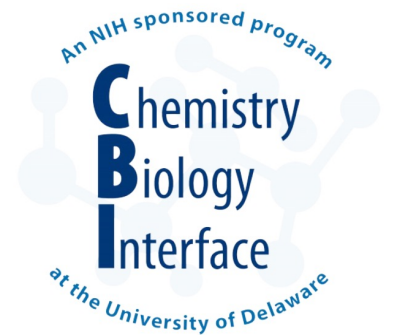


Web Resources

FAIR Data Practices for Omics Analysis Workshop
University of Delaware
April 18 (Day 1)



National Center for Biotechnology Information (NCBI) Resources

- NIH National Library of Medicine
- Repository for biomedical and genomic information
- Resources include:
 - PubMed
 - PubChem
 - GenBank
 - RefSeq
 - Sequence Read Archive (SRA)
 - Many others...



All Databases

Search

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts
into NCBI databases



Download

Transfer NCBI data to your
computer



Learn

Find help documents, attend a
class or watch a tutorial



Develop

Use NCBI APIs and code
libraries to build applications



Analyze

Identify an NCBI tool for your
data analysis task



Research

Explore NCBI research and
collaborative projects



COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

New RefSeq Annotations!

13 Apr 2022

In February and March, the NCBI
Eukaryotic Genome Annotation Pipeline

New feature in the MSA viewer: Search
for a short sequence

12 Apr 2022

We're reading and incorporating your

MANE is published in Nature!

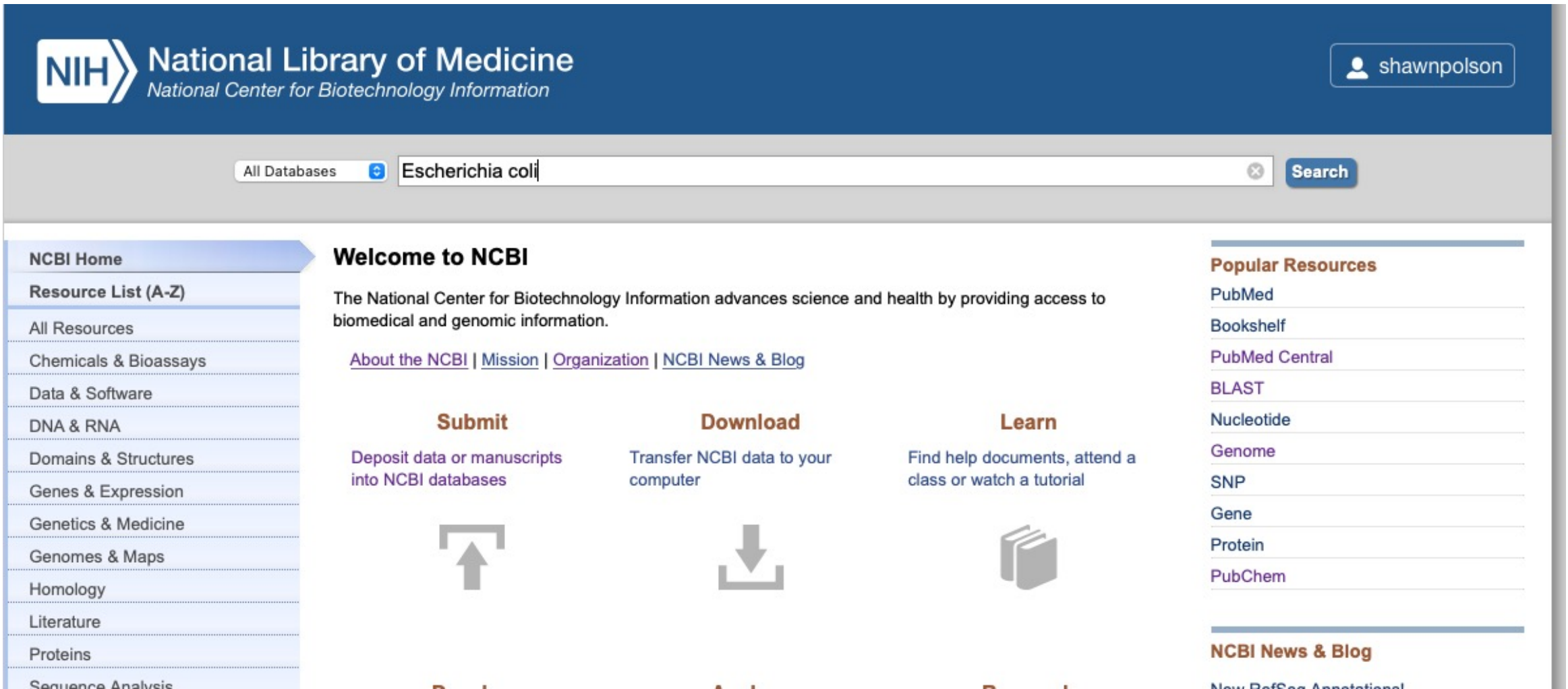
06 Apr 2022

We are delighted to announce that three
and a half years of hard work by the

[More...](#)

Entrez Search

In the search bar look for something in “All Databases” that interests you: protein, organism, compound, etc.



