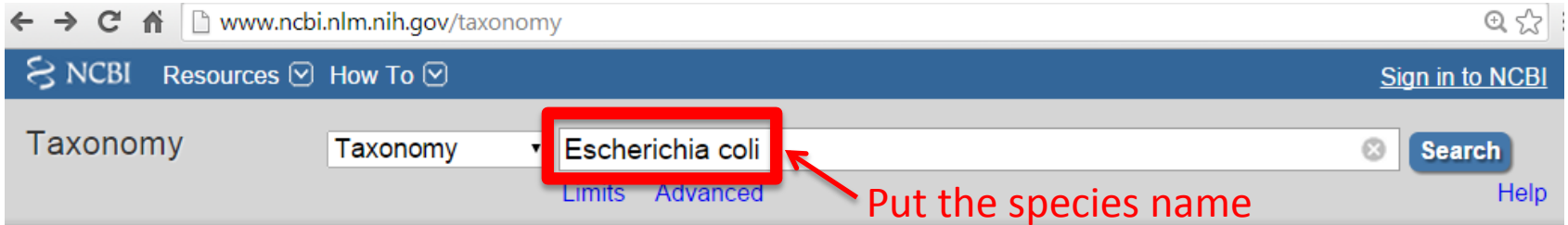


This document will show you:

With species name, how to search NCBI Taxonomy database website to locate taxonomy ID?

With protein accession number, how to search NCBI protein database to locate taxonomy ID?

Here we use *Escherichia coli* K12 MG1655 as an example to show how to search NCBI taxonomy database for its Taxonomy ID




← → ↻ 🏠 🔍 ☆

NCBI Resources How To Sign in to NCBI

Taxonomy Taxonomy Search Help

[Limits](#) [Advanced](#) Put the species name



Taxonomy

The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.

Using Taxonomy

[Quick Start Guide](#)

[FAQ](#)

[Handbook](#)

[Taxonomy FTP](#)

Taxonomy Tools

[Browser](#)

[Common Tree](#)

[Statistics](#)

[Name/ID Status](#)

[Genetic Codes](#)

[Linking to Taxonomy](#)

[Extinct Organisms](#)

Other Resources

[GenBank](#)

[LinkOut](#)

[E-Utilities](#)

[Batch Entrez](#)

[INSDC](#)

Taxonomy

Taxonomy

Escherichia coli |

[Save search](#) [Limits](#) [Advanced](#)

[Display Settings:](#) Summary

[Send to:](#)

[Escherichia coli](#)

species, enterobacteria

[Nucleotide](#) [Protein](#)

[Rel](#)
[Nuc](#)
[Pro](#)
[Ass](#)

Click on the link

[Lineage](#) (full): [root](#): [cellular organisms](#): [Bacteria](#): [Proteobacteria](#): [Gammaproteobacteria](#): [Escherichia](#)

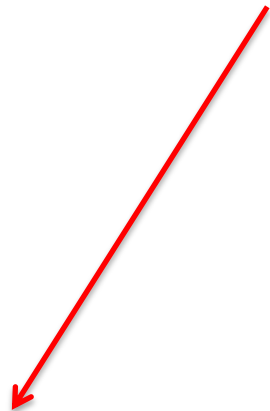
◦ [Escherichia coli](#) *Click on organism name to get more information.*

- [Escherichia coli #1/H766](#)
- [Escherichia coli 'BL21-Gold\(DE3\)pLysS AG'](#)
- [Escherichia coli 0.1197](#)
- [Escherichia coli 0.1288](#)
- [Escherichia coli 0.1304](#)
- [Escherichia coli 07798](#)
- [Escherichia coli 08BKT055439](#)
- [Escherichia coli 08BKT77219](#)
- [Escherichia coli 09BKT024447](#)
- [Escherichia coli 09BKT076207](#)
- [Escherichia coli 09BKT078844](#)
- [Escherichia coli 1-110-08 S1 C1](#)
- [Escherichia coli 1-110-08 S1 C2](#)

There are hundreds of ecoli strains
Scroll down or search in this HTML page (Ctrl+f) to
look for MG1655 strain

- [Escherichia coli LW1655F+](#)
- [Escherichia coli MC1061](#)
- [Escherichia coli NC-7](#)
- [Escherichia coli str. K-12 substr. DH10B](#)
- [Escherichia coli str. K-12 substr. MC4100](#)
- [Escherichia coli str. K-12 substr. MDS42](#)
- [Escherichia coli str. K-12 substr. **MG1655**](#)
 - [Escherichia coli C321.deltaA](#)
 - [Escherichia coli MV1994](#)
 - [Escherichia coli MV2A6](#)

mg1655 1 of 2 ^ v x



There are even many different medical isolates of MG1655
Ignore them and just hit the strain name

