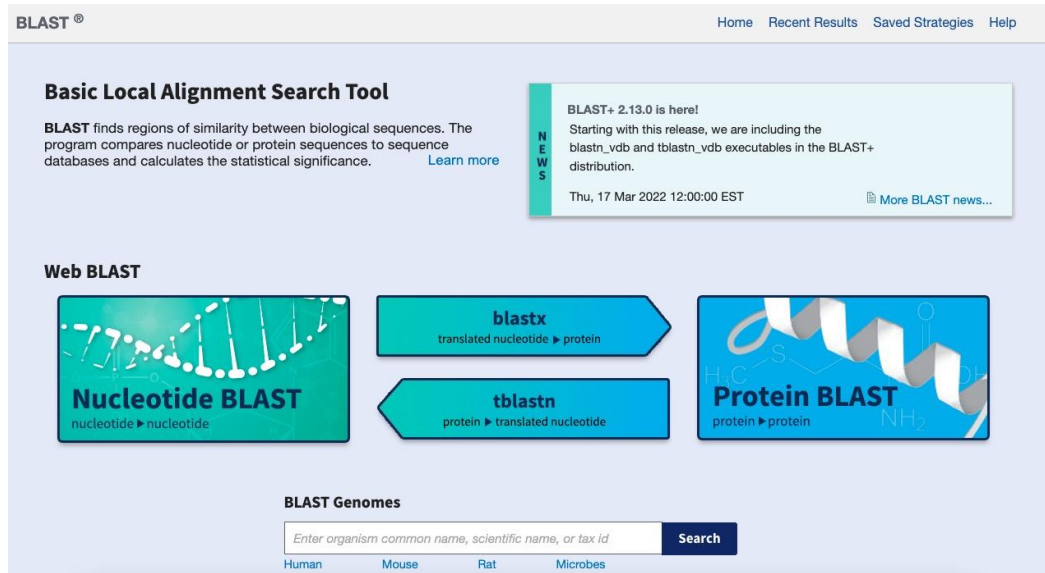


# Online BLAST Reference Guide

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

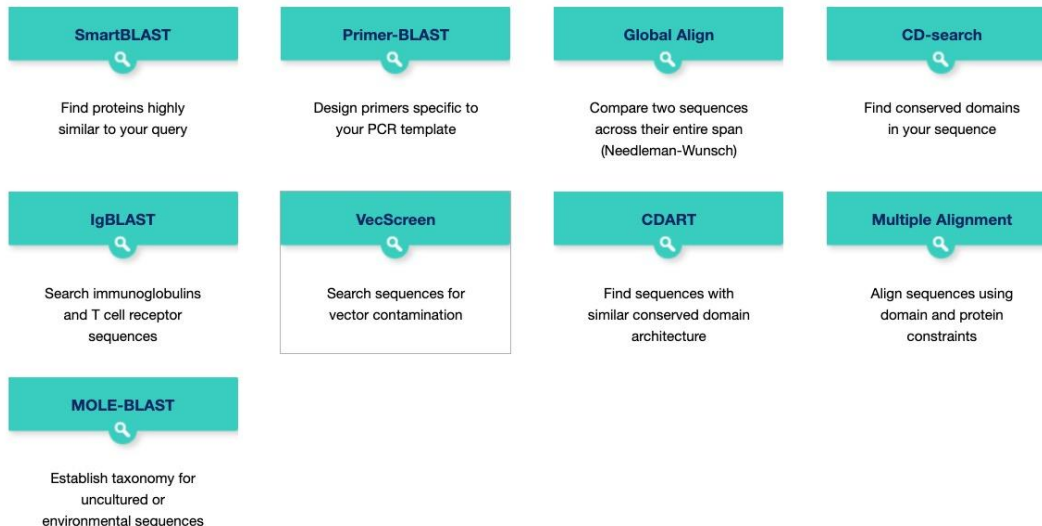


The screenshot shows the BLAST landing page. At the top, there's a navigation bar with links: Home, Recent Results, Saved Strategies, and Help. Below this, the main heading is "Basic Local Alignment Search Tool". A brief description states: "BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance." A "Learn more" link is provided. To the right, a "NEWS" box announces "BLAST+ 2.13.0 is here!" and mentions the inclusion of blastn\_vdb and tblastn\_vdb executables. Below the news box, the "Web BLAST" section features three main options: "Nucleotide BLAST" (nucleotide to nucleotide), "blastx" (translated nucleotide to protein), and "tblastn" (protein to translated nucleotide). To the right of these is "Protein BLAST" (protein to protein). At the bottom, the "BLAST Genomes" section includes a search bar for organism names and a "Search" button, with links for 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 multiple databases (just to name a few NCBI NR Database, refseq, Swissprot, etc.).

Also, to fire up the search queries we have different 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



A grid of specialized BLAST search tools. Each tool is represented by a teal box with a magnifying glass icon and a description of its function:

- 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	<b>blastNN</b>	<b>blastn</b>
peptide	peptide	peptide	<b>blastPP</b>	<b>blastp</b>
nucleotide	peptide	peptide	<b>blastNP</b>	<b>blastx</b>
peptide	nucleotide	peptide	<b>blastPN</b>	<b>tblastn</b>
nucleotide	nucleotide	peptide	<b>blastNNP</b>	<b>tblastx</b>

## 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](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](#)

Query subrange [?](#)

From

To

Or, upload file  No file chosen [?](#)

Job Title

