the right of this is a "NEWS" sidebar announcing "BLAST+ 2.8.1 is released" with details about new databases and performance improvements. Below the main heading, there is a "Web BLAST" section with three large buttons: "Nucleotide BLAST" (nucleotide to nucleotide), "blastx" (translated nucleotide to protein), and "Protein BLAST" (protein to protein). At the bottom, there is a "BLAST Genomes" section with a search input field for organism names and a "Search" button, with links for "Human", "Mouse", "Rat", and "Microbes" below it.

1. Nucleotide BLAST

Nucleotide BLAST compares a nucleotide sequence of your choice to other nucleotide sequences found on databases on the internet.

When you click on “Nucleotide BLAST” on the BLAST home page, it will take you to this page:

NIH U.S. National Library of Medicine NCBI Sign in to NCBI

BLAST » blastn suite Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#) [To](#)

Or, upload file [Choose File](#) No file chosen [Job Title](#) Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

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

Organism Optional Enter organism name or id—completions will be suggested ☐ exclude +
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional ☐ Sequences from type material

Entrez Query Optional Enter an Entrez query to limit search [YouTube](#) [Create custom database](#)

Program Selection

Optimize for ☒ Highly similar sequences (megablast)
☐ More dissimilar sequences (discontiguous megablast)
☐ Somewhat similar sequences (blastn)
Choose a BLAST algorithm

BLAST Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)
☐ Show results in a new window

Enter your sequence of interest in FASTA format in the large box on the top left. FASTA format refers to a method of representing nucleotide sequences where you give only the bases on one strand of DNA. Due to base complementarity, the other strand can be derived. You can set the subrange from your input sequence from which to blast search.

You can also indicate which database to search from - NCBI has lots of organism or species-specific sequences that you can search from.

When you are done inputting all the necessary information above, click “BLAST” to begin your search. Your search output should look something like this:

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST » [blastn suite](#) » RID-4ZYRCAXW014 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

Job title: Nucleotide Sequence (681 letters)

RID: 4ZYRCAXW014 (Expires on 01-30 14:04 pm)

Query ID: lc|Query_137897
 Description: None
 Molecule type: nucleic acid
 Query Length: 681

Database Name: nr
 Description: Nucleotide collection (nt)
 Program: BLASTN 2.8.1+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

[Graphic Summary](#)

Distribution of the top 46 Blast Hits on 45 subject sequences

Color key for alignment scores

■ <40 ■ 40-50 ■ 50-80 ■ 80-200 ■ >=200

Query

1 100 200 300 400 500 600

[Descriptions](#)

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Max score	Total score	Query cover	E value	Ident	Accession
Cloning vector p4g3R, complete sequence	1258	1258	100%	0.0	100%	KX986165.1
Cloning vector pING1001, complete sequence	1258	1258	100%	0.0	100%	KX673874.1
Chlorohydroquinone sensing module vector, complete sequence	1254	1254	99%	0.0	100%	KU746630.1
Benzoate sensing module vector, complete sequence	1254	1254	99%	0.0	100%	KU746629.1

The graph gives us a visual representation of how closely the inputted sequence aligns with sequences found in various databases.

The Descriptions box below provides more specific information about the search results. The most important value in this table is the “Query cover” value, which tells us what percent of a match the database sequence is with our sequence of interest.

If you click on any of the descriptions, it will take you to a page that looks like this:

[Download](#) [GenBank](#) [Graphics](#)

