

Biological Database Search and Retrieval of Nucleotide Sequence

Dr. Pradeep Kumar

Associate Professor

Department of Biotechnology

VBS Purvanchal University, Jaunpur

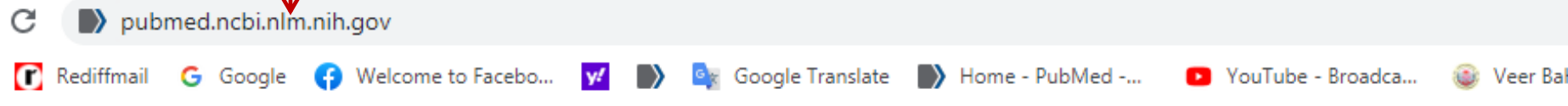
NCBI Nucleotide

(<https://www.ncbi.nlm.nih.gov/nucleotide/>)

- The National Centre for Biotechnology Information Nucleotide database has sequences from
- GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>)
- RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>)
- TPA (<https://www.ncbi.nlm.nih.gov/genbank/tpa/>)
- PDB (www.rcsb.org/pdb).

Step 1

Open any web browser and type/enter pubmed.ncbi.nlm.nih.gov in the address bar and Click enter ↵



The Google logo, consisting of the word "Google" in its characteristic multi-colored font (blue, red, yellow, blue, green, red).

Search Google or type a URL



pubmed.ncbi.nlm.nih.gov homepage will open

The screenshot shows the NCBI homepage with the following elements:

- Address Bar:** `ncbi.nlm.nih.gov` (highlighted with a red box)
- Navigation:** NCBI logo, "Resources" (checked), "How To" (checked), "Sign in to NCBI"
- Search:** "All Databases" dropdown, search input field, "Search" button
- COVID-19 Alert:** "COVID-19 is an emerging, rapidly evolving situation. Get the latest public health information from CDC: <https://www.coronavirus.gov>. Get the latest research from NIH: <https://www.nih.gov/coronavirus>. Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>."
- Left Sidebar (Resource List A-Z):**
 - 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
- Main Content:**
 - 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
- Right Sidebar:**
 - Popular Resources**
 - PubMed
 - Bookshelf
 - PubMed Central
 - BLAST
 - Nucleotide
 - Genome
 - SNP
 - Gene
 - Protein
 - PubChem
 - NCBI News & Blog**
 - NCBI Datasets now provides downloads of gene data for more than 30 thousand organisms 09 Oct 2020
 - NCBI Datasets now offers Gene tables: 08 Oct 2020
 - Recent enhancements in Genome Workbench version 3.5.0 08 Oct 2020
 - New Features Version 3.5.0 of Genome Workbench NCBI's sequence annotation 08 Oct 2020

Step 2

Click on dropdown menu(All Databases) and select Nucleotide database by clicking on Nucleotide

The screenshot shows the NCBI website interface. At the top left, the URL 'ncbi.nlm.nih.gov' is visible. The main navigation bar includes 'NCBI Resources' and 'How To'. A search bar is located on the right. A dropdown menu is open, showing a list of databases. The 'Nucleotide' option is highlighted with a red box, and a red arrow points to it from the text above. The dropdown menu includes the following items: Nucleotide, GTR, HomoloGene, Identical Protein Groups, MedGen, MeSH, NCBI Web Site, NLM Catalog, and Nucleotide. Below the dropdown menu, the main content area features a navigation sidebar on the left with categories like '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', and 'Variation'. The main content area has a central banner for COVID-19 information and several sections: 'Popular Resources' (PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem), 'NCBI News & Blog' (NCBI Datasets now provides downloads of gene data for more than 30 thousand organisms, NCBI Datasets now offers Gene tables, Recent enhancements in Genome Workbench version 3.5.0, New Features Version 3.5.0 of Genome Workbench NCBI's sequence annotation), and three action buttons: 'Submit' (Upload manuscripts and databases), 'Download' (Transfer NCBI data to your computer), and 'Learn' (Find help documents, attend a class or watch a tutorial). At the bottom, there are three more action buttons: 'Develop' (Use NCBI APIs and code libraries to build applications), 'Analyze' (Identify an NCBI tool for your data analysis task), and 'Research' (Explore NCBI research and collaborative projects).

