

NAME: ASIF ERFAN KHAN

ROLL NUMBER: 546

COURSE: MSc CS

SUBJECT: BIOINFORMATICS

**TOPIC: PERFORM A BLAST
SEARCH ON ANY GENE
SEQUENCE**

Practical No: 5

Aim: Perform a BLAST search on a specific gene sequence of a specify organism.

Steps:

Go to the National Center for Biotechnology Information Site

<https://www.ncbi.nlm.nih.gov/>

Select Nucleotide from All Databases and find any organism in a search bar

The screenshot displays the National Library of Medicine (NCBI) website. At the top, there is a navigation bar with the NCBI logo and a search bar. The search bar contains the text 'flavivirus'. Below the search bar, there is a list of search results. The first result is 'Mosquito flavivirus NS5 gene for polyprotein, partial cds, strain: YDFV/Sept/2013'. The second result is 'Mosquito flavivirus gene for polyprotein, complete cds, strain: YDFV/Oct/2013'. The third result is 'Hangzhou flavivirus 3 isolate YWL-FY92, complete genome'. The fourth result is 'Tembusu virus flavivirus polyprotein (flavivirus polyprotein gene) gene, complete cds'. The fifth result is 'Culex flavivirus gene for polyprotein, partial cds, clone: oepor_flavi_ns5_24'. The sixth result is 'Cacipacore virus flavivirus polyprotein (flavivirus polyprotein gene) and truncated polyprotein (flavivirus polyprotein gene) genes, complete cds'. Each result includes a checkbox, a title, a description, and links to various databases (GenBank, PubMed, Protein, Taxonomy, BioProject, BioSample, Assembly, FASTA, Graphics).

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 Streamlining ClinVar Submission of Assertion Criteria](#)
18 Nov 2022
ClinVar is a freely available submission-
Asian database for information about
[Re-evaluating the BLAST Nucleotide Database \(nt\)](#)
17 Nov 2022
The ongoing sequencing revolution has
resulted in exponential growth of the
[RefSeq Release 215](#)
15 Nov 2022
RefSeq Release 215 is available

☐ [Mosquito flavivirus NS5 gene for polyprotein, partial cds, strain: YDFV/Sept/2013](#)
4. 939 bp linear RNA
Accession: AB981167.1 GI: 824555718
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Mosquito flavivirus gene for polyprotein, complete cds, strain: YDFV/Oct/2013](#)
5. 10,863 bp linear RNA
Accession: AB981166.1 GI: 824555716
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Hangzhou flavivirus 3 isolate YWL-FY92, complete genome](#)
6. 9,631 bp linear RNA
Accession: MZ209680.1 GI: 2168955026
[BioProject](#) [BioSample](#) [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Tembusu virus flavivirus polyprotein \(flavivirus polyprotein gene\) gene, complete cds](#)
7. 10,990 bp linear RNA
Accession: NC_015843.2 GI: 381333820
[Assembly](#) [BioProject](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Culex flavivirus gene for polyprotein, partial cds, clone: oepor_flavi_ns5_24](#)
8. 268 bp linear RNA
Accession: LC227582.1 GI: 1247173336
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Cacipacore virus flavivirus polyprotein \(flavivirus polyprotein gene\) and truncated polyprotein \(flavivirus polyprotein gene\) genes, complete cds](#)
9. 10,284 bp linear RNA
Accession: NC_026523.1 GI: 785702599
[Assembly](#) [BioProject](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

https://www.ncbi.nlm.nih.gov/nucleotide/NC_015843.2

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

Nucleotide [Advanced](#) [Help](#)

GenBank

Tembusu virus flavivirus polyprotein (flavivirus polyprotein gene) gene, complete cds

NCBI Reference Sequence: NC_015843.2

[FASTA](#) [Graphics](#)

[Go to:](#) [\(C\)](#)

