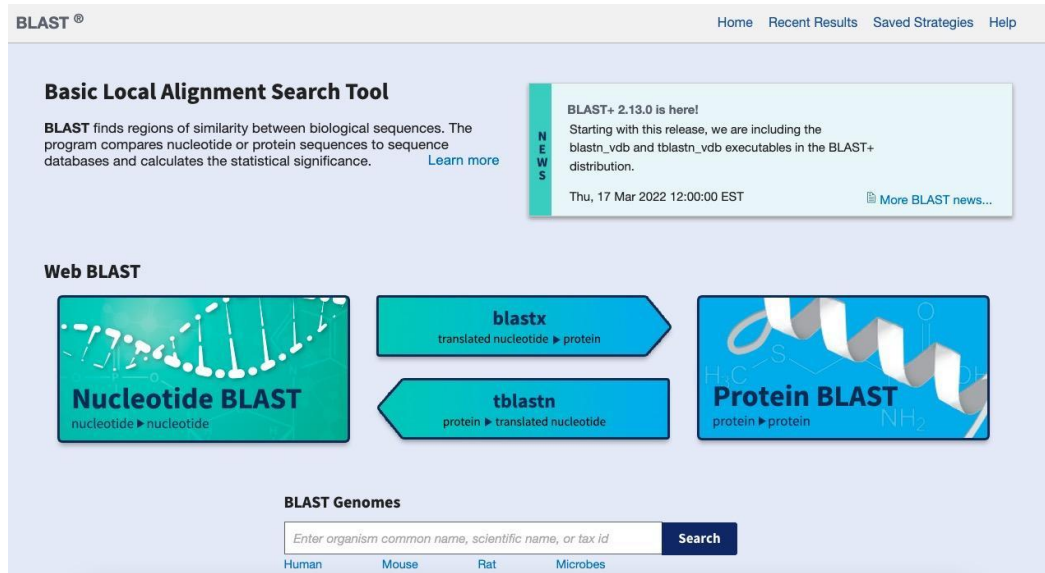


Online BLAST Guide

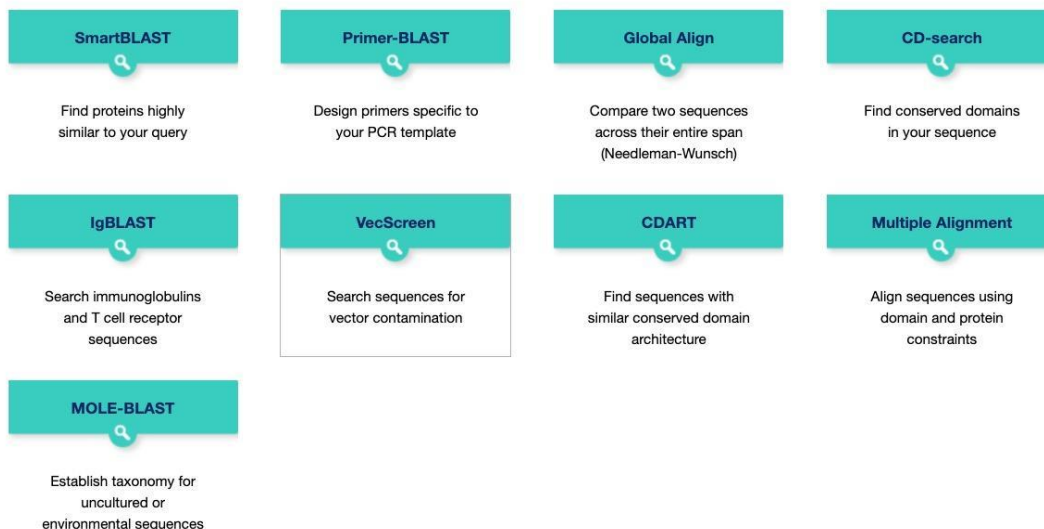
Here is the landing page for Online BLAST: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>



The screenshot shows the BLAST landing page with a navigation bar at the top containing 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. The main content area is divided into several sections. On the left, there's a 'Basic Local Alignment Search Tool' section with a brief description and a 'Learn more' link. To the right of this is a 'NEWS' box announcing BLAST+ 2.13.0, mentioning the inclusion of 'blastn_vdb' and 'tblastn_vdb' executables, dated 'Thu, 17 Mar 2022 12:00:00 EST', with a 'More BLAST news...' link. Below the 'Basic Local Alignment Search Tool' section is the 'Web BLAST' section, which features three large buttons: 'Nucleotide BLAST' (nucleotide to nucleotide), 'blastx' (translated nucleotide to protein), and 'tblastn' (protein to translated nucleotide). To the right of these buttons is a 'Protein BLAST' button (protein to protein). Below the 'Web BLAST' section is the 'BLAST Genomes' section, which includes a search bar with the placeholder text 'Enter organism common name, scientific name, or tax id' and a 'Search' button. Below the search bar are four links: 'Human', 'Mouse', 'Rat', and 'Microbes'.

