



# ***Introduction to Pathogen Genomic Databases***

Zhu Feng  
Duke-NUS Graduate Medical School


# ***What are pathogen genomic database***


- **Genomic Sequences:** Storage of complete or partial DNA/RNA sequences of pathogens.
- **Annotations:** Detailed information about genes, proteins, and other functional elements of the genome.
- **Metadata:** Contextual data including information on the source of the sample, date of collection, and clinical details.
- **Comparative Analysis:** Tools and resources for comparing genetic sequences to identify similarities and differences among strains.
- **Epidemiological Tracking:** Resources to trace the spread and outbreaks of infectious diseases.
- **Support for Diagnostics and Treatments:** Information that aids in the development of diagnostic tests, vaccines, and therapeutic strategies.
- **Data Sharing and Collaboration:** Platforms for researchers and public health officials to share data and collaborate on a global scale.

# ***What are pathogen genomic database***


- **GISAID**
- **NCBI**
- **BV-BRC**

# NCBI Virus Database

 **National Library of Medicine**  
National Center for Biotechnology Information


 **NCBI Virus**  
Sequences for discovery


About Us ▾Find Data ▾Help ▾How to Participate ▾Submit Sequences ▾

 **Quick Access to SARS-CoV-2 Data!**


- Novel Severe acute respiratory syndrome coronavirus 2 [RefSeq genomes](#), [nucleotide](#), and [protein](#) sequences.
- View our new [SARS-CoV-2 interactive dashboard](#).
- How to [submit SARS-CoV-2 sequences](#).
- Visit our new SARS-CoV-2 [Variants Overview](#) **New!**


NCBI Virus is a community portal for viral sequence data from RefSeq, GenBank and other NCBI repositories. To find, retrieve and analyze data, please select an option below.

 **Search by sequence**  
Use the NCBI BLAST™ tool to find similar viral nucleotide and protein sequences.

 **Search by virus**  
Use virus name or taxid to find viral nucleotide and protein sequences.

- By sequence
- By virus

 **Search by sequence**  
Use the NCBI BLAST™ tool to find similar viral nucleotide and protein sequences.

 **Search by virus**  
Use virus name or taxid to find viral nucleotide and protein sequences.

Nucleotide SearchProtein Search

Please enter sequence in plain text or FASTA format or NCBI sequence accession.

Start search

Search up-to-date Betacoronavirus DB

Try a Human orthopneumovirus example: [NC\\_001781](#)

 **Search by sequence**  
Use the NCBI BLAST™ tool to find similar viral nucleotide and protein sequences.

 **Search by virus**  
Use virus name or taxid to find viral nucleotide and protein sequences.

**Search by virus name or taxonomy**  
Begin typing a virus name, viral taxonomy group, or taxid to select from a list of suggestions.

virus name or taxid

Popular searches

All viruses

Human viruses

Bacteriophages

New sequences (past one month)

Up-to-date SARS-CoV-2



# Multiple filters can be activated

Refine Results Reset

Virus

West Nile virus, taxid:11082

Accession

Sequence Length

Ambiguous Characters

Sequence Type

RefSeq Genome Completeness

Nucleotide Completeness

complete

Isolate

Genotype New!

Proteins

Provirus

Geographic Region

Host

Homo sapiens (human), taxid:9606

Applied Filters: Virus (1) Nucleotide Completeness (1) Host (1)