LOCUS NC_015843 10990 bp ss-RNA linear VRL 13-AUG-2018
DEFINITION Tembusu virus flavivirus polyprotein (flavivirus polyprotein gene) gene, complete cds.
ACCESSION NC_015843 NC_016958 NC_018670
VERSION NC_015843.2
DBLINK BioProject: [PRJNA485481](#)
KEYWORDS RefSeq.
SOURCE Tembusu virus (THUV)
ORGANISM [Tembusu virus](#)
Viruses; Riboviria; Orthornavirae; Kitrinoviricota; Flusuviricetes; Amarilovirales; Flaviviridae; Flavivirus.
REFERENCE 1 (bases 1 to 10990)
AUTHORS Han,K., Huang,X., Li,Y., Zhao,D., Liu,Y., Zhou,X., You,Y. and Xie,X.
TITLE Complete genome sequence of goose tembusu virus, isolated from jiangnan white geese in jiangsu, china
JOURNAL Genome Announc 1 (2), E0023612 (2013)
PUBMED [23516233](#)
REMARK Publication Status: Online-Only
REFERENCE 2 (bases 1 to 10990)
AUTHORS Huang,X., Han,K., Zhao,D., Liu,Y., Zhang,J., Niu,H., Zhang,K.,

Analyze this sequence
[Run BLAST](#)
[Pick Primers](#)
[Highlight Sequence Features](#)
[Find in this Sequence](#)

Articles about the flavivirus polyprotein gene
Substantial Attenuation of Virulence of Tembusu Virus Strain PS Is Determined by a [J Virol. 2021]
Identification of a Neutralizing Monoclonal Antibody That Recognizes a Uniq [Viruses. 2020]
A Single Mutation at Position 156 in the Envelope Protein of Tembusu Virus Is Respo [J Virol. 2018]
[See all...](#)

Reference sequence information
[RefSeq protein product](#)

Run BLAST option we have to select

Enter accession number(s), gi(s), or FASTA sequence(s)

Or, upload file No file chosen

☐ Align two or more sequences

Choose Search Set

Database ☒ Standard databases (nr etc.) ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

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

Exclude ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to ☐ Sequences from type material

Entrez Query

Program Selection

Optimize for ☒ Highly similar sequences (megablast) ☐ More dissimilar sequences (discontiguous megablast) ☐ Somewhat similar sequences (blastn)
Choose a BLAST algorithm

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)
☐ Show results in a new window

[+ Algorithm parameters](#)

BLAST

Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download Select columns Show 100								
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"/> Tembusu virus strain JS804, complete genome	Tembusu virus	20295	20295	100%	0.0	100.00%	10990	JF895923.2
<input checked="" type="checkbox"/> Tembusu virus isolate SD2010, complete genome	Tembusu virus	20098	20098	100%	0.0	99.67%	10990	MN649260.1
<input checked="" type="checkbox"/> Tembusu virus strain JS/2010, complete genome	Tembusu virus	20064	20064	99%	0.0	99.64%	10990	JX273153.1
<input checked="" type="checkbox"/> Duck egg-drop syndrome virus strain byd1, complete genome	Duck egg-drop syndrome virus	20048	20048	99%	0.0	99.61%	10990	JQ920420.1
<input checked="" type="checkbox"/> Tembusu virus isolate Tembusu virus strain, complete genome	Tembusu virus	20026	20026	99%	0.0	99.57%	10989	KF192951.1
<input checked="" type="checkbox"/> Duck Tembusu virus isolate df-2, complete genome	Duck Tembusu virus	20020	20020	99%	0.0	99.56%	10990	KJ489355.1
<input checked="" type="checkbox"/> Duck egg-drop syndrome virus strain JXSP, complete genome	Duck egg-drop syndrome virus	20020	20020	99%	0.0	99.56%	10990	JQ920423.1
<input checked="" type="checkbox"/> Tembusu virus isolate HB2010, complete genome	Tembusu virus	20018	20018	100%	0.0	99.55%	10990	MN649262.1
<input checked="" type="checkbox"/> Tembusu virus isolate YY5, complete genome	Tembusu virus	20015	20015	99%	0.0	99.55%	10990	JF270480.1
<input checked="" type="checkbox"/> Tembusu virus isolate SDMS, complete genome	Tembusu virus	20009	20009	99%	0.0	99.54%	10990	KC333867.1
<input checked="" type="checkbox"/> Tembusu virus isolate ZJ-6, complete genome	Tembusu virus	20009	20009	99%	0.0	99.54%	10990	JF459991.1
<input checked="" type="checkbox"/> Tembusu virus strain AH-F10 from China, complete genome	Tembusu virus	20004	20004	99%	0.0	99.54%	10990	KM102539.1
<input checked="" type="checkbox"/> Duck egg-drop syndrome virus strain pigeon, complete genome	Duck egg-drop syndrome virus	20004	20004	99%	0.0	99.54%	10990	JQ920425.1
<input checked="" type="checkbox"/> Tembusu virus genomic RNA, complete genome, strain: TMUV-YY10du	Tembusu virus	19998	19998	99%	0.0	99.53%	10990	AB917088.1
<input checked="" type="checkbox"/> Duck Tembusu virus strain BZ_2010, complete genome	Duck Tembusu virus	19998	19998	99%	0.0	99.53%	10990	KC990540.1
<input checked="" type="checkbox"/> Duck egg-drop syndrome virus strain duan, complete genome	Duck egg-drop syndrome virus	19998	19998	99%	0.0	99.53%	10990	JQ920421.1
<input checked="" type="checkbox"/> Duck Tembusu virus strain GDLH01, complete genome	Duck Tembusu virus	19989	19989	99%	0.0	99.51%	10990	KT824876.1
<input checked="" type="checkbox"/> Tembusu virus isolate zYY150902 polyprotein gene, complete cds	Tembusu virus	19981	19981	99%	0.0	99.50%	10990	MF522175.1