Once here one has to know what to query as there are a suite of BLAST tools available for searching multifarious databases (just to name a few NCBI NR Database, refseq, Swissprot, etc.). Additionally, to fire up the search queries there are various algorithms handling the BLAST searches and providing multiple flavors to fetch the results. There are even more tool(s) available in the landing page mentioned below, to streamline the BLAST searches;

Specialized searches



The 'Specialized searches' section displays a grid of seven specialized BLAST tools, each with a magnifying glass icon. The tools are arranged in three rows: the first row contains 'SmartBLAST', 'Primer-BLAST', 'Global Align', and 'CD-search'; the second row contains 'IgBLAST', 'VecScreen', 'CDART', and 'Multiple Alignment'; and the third row contains 'MOLE-BLAST'. Each tool has a brief description of its function.

Tool	Description
SmartBLAST	Find proteins highly similar to your query
Primer-BLAST	Design primers specific to your PCR template
Global Align	Compare two sequences across their entire span (Needleman-Wunsch)
CD-search	Find conserved domains in your sequence
IgBLAST	Search immunoglobulins and T cell receptor sequences
VecScreen	Search sequences for vector contamination
CDART	Find sequences with similar conserved domain architecture
Multiple Alignment	Align sequences using domain and protein constraints
MOLE-BLAST	Establish taxonomy for uncultured or environmental sequences

But for our use-cases, we will be only concentrating on the blastn and blastp. However, one can refer to the [official guide](#) to get a detailed understanding about the capabilities of BLAST.

Here is a reference to which BLAST tool to specifically use based on the query sequence and the BLAST DB available, it's quite confusing at times but the below image should suffice:

Query sequence type	Database sequence type	Alignment level type	What the program should be called	What the program is actually called
nucleotide	nucleotide	nucleotide	blastNN	blastn
peptide	peptide	peptide	blastPP	blastp
nucleotide	peptide	peptide	blastNP	blastx
peptide	nucleotide	peptide	blastPN	tblastn
nucleotide	nucleotide	peptide	blastNNP	tblastx

Figure: Reference from Biostar Handbook

Example of querying a peptide sequence against NCBI NR Database – Online BLAST

One can query the Online BLAST protein sequences here ⇒

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LC=blasthome

Follow the steps as highlighted in black from the below images;

BLAST® » blastp suite

Home Recent Results Saved Strategies Help

blastn **blastp** blastx tblastn tblastx

Standard Protein BLAST

BLASTP programs search protein databases using a protein query, more...

Reset page Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

unnamed protein product ACDEFGHI

Query subrange

From To

Or, upload file

Choose file No file chosen

Job Title

unnamed protein product

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Databases

☒ Standard databases (nr etc.) [New](#) ☐ Experimental databases

Compare

☐ Select to compare standard and experimental database

Standard

Database

Non-redundant protein sequences (nr)

Organism

Optional

Enter organism name or id—completions will be suggested

exclude Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude

Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

And select only the Standard NR Database.

Program Selection

Algorithm

☐ Quick BLASTP (Accelerated protein-protein BLAST)

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

BLAST

Search database nr using Blastp (protein-protein BLAST)

☒ Show results in a new window

Filtering the results based on the percent identity:

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastp suite » results for RID-DZGC979T013

Home Recent Results Saved Strategies Help

[Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Filter Results

Organism only top 20 will appear ☐ exclude
Type common name, binomial, taxid or group name
[Add organism](#)

Percent Identity **E value** **Query Coverage**

90 to 100 to to

[Filter](#) [Reset](#)

Compare these results against the new Clustered nr database [BLAST](#)

Job Title Protein Sequence
RID DZGC979T013 Search expires on 07-27 18:27 pm [Download All](#)
Program BLASTP [Citation](#)
Database nr [See details](#)
Query ID lc|Query_6935
Description unnamed protein product
Molecule type amino acid
Query Length 8
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastp suite » results for RID-E1JRA600016

Home Recent Results Saved Strategies Help

[Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Filter Results

Organism only top 20 will appear ☐ exclude
Type common name, binomial, taxid or group name
[Add organism](#)

Percent Identity **E value** **Query Coverage**

90 to 100 to to

[Filter](#) [Reset](#)