	Nucleotide (146)	Protein (142)	RefSeq Genome (0)								Select Columns
	<input type="checkbox"/>	Accession	Organism Name	Submitters	Organization	Release Date	Isolate	Species	Length	Nuc Completeness	Ge
Expand Table	<input type="checkbox"/>	<a href="#">PP212878</a>	West Nile virus	Nagy,A.	National Center for Public...	2024-02-03	3717_2023_WNV_HU	Orthoflavivirus nilense	11009	complete	H
	<input type="checkbox"/>	<a href="#">PP212879</a>	West Nile virus	Nagy,A.	National Center for Public...	2024-02-03	3919_2023_WNV_HU	Orthoflavivirus nilense	11000	complete	H
	<input type="checkbox"/>	<a href="#">PP212880</a>	West Nile virus	Nagy,A.	National Center for Public...	2024-02-03	4429_2023_WNV_HU	Orthoflavivirus nilense	11003	complete	H
	<input type="checkbox"/>	<a href="#">PP212881</a>	West Nile virus	Nagy,A.	National Center for Public...	2024-02-03	4106_2023_WNV_HU	Orthoflavivirus nilense	10994	complete	H
	<input type="checkbox"/>	<a href="#">PP212882</a>	West Nile virus	Nagy,A.	National Center for Public...	2024-02-03	5226_2023_WNV_HU	Orthoflavivirus nilense	10997	complete	H
	<input type="checkbox"/>	<a href="#">PP212883</a>	West Nile virus	Nagy,A.	National Center for Public...	2024-02-03	5903_2023_WNV_HU	Orthoflavivirus nilense	10991	complete	H
	<input type="checkbox"/>	<a href="#">OR757508</a>	West Nile virus	Antonov,A.S., et al.	Volgograd Plague Control ...	2023-11-19	KBR 277/23	Orthoflavivirus nilense	11038	complete	R
	<input type="checkbox"/>	<a href="#">OR757510</a>	West Nile virus	Antonov,A.S., et al.	Volgograd Plague Control ...	2023-11-19	Chuvashiya 581/23	Orthoflavivirus nilense	11038	complete	R
	<input type="checkbox"/>	<a href="#">OR757511</a>	West Nile virus	Antonov,A.S., et al.	Volgograd Plague Control ...	2023-11-19	Krasnodar 858/23	Orthoflavivirus nilense	11038	complete	R
	<input type="checkbox"/>	<a href="#">OP617270</a>	West Nile virus	Bayrakdar,F.	Public Health Institution ...	2023-05-02	WNV/Turkey/HS GM136s/2...	Orthoflavivirus nilense	10978	complete	T
	<input type="checkbox"/>	<a href="#">OP345107</a>	West Nile virus	Antonov,A.S., et al.	Volgograd Plague Control ...	2023-04-10	Volgograd 810/19	Orthoflavivirus nilense	10966	complete	R
	<input type="checkbox"/>	<a href="#">OP179287</a>	West Nile virus	Nagy,A., et al.	National Public Health Ce...	2023-03-29	HUN_WNV2324_2021	Orthoflavivirus nilense	11009	complete	H
	<input type="checkbox"/>	<a href="#">OP179288</a>	West Nile virus	Nagy,A., et al.	National Public Health Ce...	2023-03-29	HUN_WNV2947_2021	Orthoflavivirus nilense	11021	complete	H
	<input type="checkbox"/>	<a href="#">OQ053515</a>	West Nile virus	Tsioka,K., et al.	Aristotle University of The...	2023-02-01		Orthoflavivirus nilense	10992	complete	G
	<input type="checkbox"/>	<a href="#">OQ053516</a>	West Nile virus	Tsioka,K., et al.	Aristotle University of The...	2023-02-01		Orthoflavivirus nilense	10989	complete	G

# Simple Visual Dashboard

## Explore Virus Data

West Nile virus

Filters (0)

View the Results Table and download

Quick Links

All viruses

Our Data Model

Our publications - cite us!

Advanced Filters for  
GenBank Sequences

Visual Filters for  
GenBank Sequences

### Sequence Type

2  
RefSeq Nucleotides

7,725  
All Proteins

23,290  
All Nucleotides

34  
RefSeq Proteins

1,993  
Complete Nucleotides

### Geographic and Time Distribution

Choose locations to select sequence records with that collection location. Use the sliders or click date columns to select records by their sample collection date and/or their GenBank release date.