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  [?](#)

Organism [Optional](#)

☐ 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 **90** to **100** E value  to  Query Coverage  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 **90** to **100** E value  to  Query Coverage  to

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

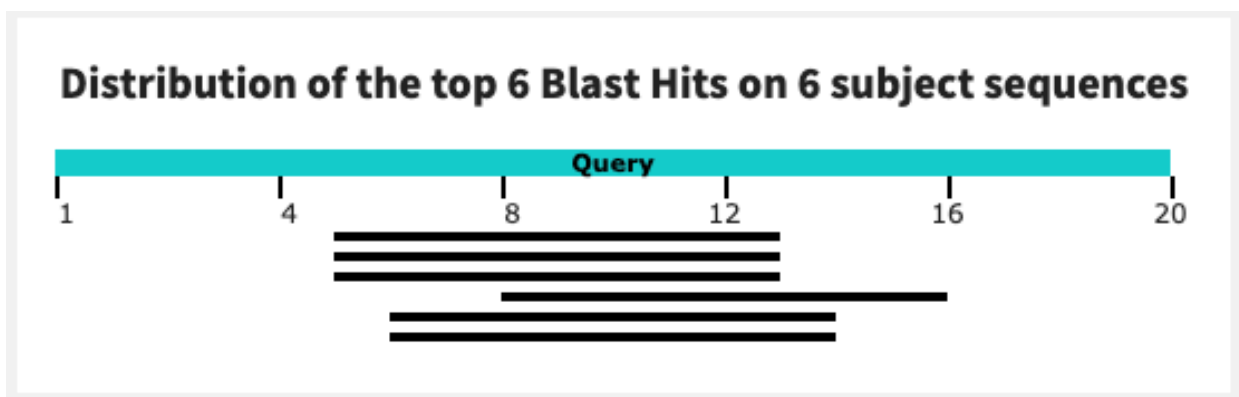
**Job Title** unnamed\_protein\_product  
**RID** E1JRA600016 Search expires on 07-28 13:19 pm [Download All](#)  
**Program** BLASTP [Citation](#)  
**Database** nr [See details](#)  
**Query ID** lc|Query\_28655  
**Description** unnamed\_protein\_product  
**Molecule type** amino acid  
**Query Length** 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 <span>▼</span> Select columns <span>▼</span> Show <span>100</span> <span>▼</span> <span>?</span>								
<input checked="" type="checkbox"/> select all 6 sequences selected <span>GenPept</span> <span>Graphics</span> <span>Distance tree of results</span> <span>Multiple alignment</span> <span>MSA Viewer</span>								
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	<a href="#">MCI9473837.1</a>
<input checked="" type="checkbox"/> NAD(P)-dependent oxidoreductase [Oscillospiraceae bacterium]	Oscillospiraceae bacterium	30.8	30.8	40%	217	100.00%	327	<a href="#">MCI9037848.1</a>
<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	<a href="#">QQB21908.1</a>
<input checked="" type="checkbox"/> uncharacterized protein LOC125669061 [Ostrea edulis]	Ostrea edulis	30.3	30.3	40%	303	100.00%	170	<a href="#">XP_048759465.1</a>
<input checked="" type="checkbox"/> DUF1624 domain-containing protein [Methanomassiliicoccales archaeon]	Methanomassiliicoccales archaeon	29.9	29.9	40%	431	100.00%	252	<a href="#">MCJ7516422.1</a>
<input checked="" type="checkbox"/> glycosyltransferase [Candidatus Saccharibacteria bacterium]	Candidatus Saccharibacteria bacterium	29.9	29.9	40%	434	100.00%	376	<a href="#">MBR3252638.1</a>