The screenshot shows the NCBI Entrez Search homepage. At the top, the NIH logo and "National Library of Medicine" text are on the left, and a user profile "shawnpolson" is on the right. Below this is a search bar with a dropdown menu set to "All Databases" and the text "Escherichia coli" entered. A "Search" button is to the right of the search bar. On the left side, there is a vertical navigation menu with links: "NCBI Home", "Resource List (A-Z)", "All Resources", "Chemicals & Bioassays", "Data & Software", "DNA & RNA", "Domains & Structures", "Genes & Expression", "Genetics & Medicine", "Genomes & Maps", "Homology", "Literature", "Proteins", and "Sequence Analysis". The main content area is titled "Welcome to NCBI" and includes a paragraph about the center's mission. Below this are three columns: "Submit" (with an icon of an arrow pointing up), "Download" (with an icon of an arrow pointing down), and "Learn" (with an icon of books). On the right side, there is a "Popular Resources" section with links to PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. At the bottom right, there is a "NCBI News & Blog" section.

NIH National Library of Medicine
National Center for Biotechnology Information

shawnpolson

All Databases Escherichia coli Search

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis

Welcome to NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Popular Resources
PubMed
Bookshelf
PubMed Central
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI News & Blog
New RefSeq Annotation

Results give a good summary of the different types of information that NCBI houses.

Take a few minutes and explore some of these links to see where they go...

Search NCBI

Escherichia coli

×

Search

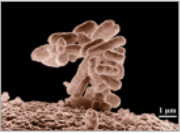
Results found in 32 databases

TAXONOMY


Was this helpful?

👍

👎



Escherichia coli
E. coli (*Escherichia coli*) is a species of enterobacteria in the family *Enterobacteriaceae* (enterobacteria).
Taxonomy ID: 562

 Genomes
Browse and download

Literature

Bookshelf4,483

MeSH66

NLM Catalog677

PubMed410,988

PubMed Central569,531

Genes

Gene108,352

GEO DataSets40,922

GEO Profiles382,402

HomoloGene42

PopSet5,743

Proteins

Conserved Domains1,411

Identical Protein Groups13,497,230

Protein84,210,753

Protein Family Models7,918

Structure65,447

Genomes

Assembly161,063

BioCollections0

BioProject11,058

BioSample313,471

Genome1

Nucleotide12,471,104

SRA311,592

Taxonomy1

Clinical

ClinicalTrials.gov560

ClinVar15

dbGaP13

dbSNP0

dbVar0

GTR3

MedGen40

OMIM162

PubChem



BioAssays53,088

Compounds12

Pathways2,986

Substances379

The taxonomy database gives a good overview of available information for a given organism or taxonomic rank



EntrezPubMedNucleotideProteinGenomeStructurePMCTaxonomyBioCollections

Search for as ☐ complete name ☐ lock

Display 3 levels using filter: none

Escherichia coli

Taxonomy ID: 562 (for references in articles please use NCBI:txid562)

current name

Escherichia coli (Migula 1895) Castellani and Chalmers 1919
neotype strain of *Escherichia coli*: [ATCC:11775](#), [CCUG:24](#), [CCUG:29300](#), [CIP:54.8](#), [DSM:30083](#), [JCM:1649](#), [BCCM/LMG:2092](#), [NBRC:102203](#), [NCCB:54008](#), [NCTC:9001](#), [IAM:12119](#), personal::U5/41
homotypic synonym: ☐ "***Bacillus coli***" Migula 1895, effective name [1](#)
"***Bacterium coli***" (Migula 1895) Lehmann and Neumann 1896, effective name [1](#)
Enterococcus coli
includes: ☐ *Achromobacter* sp. ATCC 35328
equivalent: ***Escherichia/Shigella coli***

NCBI BLAST name: **enterobacteria**
Rank: **species**
Genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)
Host: bacterialvertebrates
Other names:

heterotypic synonym

"***Bacterium coli commune***" Escherich 1885, effective name [1](#)

common name(s)

E. coli

[Lineage\(full \)](#)
[cellular organisms](#); [Bacteria](#); [Proteobacteria](#); [Gammaproteobacteria](#); [Enterobacterales](#); [Enterobacteriaceae](#); [Escherichia](#)

Entrez records			
Database name	Subtree links	Direct links	Links from type
Nucleotide	10,396,505	9,110,887	417
Protein	67,753,668	54,761,534	-
Structure	11,486	6,572	-
Genome	1	1	-
Popset	5,413	5,402	-
Conserved Domains	53	39	-
GEO Datasets	19,572	11,872	-
PubMed Central	392,486	392,479	-
Gene	383,511	58,900	-
SRA Experiments	277,605	249,713	-
GEO Profiles	313,301	298,677	-
Protein Clusters	17,571	17,571	-
Identical Protein Groups	13,496,139	13,114,092	-
Bio Project	10,227	6,432	-
Bio Sample	305,984	279,213	47
Bio Systems	22,561	1,814	-
Assembly	161,063	153,881	-
Genetic Testing Registry	3	-	-
Probe	2,210	1,804	-
PubChem BioAssay	20,365	18,536	-
Taxonomy	3,378	1	-