#### Geographic Distribution

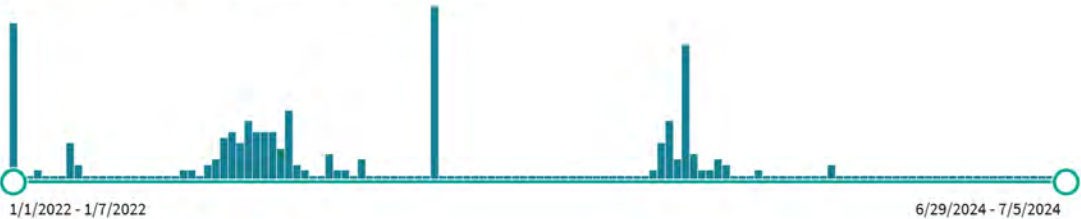
Search for a country



#### Collection Time

Weekly

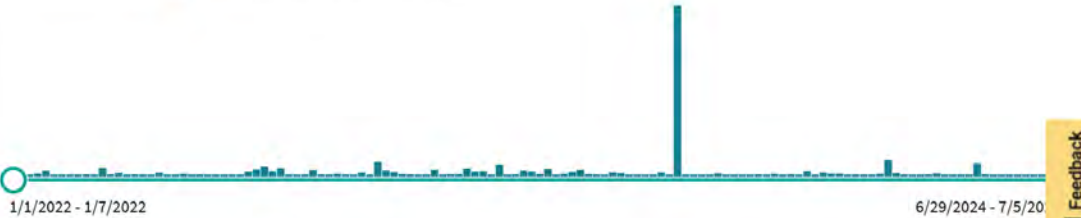
2022 - 2024



#### Release Time

Weekly

2022 - 2024



Feedback

# Data Export

Explore Virus Data Download

Popular Searches: Influenza virus, Dengue virus, Zika virus, Ebola virus

Advanced Filters for GenBank Sequences | Visual Filters for GenBank Sequences

**New! Randomized subsets in Downloads**  
You now have the option of downloading a smaller, randomized subset of all records. Follow the prompts in the Download menu. Our Help page has more information.

Refine Results: Virus, Accession

Applied Filters: Nucleotide, Coding Region, Protein

**Download Results**

**Step 1 of 3: Select Data Type**

**Sequence Data (FASTA format)**

☒ Nucleotide  
☐ Coding Region  
☐ Protein

**Accession List**

☐ Nucleotide  
☐ Protein  
☐ Assembly

**Results Table**

☐ CSV format  
☐ XML format

**Next**

West Nile virus, GenBank: PF117879, West Nile virus, NigdyA, National Center for Public Health, 2024-02-03, 3915\_2023\_WNV\_HI, Orthohavivirus nirse

**Download Results**

**Step 2 of 3: Select Records**

☒ Download Selected Records (146)  
☐ Download All Records (146)  
☐ Download a randomized subset of all records (up to 146) Reset

**Back** **Next**

**Download Results**

**Step 3 of 3: Select FASTA definition line**

☐ Use default: Accession GenBank Title

☒ Build custom: Accession GenBank Title Organism Name Species Isolate

Assembly  
SRA Accession  
Submitters  
Organization  
Release Date  
Genus  
Family  
Molecule type

**Add** **Remove**

Accession  
GenBank Title  
Organism Name  
Species  
Isolate

**Back** **Download**



# Data submission

## Submission Portal

Submit to the world's largest public repository of biological and scientific information

Type a few words about the sequence data you are submitting and select an option to learn more. You can browse submission information below.

### What do you want to submit?

Enter a few words about your sequence data.

virus genome



Suggest tool

#### Suggested tools

##### Genome >

Submit assembled complete or incomplete/draft prokaryotic and eukaryotic genomes. Not for viral, phage, or single locus sequences (for example: 16S rRNA). Submit those to regular [GenBank](#).