Synthetic construct mCherry gene, complete cds

Sequence ID: [MH070102.1](#) Length: 711 Number of Matches: 1

Range 1: 37 to 671 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
725 bits(392)	0.0	554/635(87%)	0/635(0%)	Plus/Plus
Query 22	ATCAAAGAGTT	CATGCGTTTCAAAGTTC	CGTATGGAAGGTTCCGTTAACGGTCACGAGTTC	81
Sbjct 37	ATCAAAGAGTT	CATGCGTTTCAAAGTGCACATGGAGGGTAGCGTCAACGGTCACGAATTT	96	
Query 82	GAAATCGAAGGTGAAGGTGAAGGT	CGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAA	141	
Sbjct 97	GAAATCGAAGGTGAGGGTGAAGGT	CGCCGTACGAAGGTACCCAAACCGCTAAACTGAAA	156	
Query 142	GTTACCAAAGGTGGTCCGCTGCCGTT	CGCTTGGGACATCCTGTCCCCGAGTTCAGTAC	201	
Sbjct 157	GTGACGAAAGGTGGTCCGCTGCCATT	CGCATGGGATATCCTGTCTCCACAGTTCATGTAC	216	
Query 202	GGTTCCAAAGCTTACGTTAAACACCCGGCT	GACATCCCGGACTACCTGAAACTGTCCTTC	261	
Sbjct 217	GGTTCTAAAGCGTACGTGAAACACCCGGCT	GACATTCTGACTACCTGAAACTGTCCTTC	276	
Query 262	CCGGAAGGTTTCAAATGGGAACGTGTTAT	GAACTTCGAAGACGGTGGTGTGTTACCGTT	321	
Sbjct 277	CCGGAAGGTTTCAAATGGGAACGTGTGAT	GAACTTCGAGGACGGTGGCGTAGTTACTGTT	336	
Query 322	ACCCAGGACTCCTCCCTGCAAGACGGTGAGTT	CATCTACAAAGTTAAACTGCGTGGTACC	381	
Sbjct 337	ACCCAGGACTCTTCCCTGCAGGATGGTGAGTT	TATCTACAAGGTTAAACTGCGTGGCACT	396	
Query 382	AACTTCCCGTCCGACGGTCCGGTTATGCAG	aaaaaaaaCCATGGGTTGGGAAGCTTCCACC	441	
Sbjct 397	AACTTTCCGTCCGACGGCCCGGTTATGCAGA	AAGACTATGGGCTGGGAAGCATCTAGC	456	
Query 442	GAACGTATGTACCCGGAAGACGGTGCTCT	GAAAGGTGAAATCAAAATGCGTCTGAAACTG	501	
Sbjct 457	GAACGTATGTATCCGGAAGATGGTGCTCT	GAAAGGCGAAATCAAACAGCGTCTGAAACTG	516	
Query 502	AAAGACGGTGGTCACTACGACGCTGAAGTT	AAAACACCTACATGGCTAAAAAACCGGTT	561	
Sbjct 517	AAAGACGGCGGCCATTATGATGCGGAAGTT	AAGACGACCTACAAAGCCAAGAAACCGGTT	576	
Query 562	CAGCTGCCGGGTGCTTACAAAACCGACAT	CAAACTGGACATCACCTCCCAACGAAGAC	621	
Sbjct 577	CAGCTGCCGGGCGCCTATAATGTAACAT	CAAACTGGATATTACCTCCCAACGAAGAT	636	
Query 622	TACACCATCGTTGAACAGTACGAACGTGCT	GAAGG	656	
Sbjct 637	TACACCATTGTAGAACAATATGAACGCG	CGGAAGG	671	

This map shows you where the differences are between your input sequence and the result sequence. The lines indicate the same base, whereas the spaces indicate different base.

If you click Sequence ID at the top, you can be taken to the GenBank page for that sequence.

Synthetic construct mCherry gene, complete cds

GenBank: MH070102.1

[FASTA](#) [Graphics](#)
Go to: ☐

```

LOCUS       MH070102                711 bp    DNA     linear   SYN 06-JUN-2018
DEFINITION   Synthetic construct mCherry gene, complete cds.
ACCESSION    MH070102
VERSION      MH070102.1
KEYWORDS     .
SOURCE       synthetic construct
  ORGANISM   synthetic construct
             other sequences; artificial sequences.
REFERENCE    1 (bases 1 to 711)
  AUTHORS    Hui,C.Y., Guo,Y., Zhang,W. and Huang,X.Q.
  TITLE      Rapid monitoring of the target protein expression with a
             fluorescent signal based on a dicistronic construct in Escherichia
             coli
  JOURNAL    AMB Express 8 (1), 81 (2018)
  PUBMED     29785487
  REMARK     Publication Status: Online-Only
REFERENCE    2 (bases 1 to 711)
  AUTHORS    Hui,C., Guo,Y., Zhang,W. and Huang,X.
  TITLE      Direct Submission
  JOURNAL    Submitted (14-MAR-2018) Department of Pathology & Toxicology,
             Shenzhen Prevention and Treatment Center for Occupational Diseases,
             No. 2019 Buxin Road, Luohu District, Shenzhen, Guangdong 518020,
             China
COMMENT      ##Assembly-Data-START##
             Sequencing Technology :: Sanger dideoxy sequencing
             ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source           1..711
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS              1..711
                     /note="codon optimized for E. coli"
                     /codon_start=1
                     /transl_table=11
                     /product="mCherry"
                     /protein_id="AW076971.1"

```

GenBank is another tool made by NCBI that compiles DNA sequences and their origin of publication. On this page, you can find an annotated sequence, as well as the paper and authors who first published that sequence.

2. Blastx, Tblastn, pblast

BLAST is not exclusive to nucleotide sequences - it can also be used with amino acid sequences, or between nucleotides and amino acid sequences.

Blastx is a tool that allows you to search for your translated nucleotide sequence in comparison to available amino acid sequences. This is useful if you have a sequence but you do not know what it codes for.

Tblastn is a tool that matches your amino acid sequence to RNA nucleotide sequences in the database. This may be useful when you're trying to find sequences of the same protein that are codon-optimized to various organisms.

Pblast is a tool that compares amino acid sequences. You may be able to find other proteins with similar domains, or mutations of your protein with this tool.

3. Other available BLAST tools

There are many more tools that BLAST offers aside from the above four main search tools, including a conserved domain search or a search for markers. If you want to find things similar to any biological molecule, be it by function, morphology, chemical structure, etc. BLAST is probably the place to go. Find these extra functions at the bottom of the BLAST home page.

4. General Tips and tl;dr

Before any sequence is used or ordered in the lab, it must be put through BLAST. This is to make sure that we are using a sequence that has been used and verified before. The more matches you get on BLAST, the better. It might be a good idea to look through source papers on any information about that part that might help inform your experiments.

Tl;dr blast has lots of sequences, use them.

Feel free to DM us (Sara, Cassie, or Chris) if you have any questions ;)