BLAST Search

- Perhaps the most commonly used analysis tool at NCBI
- Allows you to search by sequence
- Performs a “local alignment” allowing partial matches to be detected
 - Why would this be important?
- Several different types of BLAST are available that allow nucleotide or protein sequences against nucleotide or sequence databases



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

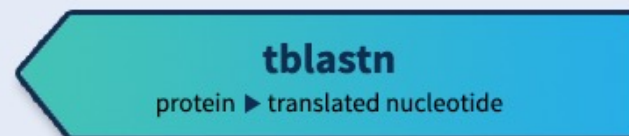
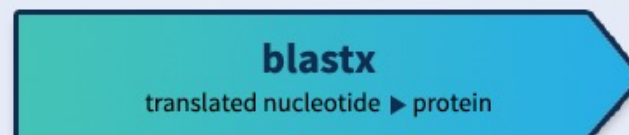
BLAST+ 2.13.0 is here!

Starting with this release, we are including the blastn_vdb and tblastn_vdb executables in the BLAST+ distribution.

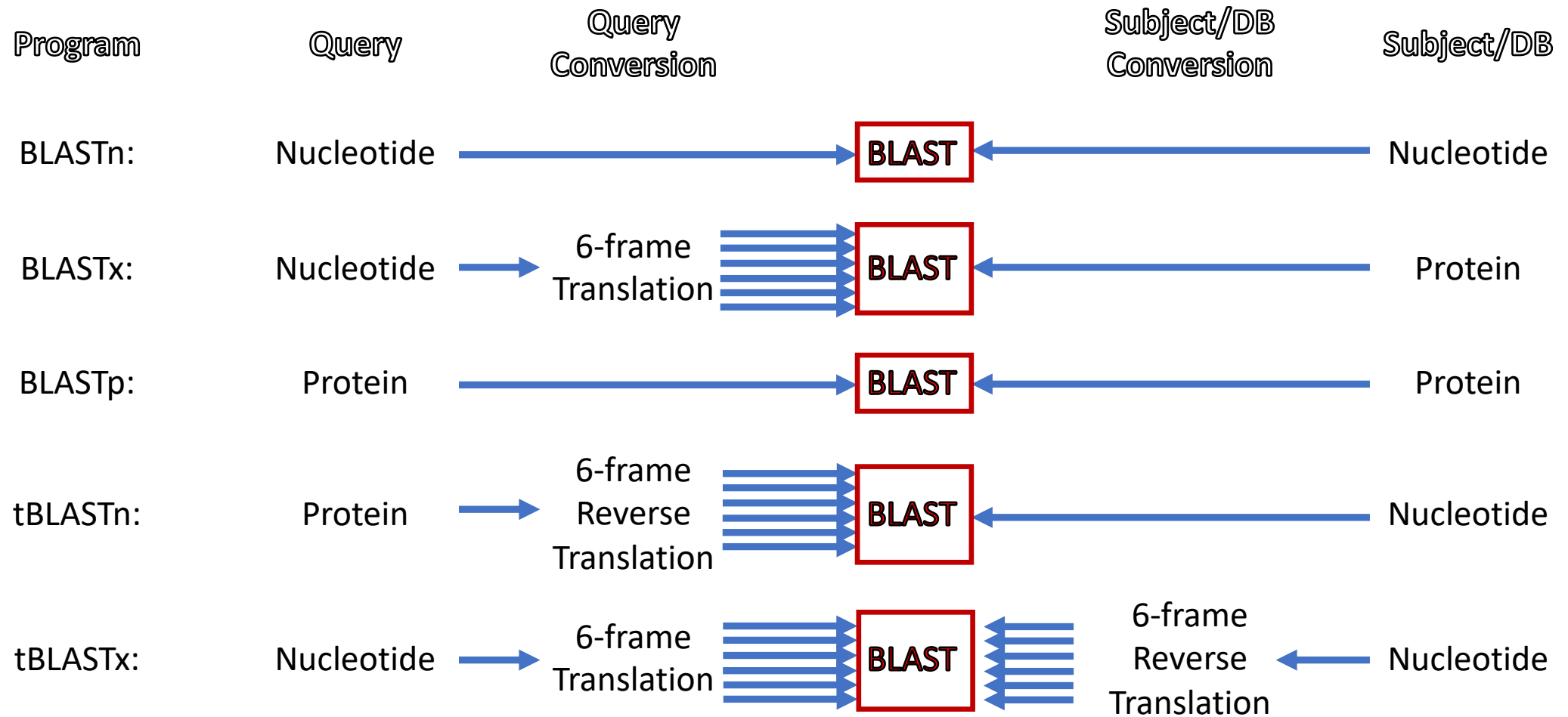
Thu, 17 Mar 2022 12:00:00 EST

[More BLAST news...](#)

