

Name: Ayush S. Tiwari

Class- MSc CS – I

Roll No.- 511

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 specific 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 NCBI (National Center for Biotechnology Information) website. At the top, the NIH logo and 'National Library of Medicine' are visible. A search bar is set to 'Nucleotide' with the query 'flavivirus'. Below the search bar, a 'Welcome to NCBI' section provides links to 'About the NCBI', 'Mission', 'Organization', and 'NCBI News & Blog'. A central area features six icons representing different functions: Submit, Download, Learn, Develop, Analyze, and Research. On the right, 'Popular Resources' include PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. Below this, 'NCBI News & Blog' highlights recent updates like 'NEW! Streamlining ClinVar Submission of Assertion Criteria' and 'Re-releasing the BLAST Nucleotide Database (nt)'. The main search results section lists several entries for 'flavivirus' sequences, including 'Mosquito flavivirus NS5 gene for polyprotein, partial cds, strain YDFV/Sept2013', 'Mosquito flavivirus gene for polyprotein, complete cds, strain YDFV/Oct2013', 'Hangzhou flavivirus 3 isolate YWLFY92, complete genome', 'Tembusu virus flavivirus polyprotein (flavivirus polyprotein gene) gene, complete cds', 'Culex flavivirus gene for polyprotein, partial cds, clone: ocpex_flavi_ns5_24', and 'Cacipacore virus flavivirus polyprotein (flavivirus polyprotein gene) and truncated polyprotein (flavivirus polyprotein gene) genes, complete cds'. Each entry provides details such as accession number, length, and links to various databases like GenBank, PubMed, and Protein.

As 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 [Help](#)

Advanced

GenBank

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

NCBI Reference Sequence: NC_015843.2

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

[Go to:](#)

LOCUS NC_015843 10959 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: PRJNA465481

KEYWORDS RefSeq.

SOURCE Tembusu virus (TMV)

ORGANISM Tembusu virus

Virus; Riboviria; Orthorvirae; Ribooviricetes; Flaviviricetes; Amarillovirales; Flavivirinae; Flavivirus.

REFERENCE 1 (bases 1 to 10959)

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 goose in Jiangsu, China

JOURNAL Genome Announcements 1 (2), e0022022 (2012)

PUBMED 23246232

PMID 23246232

REFERENCE 2 (bases 1 to 10959)

AUTHORS Huang,X., Han,K., Zhao,D., Liu,Y., Zhang,J., Liu,H., Zhang,K.,

Articles about the flavivirus polyprotein gene 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 Unique [Viruses. 2020]
A Single Mutation at Position 156 in the Envelope Protein of Tembusu Virus Is Responsible [J Virol. 2016]

See all

Reference sequence information

Run BLAST option we have to select

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

From To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ☒ Standard databases (in etc.) ☐ tRNA/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 (X-MXP) ☐ 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									
<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"/>	Tembusu virus strain JS64 .complete genome	Tembusu virus	20295	20295	100%	0.0	100.00%	10990	JF895913.1
<input checked="" type="checkbox"/>	Tembusu virus isolate SD2010 .complete genome	Tembusu virus	20090	20090	100%	0.0	99.67%	10990	MN619260.1
<input checked="" type="checkbox"/>	Tembusu virus strain JS2010 .complete genome	Tembusu virus	20064	20064	99%	0.0	99.64%	10990	JX273153.1
<input checked="" type="checkbox"/>	Duck egg-drop syndrome virus strain Jyoti .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	99.67%	10989	KF162961.1
<input checked="" type="checkbox"/>	Duck Tembusu virus isolate HZ .complete genome	Duck Tembusu virus	20020	20020	99%	0.0	99.56%	10990	KJ408365.1
<input checked="" type="checkbox"/>	Duck egg-drop syndrome virus strain JXSF .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	MN619262.1
<input checked="" type="checkbox"/>	Tembusu virus isolate YNS .complete genome	Tembusu virus	20015	20015	99%	0.0	99.55%	10990	JF279466.1
<input checked="" type="checkbox"/>	Tembusu virus isolate SDMS .complete genome	Tembusu virus	20009	20009	99%	0.0	99.54%	10990	KC313067.1
<input checked="" type="checkbox"/>	Tembusu virus isolate ZH6 .complete genome	Tembusu virus	20009	20009	99%	0.0	99.54%	10990	JF499991.1
<input checked="" type="checkbox"/>	Tembusu virus strain AHE10 from China .complete genome	Tembusu virus	20004	20004	99%	0.0	99.54%	10990	KM182539.1
<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-YY1Du	Tembusu virus	19990	19990	99%	0.0	99.53%	10990	AB917038.1
<input checked="" type="checkbox"/>	Duck Tembusu virus strain BZ_2010 .complete genome	Duck Tembusu virus	19990	19990	99%	0.0	99.53%	10990	KC906540.1
<input checked="" type="checkbox"/>	Duck egg-drop syndrome virus strain duan .complete genome	Duck egg-drop syndrome virus	19990	19990	99%	0.0	99.53%	10990	JQ920421.1
<input checked="" type="checkbox"/>	Duck Tembusu virus strain GDJH01 .complete genome	Duck Tembusu virus	19989	19989	99%	0.0	99.51%	10990	KT624076.1
<input checked="" type="checkbox"/>	Tembusu virus isolate pY150002 polyprotein gene .complete cds	Tembusu virus	10901	10901	0%	0.0	0.00%	10990	MF622175.1