##### GenBank >

Submit ribosomal RNA (rRNA), rRNA-ITS, SARS-CoV-2, Influenza, Norovirus, Dengue, metazoan COX1 or eukaryotic nuclear mRNA

##### BankIt >

Submit genomic DNA, organelle, ncRNA, plasmids, other viruses, phages, mRNA and synthetic constructs.

##### SRA >

SRA accepts unassembled reads from high throughput sequencing platforms. Submitted data files should generally be minimally processed and include per-base quality scores.

## GenBank

GenBank is the world's largest nucleotide archive containing sequences from all branches of life. The archive is a foundation for medical and biological discovery.

Started 2024-05-14

Submit assembled [SARS-CoV-2](#), Influenza, Norovirus, Dengue virus, rRNA, rRNA-ITS, metazoan COX1, Eukaryotic nuclear mRNA sequences.

Learn more

Submit

Submit genomic DNA, organelle, ncRNA, plasmids, other viruses, phages, mRNA, synthetic constructs.

Learn more

Submit

Submit assembled eukaryotic and prokaryotic genomes (WGS or Complete).

Learn more

Submit

## Sequence Read Archive (SRA)

SRA is the largest publicly-available repository of high throughput sequencing data. The archive accepts data from all branches of life as well as metagenomic and environmental surveys.

Submit unassembled, high throughput sequencing reads

[SARS-CoV-2 submission instructions](#)

Learn more

Submit

## [How to submit data to GenBank \(nih.gov\)](#)



- [BankIt](#), a WWW-based submission tool
- [Submission Portal](#), a unified system for multiple submission types.



## BankIt

### Submit new sequences to GenBank

#### What type of sequence data do you have?

- ☐ SARS-CoV-2 
- ☐ Ribosomal RNA (rRNA) or rRNA-ITS
- ☐ Metazoan (multicellular animal) COX1
- ☐ Eukaryotic nuclear mRNA  **Updated (Jul 2024)**
- ☐ Influenza virus
- ☐ Norovirus
- ☐ Dengue virus
- ☐ Eukaryotic and Prokaryotic Genomes (WGS or Complete)
- ☐ Transcriptome Shotgun Assembly (TSA)
- ☐ Unassembled sequence reads (SRA)
- ☐ Sequence data not listed above (through BankIt): genomic DNA, organelle, ncRNA, plasmids, other viruses, phages, other mRNA, synthetic constructs

Start

Need help? Contact GenBank user services at [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov).

#### Complete or view existing BankIt submissions

#### Need to update a GenBank record?

Request updates to accessioned records per the [GenBank update page](#).

# BV-BRC Database

3.36.16.5

ORGANISMS

SEARCHES

TOOLS & SERVICES

WORKSPACES

HELP

ABOUT

Bacterial Pathogens

Acinetobacter

Bacillus

Bartonella

Borrelia

Brucella

Burkholderia

Campylobacter

Chlamydia

Clostridium

Coxiella

Ehrlichia

Escherichia

All Bacteria

All Viruses

Francisella

Helicobacter

Listeria

Mycobacterium

Pseudomonas

Rickettsia

Salmonella

Shigella

Staphylococcus

Streptococcus

Vibrio

Yersinia

All Archaea

Eukaryotic Hosts

Viral Families

Adenoviridae

Asfarviridae

Bunyavirales

Caliciviridae

Coronaviridae

Filoviridae

Flaviviridae

Hepadnaviridae

Hepeviridae

Herpesvirales

Featured Viruses

Bacteriophages

Dengue virus

Ebolavirus

Enterovirus

Hepacivirus

Orthomyxoviridae

Paramyxoviridae

Parvoviridae

Picornaviridae

Pneumoviridae

Polyomaviridae

Poxviridae

Reovirales

Rhabdoviridae

Togaviridae

Influenza A virus

Mammarenavirus lassaense

Monkeypox virus

SARS-CoV-2

Zika virus

RESOURCE CENTER

Q All terms

VIRUSES

EUKARYOTIC HOSTS

OUTBREAKS

Influenza H5N1 2024

SARS-CoV-2

NEWS & ANNOUNCEMENTS

BV-BRC Viral Team at ASV

The BV-BRC team will be at ASV 2024. Stop by and see us! ... [read more](#)

BV-BRC

ASV 2024

We are looking forward to seeing you!