Web BLAST



BLAST Programs

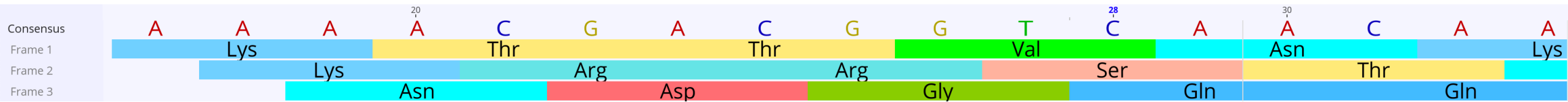


6-frame translation

Consensus A A A A²⁰ C G A C G G T | C²⁸ A A³⁰ C A A

UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA } Stop UGG } Trp
CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }
AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }
GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }

6-frame translation



UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA } Stop UGG } Trp
CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }
AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }
GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }

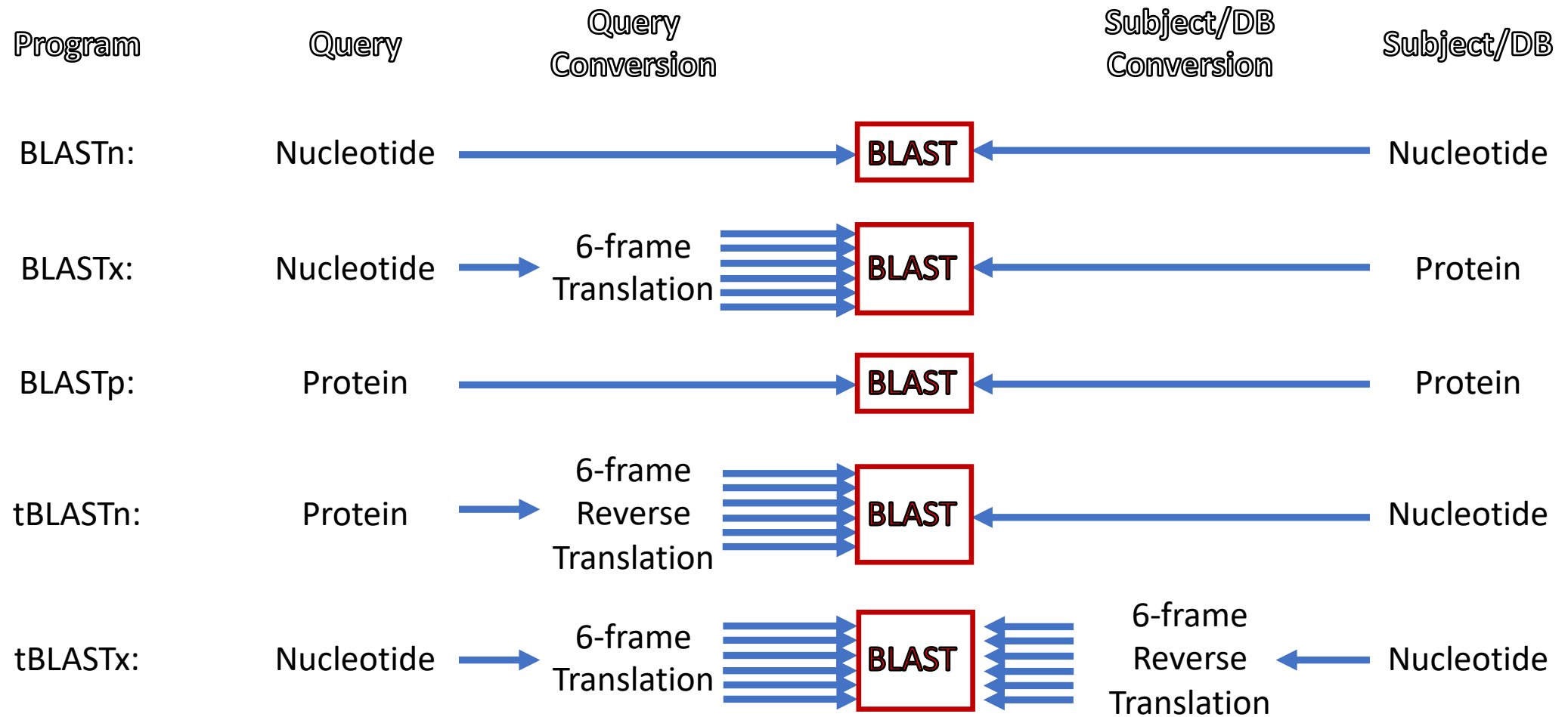
6-frame translation

Consensus	A A A A C G A C G G T C A												A C A A				
Frame 1	Lys			Thr				Thr		Val			Asn		Lys		
Frame 2	Lys			Arg				Arg			Ser			Thr			Lys
Frame 3	Asn			Asp				Gly			Gln			Gln			
Complement	T	T	T	T	G	C	T	G	C	C	A	G	T	T	G	T	T

6-frame translation

Consensus	A A A A ²⁰ C G A C G G T C ²⁸ A A ³⁰ C A A																		
Frame 1	Lys				Thr				Val				Asn		Lys				
Frame 2	Lys				Arg				Arg				Ser		Thr				Lys
Frame 3	Asn				Asp				Gly				Gln		Gln				
Complement	T T T T G C T G C C A G T T G T T																		
Reverse Frame 1	Phe				Arg				Asp				Val		Phe				
Reverse Frame 2	Val				Val				Thr				Leu				Leu		
Reverse Frame 3	Phe				Ser				Pro				*		Cys				