Nucleotide database is selected

The screenshot shows the NCBI website interface. At the top, the URL is `ncbi.nlm.nih.gov`. The navigation bar includes the NCBI logo, "Resources", "How To", and "Sign in to NCBI". A search bar is present with a dropdown menu currently set to "Nucleotide", which is highlighted by a red box and a red arrow pointing from the title above. Below the search bar is a red banner with COVID-19 information. The main content area is divided into three columns: "Welcome to NCBI" with a navigation menu on the left, a central grid of action buttons (Submit, Download, Learn, Develop, Analyze, Research), and "Popular Resources" on the right. The "Nucleotide" link is visible in the "Popular Resources" list.

for Biotechnology X +

ncbi.nlm.nih.gov

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information

Nucleotide Search

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

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

Popular Resources
[PubMed](#)
[Bookshelf](#)
[PubMed Central](#)
[BLAST](#)
[Nucleotide](#)
[Genome](#)
[SNP](#)
[Gene](#)
[Protein](#)
[PubChem](#)

NCBI News & Blog
NCBI Datasets now provides downloads of gene data for more than 30 thousand organisms
09 Oct 2020
NCBI Datasets now offers Gene tables:
Recent enhancements in Genome Workbench version 3.5.0
08 Oct 2020
New Features Version 3.5.0 of Genome Workbench: NCBI's sequence annotation

Step 3

Enter the name of the gene in the search bar (MTHFR in this example) and click on search

The screenshot shows the NCBI website interface. At the top, the URL is ncbi.nlm.nih.gov. The navigation bar includes 'NCBI Resources' and 'How To'. The search bar is set to 'Nucleotide' and contains the text 'MTHFR'. The 'Search' button is circled in red. A red arrow points from the text 'Enter the name of the gene in the search bar' to the search bar, and another red arrow points from 'click on search' to the 'Search' button. Below the search bar is a COVID-19 notice. The main content area is titled 'Welcome to NCBI' and features a 'Resource List (A-Z)' sidebar on the left, a central 'Submit', 'Download', and 'Learn' section, and a 'Popular Resources' list on the right. The 'Submit' section includes 'Develop' and 'Analyze' options, while the 'Learn' section includes 'Research'.

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

Popular Resources
[PubMed](#)
[Bookshelf](#)
[PubMed Central](#)
[BLAST](#)
[Nucleotide](#)
[Genome](#)
[SNP](#)
[Gene](#)
[Protein](#)
[PubChem](#)

NCBI News & Blog
NCBI Datasets now provides downloads of gene data for more than 30 thousand organisms
09 Oct 2020
NCBI Datasets now offers Gene tables

<https://www.ncbi.nlm.nih.gov/nuccore/?term=MTHFR> page will open

Step 4

Click on the FASTA to obtain the nucleotide sequence of the *Mus musculus* MTHFR gene (Sequence no. 2 i.e. sequence of your interest in this example)

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide MTHFR Search

Create alert Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Species Summary 20 per page Sort by Default order Send to Filters: Manage Filters

Animals (2,437)
Plants (251)
Fungi (2,408)
Protists (256)
Bacteria (40,405)
Archaea (704)
Viruses (12)
Customize ...

Molecule types
genomic DNA/RNA (44,543)
mRNA (2,147)
Customize ...

Source databases
INSDC (GenBank) (44,268)
RefSeq (2,618)
Customize ...

Sequence Type
Nucleotide (46,893)
EST (4)
GSS (4)

Genetic compartments
Mitochondrion (1)
Plasmid (94)

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all
Show additional filters

GENE Was this helpful?

MTHFR – methylenetetrahydrofolate reductase

[Homo sapiens \(human\)](#)
GeneID: 4524

[RefSeq transcripts \(10\)](#) [RefSeq proteins \(9\)](#) [RefSeqGene \(1\)](#) [PubMed \(3,549\)](#)

Orthologs Genome Browser BLAST Download

RefSeq Sequences +

Items: 1 to 20 of 46901

<< First < Prev Page 1 of 2346 Next > Last >>

[Cervus elaphus hippelaphus isolate Hungarian chromosome 14, whole genome shotgun sequence](#)

1. 103,589,943 bp linear DNA
Accession: MKHE01000014.1 GI: 1207830612
[BioProject](#) [BioSample](#) [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[Mus musculus methylenetetrahydrofolate reductase gene, complete cds](#)

2. 4,181 bp linear DNA
Accession: AH007491.2 GI: 1036030392
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Results by taxon

Top Organisms [Tree]

[Acinetobacter baumannii \(6009\)](#)
[Salmonella enterica \(2896\)](#)
[Escherichia coli \(2747\)](#)
[Clostridioides difficile \(1818\)](#)
[Klebsiella pneumoniae \(971\)](#)
All other taxa (32460)
More...

Find related data

Database: Select

Find items

Search details

MTHFR[All Fields]

Search See more...