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			

#####

## Online BLAST Sequences:

### - Online BLAST 4 Sequence (delta sleep-inducing peptide):

RecName: Full=Delta sleep-inducing peptide; Short=DSIP

UniProtKB/Swiss-Prot: P01158.1

GenPept Identical Proteins Graphics

>sp|P01158.1|DSIP\_RABIT RecName: Full=Delta sleep-inducing peptide;

Short=DSIP

WAGGDASGE

### - Online BLAST 5 Sequence (PPF-1 protein sequence):

RecName: Full=Inner membrane protein PPF-1, chloroplastic; AltName: Full=Post-floral-specific protein

UniProtKB/Swiss-Prot: Q9FY06.2

GenPept Identical Proteins Graphics

>sp|Q9FY06.2|PPF1\_PEA RecName: Full=Inner membrane protein PPF-1, chloroplastic; AltName: Full=Post-floral-specific protein 1; Flags:

Precursor

MAKTLISSPSFLGTPPLSLHRTFSPNRTLFTKVQFSFHQLPPIQSVSHSVDLSGIFARAEGLLYTLADATVA  
ADAAASTDVAAQKNGGWFGFISDGMEFVLKVLKDGLSSVHVPYSYGFAIILLTVIVKAATLPLTKQQVESTLA  
MQNLQPKIKAIQERYAGNQERIQLETSRLYTQAGVNPLAGCLPTLATIPVWIGLYQALSNVANEGLLTEGFLW  
IPSLGGPTSIAARQSGSGISWLFPPVDGHPLLGWYDTAAYLVLPVLLIVSQYVSMEIMKPPQTNDPNQKNTLL  
IFKFLPLMIGYFSLSVPSGLTIYWFTNNVLSTAQQVWLRLKLGAKPAVNENAGGIITAGQAKRSASKPEKGGE  
RFRQLKEEEKKKKLIKALPVEEVQPLASASASNDGSDVENNKEQEVTEESNTSKVSQEVQSFSRERRRSKRSKR  
KPVA

### - Online BLAST 7 Sequence (OsHT01 protein sequence):

histidine amino acid transporter (protoplast) [Oryza sativa Indica Group]

GenBank: CAD89802.1

GenPept Identical Proteins Graphics

>CAD89802.1 histidine amino acid transporter (protoplast) [Oryza sativa Indica Group]

MAKQWWQDGRSAQEKAIDDWLPITSSRNAKWWYSAFHNVTAMVGAGVLSLPYAMSELGWGPGIAVLILSWIIT  
LYTLWQMVMEMHEMVPGRKRFDRYHELGHAFGEKLGWLWIVVPQQLVVEVGVNIVYMTGGKSLKKFHDVLCCEGH  
GCKNIKLTYFIMIFASVHFVLSQLPNFNSISGVSLAAVMSLSYSTIAWGASVDKGVADVDYHLRATTSTGK  
VFGFFSALGDVAFAYAGHNVVLEIQATIPSTPEKPSKKPMWKGVVVAYIIIVALCYFPVALVGWAFGNHVDN  
ILITLSRPKWLIALANMMVVIHVIGSYQIYAMPVFMDIETVLVKKLRFPPGLTLRLIARTLYVAFTMFIAITF