BLAST Programs



Let's try a Nucleotide BLAST (BLASTn)

```
ATGAAACTGCAACAACCTGAAGTACATTGTTGAAGTCGTTAATCACAATCTAAATGTCTCGGCGACCGCCG
AAAGCTTGTATACCTCTCAGCCAGGGATCAGTAAGCAAGTCAGGCTTTTAGAAGATGAGTTAGGCATTCA
GATATTTGAACGAAGTGGTAAACACCTCACTCAAGTGAAGTCGAGCTGGGGAAGATATCGTGCGAATTTTCG
CAAGAAATTTTAGCTCGAGTTGAGAGTATTAAAGCTGTGGCAGGTGAGCATACTCATCCTGAGATGGGCA
CACTGAACATTTCTACCACCCATACACAGGCTCGTTATGCTTTACCGGATGTCATTAAAGGGTTTGTAAA
GCGTTACCCGAAGGTCTCCTTGACACATGCACCAAGGGACGCCAAGTCAAATGTCCGAAGCGATTGCAAAA
GGCACGGCGAATTTTTCGATTGCGACAGAAGCTCTGCATTTGTACCAAGATGCGATTATGTTGCCTTGTT
ACCATTTGGAATCGTTTCGATTGTTGTACCTAAAGAGCATCCATTGGCTAAGAAGGAAAAAGTGACTATTCA
AGATTTGGCTTCTTATCCATTAGTGACATATGTGTTTGGTTTTACCGGTCGCTCGGAGCTTGACACCGCA
TTTAATAGAGTAGGATTAACGCCGCGAGTGGTGTTTACTGCAACCGACGCTGATGTTATCAAAACTTATG
TACGTATGGGGATTGGTGTGGGTGTGATTGCCAGCATGGCGGTAGATATAGAGCAGGATAAAGATCTGGT
CTCGATCGATGCGAGCCACTTATTTGGCGCAACTACCACTAGTATCGGCTTCCGCCGAGGTACGTTCCCTA
CGTTCATATATGTTTGTATTTTATGGAAAGATTTGCCCTCATTTTGACAAGGCCAGTGGTAGAGCAGGCGA
TCTCGTTGAAATCAAATGCTGAGATTGATGAAATGTTTAAAGACATCGTTCTTCCCGTTCGTTAA
```


Standard Nucleotide BLAST

blastn

blastp

blastx

tblastn

tblastx

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

Enter Query Sequence

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

```
ATGAAACTGCAACAACTGAAGTACATTGTTGAAGTCGTTAATCACAATCT
AAATGTCTCGGCGACCGCCG
AAAGCTTGATACCTCTCAGCCAGGGATCAGTAAGCAAGTCAGGCTTTT
AGAAGATGAGTTAGGCATTCA
GATATTTGAACGAAGTGGTAAACACCTCACTCAAGTGACTCGAGCTGG
GGAAGATATCGTGCGAATTCG
CAAGAAATTTAGCTCGAGTTGAGAGTATTAAGCTGTGGCAGGTGAGC
ATACTCATCCTGAGATGGGCA
CACTGAACATTTCTACCACCCATACACAGGCTCGTTATGCTTTACCGGAT
GTCATTAAAGGGTTTGTA
GCGTTACCCGAAGGTCTCCTTGACACATGCACCAAGGGACGCCAAGTCA
AATGTCCGAAGCGATTGCAAAA
GGCACGGCGAATTTTGCGATTGCGACAGAAGCTCTGCATTGTACCAA
GATGCGATTATGTTGCCTTGTT
ACCATTGGAATCGTTGATTGTTGACCTAAAGAGCATCCATTGGCTAAG
AAGGAAAAAGTGACTATTCA
ACATTGGCTTCTATCCATTAGTCACTTATGTTTGCTTTACCGGTC
```

Query subrange [?](#)From To

Or, upload file

no file selected

[?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)☐ Align two or more sequences [?](#)

Choose Search Set

Database

☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ BetacoronavirusNucleotide collection (nr/nt) [?](#)

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) ☐ Uncultured/environmental sample sequences

Limit to

Optional

☐ Sequences from type material

Entrez Query

Optional

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

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

Descriptions	Graphic Summary	Alignments	Taxonomy
--------------	-----------------	------------	----------