Lineage (full): [root](#); [cellular organisms](#); [Bacteria](#); [Proteobacteria](#); [Gammaproteobacteria](#); [Enterobac](#)
[Escherichia](#); [Escherichia coli](#); [Escherichia coli K-12](#)

- [Escherichia coli str. K-12 substr. MG1655](#) *Click on organism name to get more information.*
 - [Escherichia coli C321.deltaA](#)
 - [Escherichia coli MV1994](#)
 - [Escherichia coli MV2 Δ6](#)

Here is the taxonomy lineage description page of MG1655
This is the Taxonomy ID. This should be copied and paste to
ORFanFinder job submission page

Escherichia coli str. K-12 substr. MG1655

Taxonomy ID: 511145

Inherited blast name: **enterobacteria**

Rank: no rank

Genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)

Other names:

synonym: **Escherichia coli MG1655**

equivalent name: **Escherichia coli strain MG1655**

equivalent name: **Escherichia coli str. MG1655**

equivalent name: **Escherichia coli str. K12 substr. MG1655**

[Lineage \(full\)](#)

[cellular organisms](#); [Bacteria](#); [Proteobacteria](#); [Gammaproteobacteria](#);
[Enterobacteriales](#); [Enterobacteriaceae](#); [Escherichia](#); [Escherichia coli](#);
[Escherichia coli K-12](#)

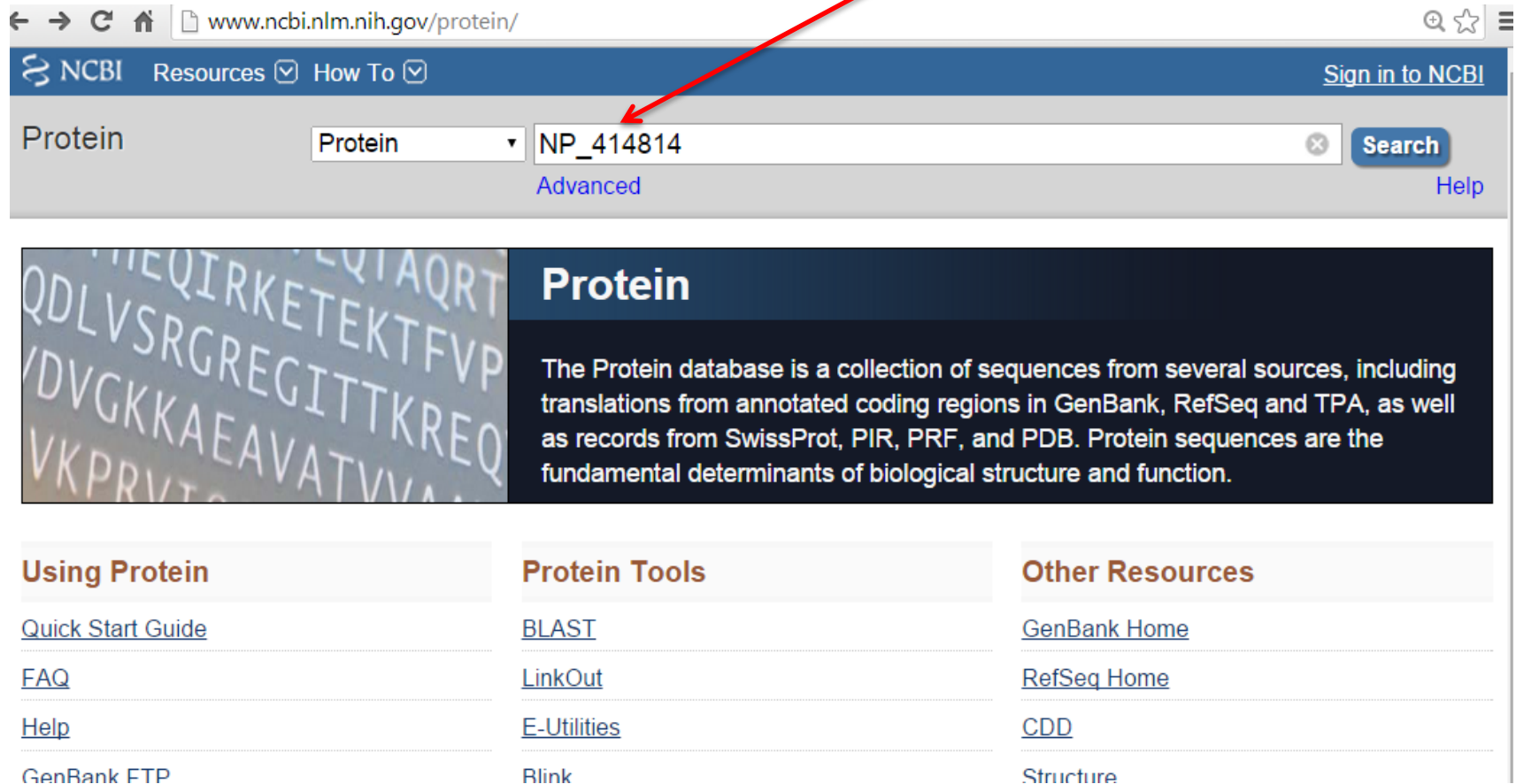
Entrez records	
Database name	Subtree li
Nucleotide	4
Nucleotide GSS	1
Protein	14.2
Structure	
Genome	
Popset	
GEO Datasets	1.9
PubMed Central	
Gene	4.5
SRA Experiments	9
Probe	2
Assembly	