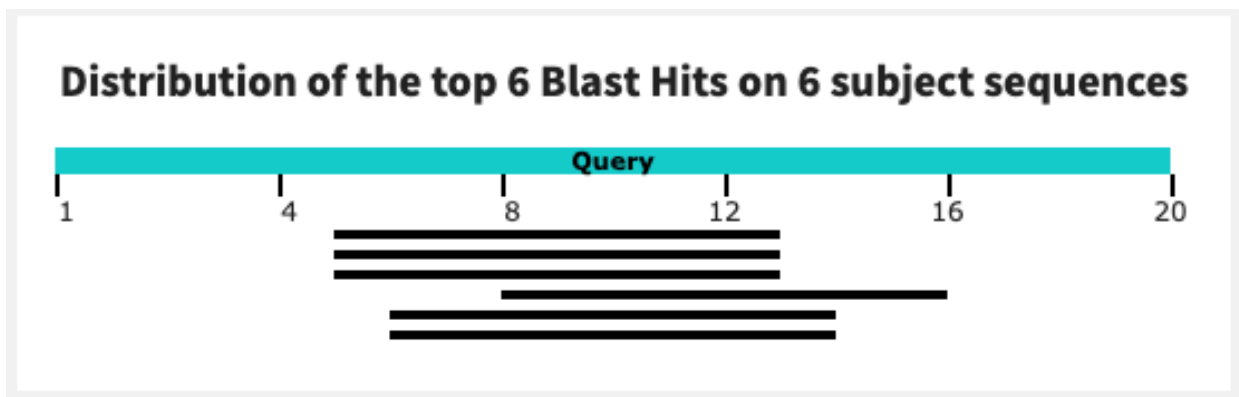
Job Title unnamed_protein_product
RID E1JRA600016 Search expires on 07-28 13:19 pm [Download All](#)
Program BLASTP [Citation](#)
Database nr [See details](#)
Query ID lc|Query_28655
Description unnamed_protein_product
Molecule type amino acid
Query Length 20
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

The search result gets curated upon filtering and we can switch between “Descriptions, Graphic Summary, Alignments and Taxonomy” for further understanding on the query and subject matches.

“Description” from the report:

Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download ▼ Select columns ▼ Show 100 ▼ ?								
<input checked="" type="checkbox"/> select all 6 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer								
Description ▼	Scientific Name ▼	Max Score ▼	Total Score ▼	Query Cover ▼	E value ▼	Per. Ident	Acc. Len ▼	Accession
<input checked="" type="checkbox"/> NAD(P)-dependent oxidoreductase [Flavonifractor sp.]	Flavonifractor sp.	30.8	30.8	40%	217	100.00%	289	MCI9473837.1
<input checked="" type="checkbox"/> NAD(P)-dependent oxidoreductase [Oscillospiraceae bacterium]	Oscillospiraceae bacterium	30.8	30.8	40%	217	100.00%	327	MCI9037848.1
<input checked="" type="checkbox"/> Sensor protein ZraS [candidate division BRC1 bacterium ADurb.Bin183]	candidate division BRC1 bacterium ADurb.Bin183	30.8	30.8	40%	218	100.00%	430	QQB21908.1
<input checked="" type="checkbox"/> uncharacterized protein LOC125669061 [Ostrea edulis]	Ostrea edulis	30.3	30.3	40%	303	100.00%	170	XP_048759465.1
<input checked="" type="checkbox"/> DUF1624 domain-containing protein [Methanomassiliicoccales archaeon]	Methanomassiliicoccales archaeon	29.9	29.9	40%	431	100.00%	252	MCJ7516422.1
<input checked="" type="checkbox"/> glycosyltransferase [Candidatus Saccharibacteria bacterium]	Candidatus Saccharibacteria bacterium	29.9	29.9	40%	434	100.00%	376	MBR3252638.1

For the protein sequence(ADIMWQVRSFCYLGHTKEPN) here is the “Graphic Summary”:



“Alignments” ⇒ gives us **Query – Subject** mapping along with **Identities** and **Gaps**


[Download](#)

[GenPept](#)
[Graphics](#)

NAD(P)-dependent oxidoreductase [Flavonifractor sp.]

Sequence ID: [MCI9473837.1](#) Length: 289 Number of Matches: 1

Range 1: 193 to 200 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Positives	Gaps
30.8 bits(65)	217	8/8(100%)	8/8(100%)	0/8(0%)

Query

6

QVRSFCYL

13

Sbjct

193

QVRSFCYL

200

[Download](#) [GenPept](#) [Graphics](#)

NAD(P)-dependent oxidoreductase [Flavonifractor sp.]

Sequence ID: [MCI9473837.1](#) Length: 289 Number of Matches: 1

Range 1: 193 to 200 [GenPept](#) [Graphics](#)

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

Score	Expect	Identities	Positives	Gaps
30.8 bits(65)	217	8/8(100%)	8/8(100%)	0/8(0%)
Query 6	QVRSFCYL 13			
	QVRSFCYL			
Sbjct 193	QVRSFCYL 200			

[Download](#) [GenPept](#) [Graphics](#)

NAD(P)-dependent oxidoreductase [Oscillospiraceae bacterium]

Sequence ID: [MCI9037848.1](#) Length: 327 Number of Matches: 1

Range 1: 231 to 238 [GenPept](#) [Graphics](#)

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

Score	Expect	Identities	Positives	Gaps
30.8 bits(65)	217	8/8(100%)	8/8(100%)	0/8(0%)
Query 6	QVRSFCYL 13			
	QVRSFCYL			
Sbjct 231	QVRSFCYL 238			