Sequences producing significant alignments									
Download ▾					Select columns ▾		Show	100 ▾	?
<input checked="" type="checkbox"/> select all 100 sequences selected					GenBank Graphics Distance tree of results MSA Viewer				
	Description ▾	Scientific Name ▾	Max Score ▾	Total Score ▾	Query Cover ▾	E value ▾	Per. Ident ▾	Acc. Len ▾	Accession
<input checked="" type="checkbox"/>	Vibrio vulnificus strain FORC_037 chromosome I, complete sequence	Vibrio vulnificus	1801	1801	100%	0.0	100.00%	3224939	CP016321.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain 06-2410 chromosome 1, complete sequence	Vibrio vulnificus	1796	1796	100%	0.0	99.90%	3225605	CP046832.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain 07-2444 chromosome 1, complete sequence	Vibrio vulnificus	1773	1773	100%	0.0	99.49%	3472667	CP046835.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain CECT 4999 chromosome I, complete sequence	Vibrio vulnificus	1773	1773	100%	0.0	99.49%	3394464	CP014636.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain 2015AW-0208 chromosome 1, complete sequence	Vibrio vulnificus	1762	1762	100%	0.0	99.28%	3418330	CP046758.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain FORC_054 chromosome 1, complete sequence	Vibrio vulnificus	1762	1762	100%	0.0	99.28%	3311092	CP019121.1
<input checked="" type="checkbox"/>	Vibrio vulnificus Env1 chromosome I, complete sequence	Vibrio vulnificus Env1	1757	1757	100%	0.0	99.18%	3241343	CP017635.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain 2497-87 chromosome 1, complete sequence	Vibrio vulnificus	1757	1757	100%	0.0	99.18%	3260230	CP060047.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain 2009V-1035 chromosome 1	Vibrio vulnificus	1751	1751	100%	0.0	99.08%	3317415	CP035784.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain FDAARGOS_663 chromosome 2, complete sequence	Vibrio vulnificus	1751	1751	100%	0.0	99.08%	3249353	CP044069.1
<input checked="" type="checkbox"/>	Vibrio vulnificus NBRC 15645 = ATCC 27562 chromosome 1, complete sequence	Vibrio vulnificus NBRC 1...	1751	1751	100%	0.0	99.08%	3266118	CP012881.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain FORC_077 chromosome 1, complete sequence	Vibrio vulnificus	1613	1613	100%	0.0	96.51%	3217678	CP027030.1
<input checked="" type="checkbox"/>	Vibrio vulnificus MO6-24/O chromosome I, complete sequence	Vibrio vulnificus MO6-24/O	1607	1607	100%	0.0	96.41%	3194232	CP002469.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain FORC_017 chromosome 1, complete sequence	Vibrio vulnificus	1591	1591	100%	0.0	96.10%	3253417	CP012739.1
<input checked="" type="checkbox"/>	Vibrio vulnificus strain Vv180806 chromosome 1, complete sequence	Vibrio vulnificus	1591	1591	100%	0.0	96.10%	3400045	CP044206.1

We'll briefly discuss what some of the scores and other metrics mean, but for more information check out the YouTube videos that NCBI has put together at the link below. Highly recommend the short Expect Value videos.

<https://www.youtube.com/playlist?list=PL7dF9e2qSW0azL2xOKAtxDW7QI8UU4XZ6>

BLAST Terminology

- **Query** – Sequence you searched for
- **Subject** – Sequence that a query aligns to (i.e. sequence from database)
- **Scoring Matrix** – A table for how likely different substitutions are. Allows customization in how different mismatches are penalized. (e.g. lower penalty for substitution of biochemically similar amino acids versus dissimilar)
- **Raw Score** – Simple scoring of an alignment to reward matches/penalize mismatches for each position (uses a substitution matrix)

BLAST Terminology

- **Bit Score** – Score normalized for database size and use of different scoring matrices. Allow scores of the same query to be compared across searches. (High is good)
- **Expect (E) value** – The number of alignments with scores \geq to a given match that would occur by random chance. (Very low is good)
- **Coverage** – Percentage of Query/Subject that were covered by the alignment
- **Identity** – Percentage of aligned nucleotides/amino acids that were exact matches
- **Similarity** - Percentage of aligned amino acids that were positive matches (exact matches or a substitution deemed likely (similar) by the substitution matrix)

Now try these variations

- Repeat the BLASTn with only part of the query sequence
- Repeat the BLASTn with the RefSeq Representative Genome database for Organism “Bacteria (taxid:2)”

For discussion: Make note of how these variations change the results (both the hits themselves and the scores)

Let's try a BLASTx with the same sequence...

- TIP: Since we know something about what type of organism has this protein we can speed up the search by limiting the “Organism”
- What do these results tell you that was not apparent with BLASTn?

Optional Exercise:

Try BLASTp, tBLASTx, and tBLASTn

MKTYNVAIVGASGAVGQELIKGLENSFFPIKKFVPLASTRSAGKKIKAFNKDYEILETTH
EVFEREKIDIAFFSAGGSVSEEFATSASKTALVVDNTSFFRLNKDVPLVVPEINAKEIFN
APLNIIANPNCSTIQMTQILNPLHLHFKIKSVIVSTYQAVSGAGNKGIESLKNELKTALE
CLEKDPTIDLNQVLQAGAFAYPIAFNAIAHIDTFKENGYTKEELKMLHETHKIMGVDFPI
SATCVRVPVLRSHSESLSIAFEKEFDLKEVYEVLKNAPSVAVCDDPSHNLYPTPLKASHT
DSVFIGRLRKDLFDKKTLHGFCVADQLRVGAATNALKIALHYIKNA

Universal Protein Resource (UniProt)


- Comprehensive resource for protein sequence and annotation data
- Collaboration between the European Bioinformatics Institute (EMBL-EBI), the SIB Swiss Institute of Bioinformatics and the Protein Information Resource (PIR)
- We will not do a comprehensive overview of UniProt resources today, please check their training resources:
<https://www.youtube.com/user/uniprotvideos>

The mission of [UniProt](#) is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB


UniProt Knowledgebase

Swiss-Prot (566,996)

 Manually annotated and reviewed.