Recent activity

Turn Off Clear

MTHFR (46901) Nucleotide

abo (8921) Nucleotide

ABO (8921) Nucleotide

Total number of items available in the nucleotide database related to our search i.e. "MTHFR"

<https://www.ncbi.nlm.nih.gov/nuccore/?term=MTHFR>

Step 6

First click on file under **Choose destination** section and then on create file

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search

Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

FASTA

Mus musculus methylenetetrahydrofolate reductase gene, complete cds

GenBank: AH007491.2
[GenBank](#) [Graphics](#)

>AH007491.2 Mus musculus methylenetetrahydrofolate reductase gene, complete cds
GGGTTTGGTACCAGCCCTATAATACCCCGGCCCCACCCTCTACAGCAGGAATCCAGCCATGGTGAACG
AGGCCAGAGGAAGTGGCAGTCCCAACCCCGCATCTGAGGGCAGCAGCAGTGGCAGCGAGAGTTCCAAGGA
CAGTTCAAGATGTTCCACCCAGCCTGGACCAGAGCGGCAGAGAGACTCCGGGAGAAGATGAGGCGC
AGAATGGACTCTGGTGACAAGTGGTTCTCCCTGGAGTTCTTCCCCCTCGAACTGCTGAGGGAGCTGTTA
ACCTCATCTCCAGGTGAGTAGGGAGGTTAATCCGCGGGGTCTGGCAGGCTTCAGGGGAGCGTGNNNNNNN
NN
NN
NN
CAGGTTTGACCGGATGGCAGCAGGGGCCCCCTCTTCGTAGATGTTACCTGGCACCCAGCTGGAGACCCT
GGCTCAGACAAGGAGACCTCCTCATGATGATCGCCAGCACAGCAGTAACTACTGCGGCTTGGAAACCA
TCCTGCATATGACCTGCTGCCAGCAGCCCGGAGGAGATCACAGGCCATGTCACAGAGCCAAGCAGCT
GGGCC TGAAGAACATAATGGCGCTGAGGGGAGGTGTGGCCAGCACCCCTCTTTGGGTTCTTGCTT
TCCTGAAGGCTTNN
NN
NN
ATGACCTCAGCAACCTGCAGACCTGTAGGTGACCCTGGGAAGCAGAGGAAGGAGGCTTCAGCTATG

Send to: **Choose Destination**

- Complete Record
- Coding Sequences
- File
- Collections
- Clipboard
- Analysis Tool

Download 1 item.

Format: FASTA

Show GI

Create File

Articles about the Mthfr gene
Impaired folate 1-carbon metabolism causes formate-preventable hydroce [J Clin Invest. 2020]
One-carbon metabolism supplementation

Step 6

The file will be downloaded in your computer and will open in notepad
click on the downloaded file to open

The screenshot shows the NCBI Nucleotide search results page for the entry AH007491.2. The page title is "Mus musculus methylenetetrahydrofolate reductase gene, complete cds". The GenBank ID is AH007491.2. The FASTA sequence is displayed, and a download button for "sequence (7).fasta" is visible at the bottom left. A red box highlights the download button. A blue arrow points from the text above to the download button. A pink banner at the top of the page contains a COVID-19 warning and links to CDC and NIH resources.