COMMUNITY NEWS


BV-BRC

271 followers

1d

We have updated our map to include the latest human case of H5N1 detected in Colorado.

2024 H5N1 Cases in Cattle and Humans



SEARCH

Taxa

Genomes

Proteins

Specialty Genes

[All Searches...](#)

ANALYZE

Assembly

Annotation

BLAST

MSA

[All Tools & Services...](#)

MANAGE DATA

Access Private Workspace

Save Search Results

Upload Data

Access Analysis Jobs

Share and Publish

BATCH ACCESS

Command-Line Interface

Data API

FTP


QUICK START

REFERENCE GUIDES

TUTORIALS


Information for PATRIC Users

The new BV-BRC system is built on the PATRIC system. Viral data and analysis tools and services from IRD/ViPR have now been integrated into the



Information for IRD/ViPR Users

The new BV-BRC system is built on the PATRIC Bacterial BRC system and incorporates data and tools from IRD and ViPR. The following materials are designed to aid IRD and ViPR users in



- Bacterial
- Virus

SEARCHES

TOOLS & S

Advanced Search

Taxa

Genomes

Strains

Features

Proteins

Specialty Genes

Domains and Motifs

Epitopes

Protein Structures

Pathways

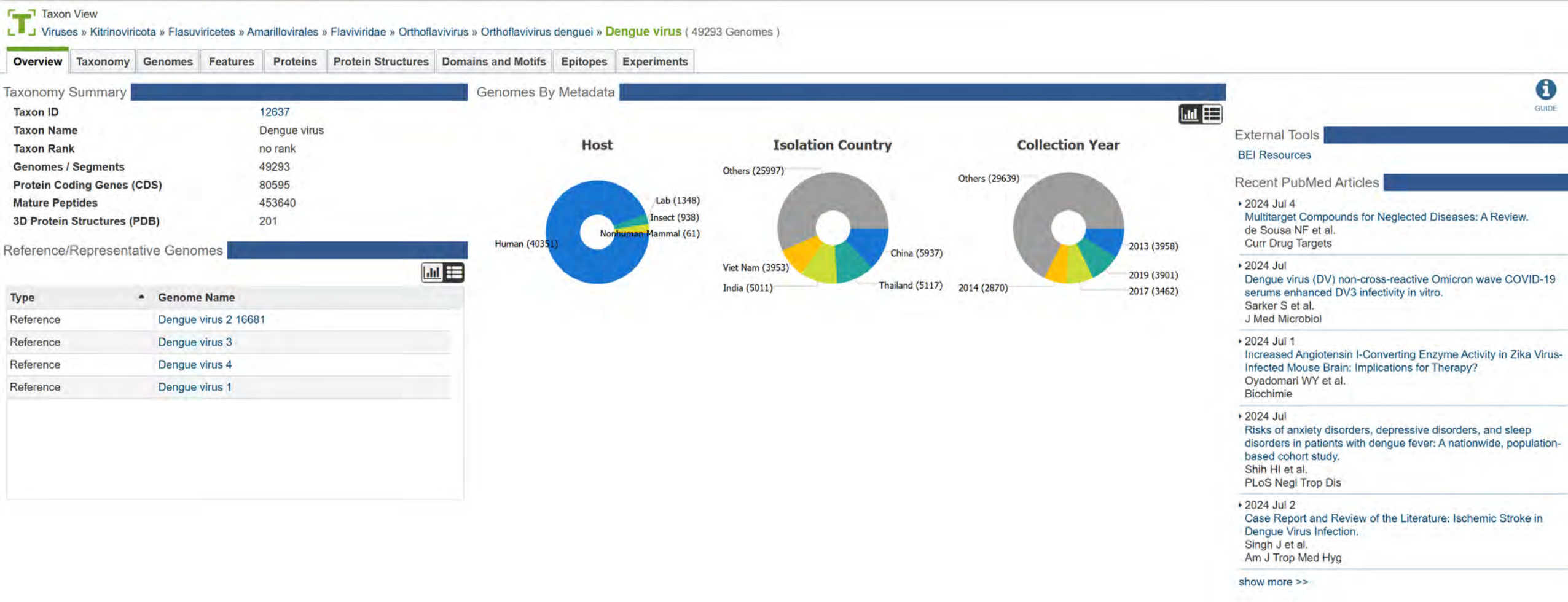
Subsystems

Surveillance

Serology

ASIA PATHOGEN  
GENOMICS Initiative

# Simple Visual Dashboard





# Multiple filters and data export

Taxon View  
Viruses » Kitrinoviricota » Flaviviricetes » Amarillovirales » Flaviviridae » Orthoflavivirus » Orthoflavivirus denguei » **Dengue virus** ( 49293 Genomes )

Overview Taxonomy **Genomes** Features Proteins Protein Structures Domains and Motifs Epitopes Experiments

Download KEYWORDS ADV Search Complete 2019 GENOME\_STATUS x COLLECTION\_YEAR x

HIDE APPLY

130 items selected.

Public Genome Status Subtype Collection Year Isolation Country Geographic Group Host Common Name Host Group

true (130) Complete (130) Partial (3770) WGS (1) 1 (52) 2 (50) 3 (8) 4 (15) 2019 (130) 1944 (5) 1953 (1) 1956 (4) 1960 (2) Australia (1) Bangladesh (4) Brazil (25) Cambodia (21) China (18) Asia (98) North America (1) Oceania (1) South America (25) Human (123) Mosquito (2) Insect (2)