Here the result will be display

Download	GenBank Graphics	Next Previous Descriptions
Tembusu virus isolate SD2010, complete genome		
Sequence ID: MN619260.1 Length: 10990 Number of Matches: 1		
Range 1: 1 to 10990 GenBank Graphics		
Score	Expect	Identities
20098 bits(10883)	0.0	10954/10990(99%)
		Gaps
		0/10990(0%)
		Strand
		Mus/Plus
Query 1	AGAAAGTTCCCTGTGTGAACCTATTCCAAACAGCTTTTGGAGTAGTGCCTGGAACGTAA	60
Subject 1	AGAAAGTTCACTGTGTGAACCTATTCCAAACAGCTTTTGGAGTAGTGCCTGGAACGTAA	60
Query 61	ACACAGTTTGAACGTTTTTGGATAGAGCAACATATGCTTAACAAAAAACCAGGAAGACC	120
Subject 61	ACACAGTTTGAACGTTTTTGGATAGAGCAACATATGCTTAACAAAAAACCAGGAAGACC	120
Query 121	CGGCTCAGCCGGGTTGTCATATGCTAAAGCGCGGAACGTCCCGGGAATCCGCTAGC	180
Subject 121	CGGCTCAGCCGGGTTGTCATATGCTAAAGCGCGGAACGTCCCGGGAATCCGCTAGC	180
Query 181	GCGGATAAAGAGGACGATTCATGGGTCCTGAGAGGAGCAGGACCCATAAGCTTTGTCT	240
Subject 181	GCGGATAAAGAGGACGATTCATGGGTCCTGAGAGGAGCAGGACCCATAAGCTTTGTCT	240
Query 241	GACCTACTBACTTTCTCAAGTTTACAGCCCTGAGGCCAACCATTTGAATGCTAAGAG	300
Subject 241	GACCTACTBACTTTCTCAAGTTTACAGCCCTGAGGCCAACCATTTGAATGCTAAGAG	300
Query 301	ATGGAAGCTGGTTGGAGTTAATGAGGCGACCAACATCTGAAGAGCTTCAAGCGTGACAT	360
Subject 301	ATGGAAGCTGGTTGGAGTTAATGAGGCGACTAATCATCTGAAGAGCTTCAAGCGTGACAT	360
Query 361	TGGACAGATGCTCGACGGACTGAATAAGCGGAAGGCGAAGCTCGGGGGGAGTTGCTC	420
Subject 361	TGGACAGATGCTCGACGGACTGAATAAGCGGAAGGCGAAGCTCGGGGGGAGTTGCTC	420
Query 421	TTGGATCATTATGTTACTCCCGATAGTTGCTGGCTGAAGCTTGGAACTATAATGGTAG	480
Subject 421	TTGGATCATTATGTTACTCCCGATAGTTGCTGGCTGAAGCTTGGAACTATAATGGTAG	480
Query 481	AGTTTGGGCTTTTAAATAGAGCGATGATCAGACTTGGTAGTCATCCAGCAAGGCG	540
Subject 481	AGTTTGGGCTTTTAAATAGAGCGATGATCAGACTTGGTAGTCATCCAGCAAGGCG	540
Query 541	TGGACCAATGGATGCTCCTACGTGCTCTAGATGTGGGACTAATGCTCAGGATGACAT	600