Here the result will be display

Download		GenBank	Graphics	Next	Previous	Descriptions
Tembusu virus isolate SD2010, complete genome						
Sequence ID: MN649260.1 Length: 10990 Number of Matches: 1						
Range 1: 1 to 10990						
Score	Expect	Identities	Gaps	Strand		
20098 bits(10883)	0.0	10954/10990(99%)	0/10990(0%)	Plus/Plus		
Query 1	AGAAGTTCGCCTGTGTGAACCTATTCCAAACAGCTTTTGGAGTAGTGCCTGTGAACGTAA	60				
Sbjct 1	AGAAGTTCATCTGTGTGAACCTATTCCAAACAGCTTTTGGAGTAGTGCCTGTGAACGTAA	60				
Query 61	ACACAGTTTGAACGTTTTTTGGATAGAGACAACATATGCTAAACAAAAACGGAAGACCC	120				
Sbjct 61	ACACAGTTTGAACGTTTTTTGGATAGAGACAACATATGCTAAACAAAAACGGAAGACCC	120				
Query 121	CGGCTCAGGCCGGGTTGTCAATATGCTAAAGCGCGGAACGTCCTCCGCGAAATCCGCTAGC	180				
Sbjct 121	CGGCTCAGGCCGGGTTGTCAATATGCTAAAGCGCGGAACGTCCTCCGCGAAATCCGCTAGC	180				
Query 181	GCGGATAAAGAGGACGATTGATGGGGCTCTGAGAGGAGCAGGACCCATAAGGTTTGTGCT	240				
Sbjct 181	GCGGATAAAGAGGACGATTGATGGGGCTCTGAGAGGAGCAGGACCCATAAGGTTTGTGCT	240				
Query 241	GGCTCTACTGACTTTCTTCAAGTTTACAGCCCTGAGGCCCAACATTGGAATGCTGAAGAG	300				
Sbjct 241	GGCTCTACTGACTTTCTTCAAGTTTACAGCCCTGAGGCCCAACATTGGAATGCTGAAGAG	300				
Query 301	ATGGAAGCTGGTTGGAGTTAATGAGGCGACCAACATCTGAAAGCTTCAAGCGTGACAT	360				
Sbjct 301	ATGGAAGCTGGTTGGAGTTAATGAGGCGACCTAAACATCTGAAAGCTTCAAGCGTGACAT	360				
Query 361	TGGACAGATGCTCGACGGACTGAATAAGCGGAAGGCGAAACGTC	420				
Sbjct 361	TGGACAGATGCTCGACGGACTGAATAAGCGGAAGGCGAAACGTC	420				
Query 421	TTGGATCATTATGTTACTCCCGATAGTTGCTGGGCTGAAGCTTGGAACATATAATGGTAG	480				
Sbjct 421	TTGGATCATTATGTTACTCCCGATAGTTGCTGGGCTGAAGCTTGGAACATATAATGGTAG	480				
Query 481	AGTTTTGGCCACTTTAAATAAGACTGATGTATCAGACTTGTAGTCATTCCAATAACGGC	540				
Sbjct 481	AGTTTTGGCCACTTTAAATAAGACCGATGTATCAGACTTGTAGTCATTCCAATAACGGC	540				
Query 541	TGGCAGCAATGGATGCTGCTACGTGCTAGATGTGGGACTAATGTGTCAAGATGACAT	600				