Next:

With protein accession number, how to search
NCBI protein database to locate taxonomy ID?

Here we use a GenBank protein ID (NP_414814) as an example to show how to search NCBI protein database to get its Taxonomy ID

Put the protein ID here



www.ncbi.nlm.nih.gov/protein/

NCBI Resources How To Sign in to NCBI

Protein Protein Search

Advanced Help

Protein

The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.

Using Protein	Protein Tools	Other Resources
Quick Start Guide	BLAST	GenBank Home
FAQ	LinkOut	RefSeq Home
Help	E-Utilities	CDD
GenBank FTP	Blink	Structure

[Display Settings:](#) GenPept

[Send to:](#) (

CP4-6 prophage; uncharacterized protein [Escherichia coli str. K-12 substr. MG1655]

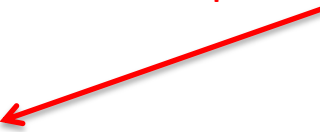
NCBI Reference Sequence: NP_414814.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS NP_414814 146 aa linear CON 29-OCT-2014
DEFINITION CP4-6 prophage; uncharacterized protein [Escherichia coli str. K-12 substr. MG1655].
ACCESSION NP_414814
VERSION NP_414814.1 GI:16128265
DBLINK BioProject: [PRJNA57779](#)
BioSample: [SAMN02604091](#)
DBSOURCE REFSEQ: accession [NC_000913.3](#)
KEYWORDS RefSeq.
SOURCE Escherichia coli str. K-12 substr. MG1655
ORGANISM [Escherichia coli str. K-12 substr. MG1655](#)
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (residues 1 to 146)
AUTHORS Riley,M., Abe,T., Arnaud,M.B., Berlyn,M.K., Blattner,F.R., Chaudhuri,R.R., Glasner,J.D., Horiuchi,T., Keseler,I.M., Kosuge,T.,

This is GenPept format view of the queried protein, which is of E. coli MG1655 strain. Click on this species name



There are even many different medical isolates of MG1655
Ignore them and just hit the strain name

Lineage (full): [root](#); [cellular organisms](#); [Bacteria](#); [Proteobacteria](#); [Gammaproteobacteria](#); [Enterobac](#)
[Escherichia](#); [Escherichia coli](#); [Escherichia coli K-12](#)

- [Escherichia coli str. K-12 substr. MG1655](#) *Click on organism name to get more information.*
 - [Escherichia coli C321.deltaA](#)
 - [Escherichia coli MV1994](#)
 - [Escherichia coli MV2 Δ6](#)

Here is the taxonomy lineage description page of MG1655
This is the Taxonomy ID. This should be copied and paste to
ORFanFinder job submission page

Escherichia coli str. K-12 substr. MG1655

Taxonomy ID: 511145

Inherited blast name: **enterobacteria**

Rank: no rank

Genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)

Other names:

synonym: **Escherichia coli MG1655**

equivalent name: **Escherichia coli strain MG1655**

equivalent name: **Escherichia coli str. MG1655**

equivalent name: **Escherichia coli str. K12 substr. MG1655**

Lineage(full)

[cellular organisms](#); [Bacteria](#); [Proteobacteria](#); [Gammaproteobacteria](#);
[Enterobacteriales](#); [Enterobacteriaceae](#); [Escherichia](#); [Escherichia coli](#);
[Escherichia coli K-12](#)

Entrez records	
Database name	Subtree li
Nucleotide	4
Nucleotide GSS	1
Protein	14.2
Structure	
Genome	
Popset	
GEO Datasets	1.9
PubMed Central	
Gene	4.5
SRA Experiments	9
Probe	2
Assembly	