Records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL (230,328,648)

 Automatically annotated and not reviewed.


Records that await full manual annotation.

UniRef




The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.

UniParc

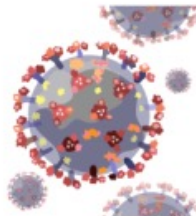


UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

Proteomes







A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.



New UniProt portal for the latest SARS-CoV-2 coronavirus protein entries and receptors, updated independent of the general UniProt release cycle.

View SARS-CoV-2 Proteins and Receptors

News

Forthcoming changes
Planned changes for UniProt

UniProt release 2022_01
A phospholipase for clear vision | Cross-references to MANE-Select

UniProt release 2021_04
ZTGC: bacteriophages reinvent the DNA alphabet

UniProt release 2021_03
The importance of being disordered | MobiDB-lite predictions for

[News archive](#)

Supporting data

Literature citations

Cross-ref. databases

Taxonomy

Diseases

Subcellular locations

Keywords

[Getting started](#)

 [YouTube](#)

[UniProt data](#)


[Protein spotlight](#)


<https://www.uniprot.org>

UniProtKB 2022_01 results

Basket

UniProtKB consists of two sections:

 **Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.

 **Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

 [Help](#)  [UniProtKB help video](#)  [Other tutorials and videos](#)  [Downloads](#)

Filter byⁱ

 **Unreviewed (2)**
TrEMBL

Popular organisms
VIBVL (1)

VIBVU (1)

BLASTAlignDownloadAdd to basketColumns

1 to 2 of 2Show250

<input type="checkbox"/>	Entry	Entry name	<input type="checkbox"/>	Protein names	Gene names	Organism	Length	
<input type="checkbox"/>	A0A087JF97	A0A087JF97_VIBVL		Cys regulon transcriptional activat...	cysB CRN52_15695, D8T54_16205, D8T65_03355, FORC53_1863, I7730_22205	Vibrio vulnificus	324	
<input type="checkbox"/>	A0A3Q0L6S0	A0A3Q0L6S0_VIBVU		Regulatory protein CysB	VV1_2954	Vibrio vulnificus (strain CMCP6)	324	

1 to 2 of 2Show250

https://www.uniprot.org/uniprot/?query=WP_011080770&sort=score

Reviewed vs Unreviewed Entries

- **Unreviewed** (TrEMBL): Have been automatically annotated by a computational algorithm.
- **Reviewed** (Swiss-Prot): Have been manually annotated by a sequence curator and determined to be of high quality based on experimental evidence. In addition to the increased quality/confidence of annotation, these entries often have more types of information than TrEMBL entries.

Broaden the Search to find Reviewed Entries



UniProtKB

Advanced

Search

Searching in UniProtKB

Help



Term

Gene name [GN]

CysB



Term

Taxonomy [OC]

Gammaproteobacteria



Term

All

Term

All

Gammaproteobacteria (9GAMM) [1236]

Gammaproteobacteria bacterium (ex Lamellibrachia satsuma) [2496632]

Gammaproteobacteria bacterium 3B [2690150]

Gammaproteobacteria bacterium AH-315-C21 [2814549]

Gammaproteobacteria bacterium AH-315-M22 [2814552]

Gammaproteobacteria bacterium 3C [2690151]

Gammaproteobacteria bacterium JGI_MCM16TB006 [2663653]

Gammaproteobacteria bacterium LE1 [2690156]


Gammaproteobacteria bacterium AH-315-K14 [2814551]

Gammaproteobacteria bacterium JGI 01_F9_750m [2738761]

Gammaproteobacteria bacterium Biozentrum3 [2690152]

Search

Filter to Swiss-Prot Entries





UniProtKB Advanced Search

BLAST Align Retrieve/ID mapping Peptide search SPARQL Help Contact

UniProtKB 2022_01 results Basket

UniProtKB consists of two sections:



**Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.


**Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Help UniProtKB help video Other tutorials and videos Downloads

Filter byⁱ

 Reviewed (7) 
Swiss-Prot

 Unreviewed (3,078)
TrEMBL

Popular organisms

E. coli K12 (1)

KLEPN (22)

SALTY (1)

Haemophilus influenzae PittII (1)






HAEI3 (1)

Other organisms Go

BLAST Align Download Add to basket Columns

1 to 250 of 3,085 Show 250

Show only exact matches for cysb

<input type="checkbox"/>	Entry	Entry name		Protein names	Gene names	Organism	Length	
<input type="checkbox"/>	P0A9F3	CYSB_ECOLI		HTH-type transcriptional regulator ...	cysB b1275, JW1267	Escherichia coli (strain K12)	324	
<input type="checkbox"/>	P06614	CYSB_SALTY		HTH-type transcriptional regulator ...	cysB STM1713	Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)	324	
<input type="checkbox"/>	P45600	CYSB_KLEPN		HTH-type transcriptional regulator ...	cysB	Klebsiella pneumoniae	324	
<input type="checkbox"/>	A0A0H3PE80	A0A0H3PE80_HAEI3		Succinate--CoA ligase [ADP-forming]...	cysB sucC, CGSHI3655_05914	Haemophilus influenzae (strain NTHi 3655)	389	
<input type="checkbox"/>	A0A0E1SPA3	A0A0E1SPA3_HAEIF		Succinate--CoA ligase [ADP-	cysB sucC, CGSHIII_00999	Haemophilus influenzae PittII	389	