PFFGGLLGFFGGFAFAPTTYFLPCIMWLAIYKPRRFSLSWFTNWICIIILGVMLMILSPIGGLRQIIIDAKTYK  
FYS

## - Online BLAST 8 Sequence (Q57997)

RecName: Full=Universal stress protein MJ0577; Short=USP MJ0577

UniProtKB/Swiss-Prot: Q57997.1

GenPept Identical Proteins Graphics

>sp|Q57997.1|Y577\_METJA RecName: Full=Universal stress protein MJ0577;  
Short=USP MJ0577

MSVMYKKILYPTDFSETAEIALKHVKAFKTLKAEVILLHVIDEREIKKRDIFSLLLGVAGLNKSVVEEFE  
NELKNKLTEEAKNMENIKKELEDVGFKVKDIIIVVGIPHEEIVKIAEDEGVDDIIMGSHGKTNLKEILLG  
SVTENVIKKSNKPVLVVKRKNS

## - Online BLAST 9 Sequence (Cytokine Induced Protein Sequence – NP149073)

SAP domain-containing ribonucleoprotein [Homo sapiens]

NCBI Reference Sequence: NP\_149073.1

GenPept Identical Proteins Graphics

>NP\_149073.1 SAP domain-containing ribonucleoprotein [Homo sapiens]  
MATETVELHKLKLAELKQECLARGLETGKIKQDLIHRLQAYLEEHAEEEEANEEDVLGDETEEEETKPIEL  
PVKEEEPPEKTVDVAAEKKVVKITSEIPQTERMQKRAERFNVFVSLESKKAARAARFGISSVPTKGLSSD  
NKPMVNLDKLERAQRFGLNVSSISRKSEDEKLKKRKERFGIVTSSAGTGTTEDEAKKRKRAERFGIA

## - Online BLAST 10 Sequence (olfactory receptor protein sequence NP\_001005182)

olfactory receptor 6C1 [Homo sapiens]

NCBI Reference Sequence: NP\_001005182.1

GenPept Identical Proteins Graphics

>NP\_001005182.1 olfactory receptor 6C1 [Homo sapiens]  
MRNHTEITEFILLGLTDDPNFQVVIFVFLITYMLSITGNLTLITITLLDSHLQTPMYFFLRNFSILEIS  
FTTVSIPKFLGNIISGDKTISFNNCIVQLFFIFILLGVTEFYLLAAMSVDYVAICKPLHCLSIMNRRVCT  
LLVFTSWLVSFLIIFPALMLLLKLHYCRSNIIDHFTCDYFPLLQLACSDTKFLEVVMGFSCAAFTLMFTLA  
LIFLSYIYIIRTILRIPSTSQRTKAFSTCSSHMVVVSISYGSCIFMYIKPSAKDRVSLSKGVAILNTSVA  
PMMNPFYISLRNQVKQAFINMARKTVFFFTST



## - Online BLAST 11 Sequence (Universal Stress Protein)

```
>Y577_METJA - Q57997, Universal stress protein, MJ0577; J Luo, 2016-08-23
MSVMYKKILYPTDFSETAEIALKHVKAFKTLKAEEVILLHVIDEREIKKRDIFSLLLGVA
GLNKSVEEFENELKNKLTEEAKNKMENIKKELEDVGFKVKDIIIVGIPHEEIVKIAEDEG
VDIIIMGSHGKTNLKEILLGSVTENVIKKSNKPVLVVKRNS
```

#####

The output and alignments for the above-mentioned sequences are saved in the [drive](#).