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Help

Advanced

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

FASTA Send to: Change region shown Customize view

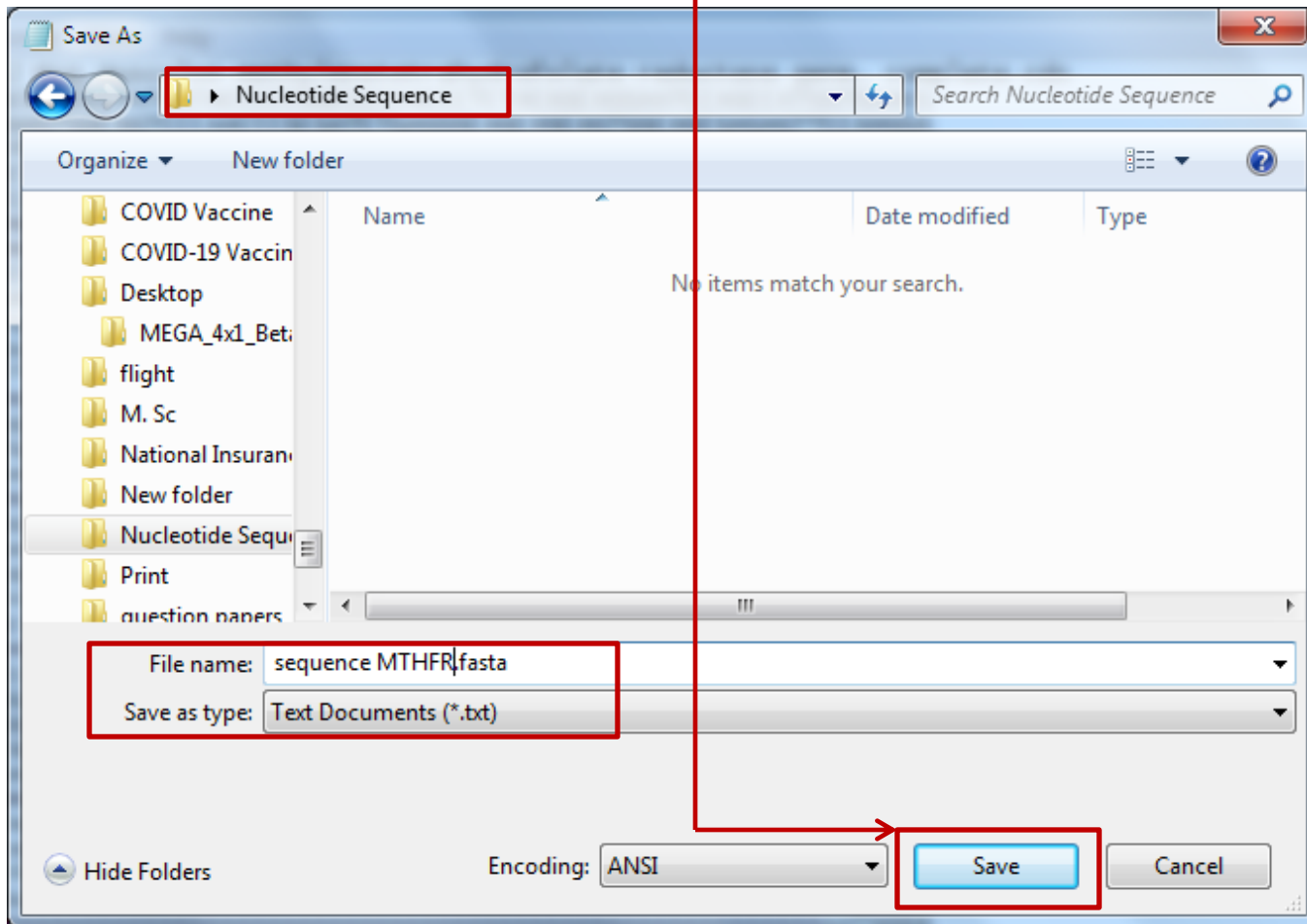
Mus musculus methylenetetrahydrofolate reductase gene, complete cds

GenBank: AH007491.2
[GenBank](#) [Graphics](#)

>AH007491.2 Mus musculus methylenetetrahydrofolate reductase gene, complete cds
GGGTTTGTTACCAGCCCTATAATACCCCGGCCCCACCCTCTACAGCAGGAATCCAGCCATGGTGAACG
AGGCCAGAGGAAGTGGCAGTCCCAACCCGCGATCTGAGGGCAGCAGCAGTGGCAGCGAGAGTTC AAGGA
CAGTTCAAGATGTTCCACCCCGAGCCTGGACCCAGAGCGGCACGAGAGACTCCGGGAGAAGATGAGGCGC
AGAATGGACTCTGGTGACAAGTGGTTCTCCCTGGAGTCTTCCCCCTCGAACTGCTGAGGGAGCTGTTA
ACCTCATCTCCAGGTGAGTAGGGAGGTTAATCCGCGGGGTGCGCAGGCTTCAGGGGAGCGTGNNNNNNN
NN
NN
NN
NN
CAGGTTTGACCGGATGGCAGCAGGGGGCCCTCTCGTAGATGTTACCTGGCACCCAGCTGGAGACCCCT
GGCTCAGACAAGGAGACTCCTCCATGATGATCGCCAGCAGCAGTAACTACGCGGCTGGAAACCA
TCCTGCATATGACCTGCTGCCAGCAGCGCCGGAGGAGATCACAGCCATCTGCACAGAGCCAAGCAGCT
GGGCTGAAGAACATAATGGCGCTGAGGGGAGGTGTGGCGCCAGCACCCCTCTTTGGGTTCTTGCTT
TCCTGAAGGCTTNN

sequence (7).fasta

The file will be saved in a specific folder in your computer.
You can rename the file (e.g “Sequence MTHFR.fasta” in this case)
Click on save to save the file



References/Acknowledgements

- pubmed.ncbi.nlm.nih.gov
- <https://www.ncbi.nlm.nih.gov/nucleotide/>
- <https://www.ncbi.nlm.nih.gov/genbank/>
- <https://www.ncbi.nlm.nih.gov/refseq/>
- <https://www.ncbi.nlm.nih.gov/genbank/tpa/>
- www.rcsb.org/pdb .
- <https://www.ncbi.nlm.nih.gov/nuccore/?term=MTHFR>