UniProtKB ▾

gene:cysb taxonomy:"Gammaproteobacteria [1236]" AND reviewed:yes



Advanced ▾

Search

[BLAST](#) [Align](#) [Retrieve/ID mapping](#) [Peptide search](#) [SPARQL](#)[Help](#) [Contact](#)

UniProtKB 2022_01 results

Basket ▾

UniProtKB consists of two sections:

**Reviewed (Swiss-Prot) - Manually annotated**

Records with information extracted from literature and curator-evaluated computational analysis.

**Unreviewed (TrEMBL) - Computationally analyzed**

Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

[? Help](#)[UniProtKB help video](#)[Other tutorials and videos](#)[Downloads](#)Filter byⁱ

Reviewed (7)

Swiss-Prot

Popular organisms

[E. coli K12 \(1\)](#)[KLEPN \(1\)](#)[HAEIN \(1\)](#)[SALTY \(1\)](#)[ECOL6 \(1\)](#)

Other organisms

Go

Search terms

[BLAST](#)[Align](#)[Download](#)[Add to basket](#)[Columns](#)

1 to 7 of 7

Show

250



<input type="checkbox"/>	Entry ▾	Entry name ▾		Protein names ▾	Gene names ▾	Organism ▾	Length ▾	
<input type="checkbox"/>	P0A9F3	CYSB_ECOLI		HTH-type transcriptional regulator ...	cysB b1275, JW1267	Escherichia coli (strain K12)	324	
<input type="checkbox"/>	P06614	CYSB_SALTY		HTH-type transcriptional regulator ...	cysB STM1713	Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)	324	
<input type="checkbox"/>	P45600	CYSB_KLEPN		HTH-type transcriptional regulator ...	cysB	Klebsiella pneumoniae	324	
<input type="checkbox"/>	P45105	CYSB_HAEIN		HTH-type transcriptional regulator ...	cysB HI_1200	Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd)	323	
<input type="checkbox"/>	P0A9F4	CYSB_ECOL6		HTH-type transcriptional regulator ...	cysB c1742	Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC)	324	
<input type="checkbox"/>	P0A9F5	CYSB_ECO57		HTH-type transcriptional regulator ...	cysB Z2535, ECs1847	Escherichia coli O157:H7	324	
<input type="checkbox"/>	P52675	CYSB_THIRO		HTH-type transcriptional regulator ...	cysB	Thiocapsa roseopersicina	136	

UniProtKB - P0A9F3 (CYSB_ECOLI)

[Basket](#) ▾

Display

[Help video](#)[BLAST](#)[Align](#)[Format](#)[Add to basket](#)[History](#)[Add a publication](#)[Feedback](#)

Entry

Publications

Feature viewer

Feature table

Protein | **HTH-type transcriptional regulator CysB**Gene | **cysB**Organism | *Escherichia coli* (strain K12)Status |  Reviewed - Annotation score: ●●●●○ - Protein inferred from homologyⁱFunctionⁱ

This protein is a positive regulator of gene expression for the cysteine regulon, a system of 10 or more loci involved in the biosynthesis of L-cysteine from inorganic sulfate. The inducer for CysB is N-acetylserine. CysB inhibits its own transcription.

Regions

Feature key	Position(s)	Description	Actions	Graphical view	Length
DNA binding ⁱ	19 – 38	H-T-H motif PROSITE-ProRule annotation ▾	Add BLAST		20

GO - Molecular functionⁱ

- DNA binding [Source: EcoCyc ▾](#)
- DNA-binding transcription factor activity [Source: InterPro](#)
- Identical protein binding [Source: EcoCyc ▾](#)
- transcription cis-regulatory region binding [Source: GO_Central ▾](#)

[Complete GO annotation on QuickGO ...](#)GO - Biological processⁱ

- cysteine biosynthetic process [Source: GO_Central ▾](#)
- regulation of transcription, DNA-templated [Source: GO_Central ▾](#)
- response to X-ray [Source: EcoCyc ▾](#)

[Complete GO annotation on QuickGO ...](#)Keywordsⁱ

Molecular function	Activator , DNA-binding
Biological process	Amino-acid biosynthesis , Cysteine biosynthesis , Transcription , Transcription regulation

Enzyme and pathway databases

BioCyc ⁱ	EcoCyc:PD00232
---------------------	--------------------------------

Names & Taxonomyⁱ

None

☒ Function☒ Names & Taxonomy☒ Subcellular location☐ Pathology & Biotech☒ PTM / Processing☐ Expression☒ Interaction☒ Structure☒ Family & Domains☒ Sequence☒ Similar proteins☒ Cross-references☒ Entry information☒ Miscellaneous[▲ Top](#)