	Genome Name	Strain	GenBank Accessions	Size	CDS	Collection Year	Isolation Country	Host Common Name
<input checked="" type="checkbox"/>	dengue virus type 3 D191267	D191267	OR053670	10707	1	2019		Human
<input checked="" type="checkbox"/>	dengue virus type 3 Dengue virus type 3 Homo sapiens/Bangl	Dengue virus type 3 Homo sapiens/Bangladesh-UF-1/2019	PP711274	10707	1	2019	Bangladesh	Human
<input checked="" type="checkbox"/>	Dengue virus type 3 DV3syn	DV3syn	OQ948474	10707	1	2019	China	Human
<input checked="" type="checkbox"/>	dengue virus type 1 DENVN19_080	DENVN19_080	OQ832610	10735	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 DENVN19_078	DENVN19_078	OQ832621	10723	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 DENVN19_089	DENVN19_089	OQ832622	10723	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 DENVN19_143	DENVN19_143	OQ832623	10723	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 DENVN19_010	DENVN19_010	OQ832617	10723	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 DENVN19_013	DENVN19_013	OQ832619	10723	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	dengue virus type 1 DENVN19_137	DENVN19_137	OQ832612	10735	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 DENVN19_011	DENVN19_011	OQ832618	10723	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	dengue virus type 1 DENVN19_075	DENVN19_075	OQ832609	10735	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 DENVN19_015	DENVN19_015	OQ832620	10723	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 DENVN19_004	DENVN19_004	OQ832616	10723	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	dengue virus type 1 100-0587	100-0587	OQ683856	10733	1	2019	Cambodia	Human
<input checked="" type="checkbox"/>	Dengue virus type 2 C2_2019_DENV-2	C2_2019_DENV-2	OQ101603	10725	1	2019	Brazil	Human
<input checked="" type="checkbox"/>	dengue virus type 1 128/SH/2019	128/SH/2019	OM281595	10735	1	2019	Viet Nam	Human
<input checked="" type="checkbox"/>	dengue virus type 1 132/SH/2019	132/SH/2019	OM281596	10735	1	2019	Maldives	Human
<input checked="" type="checkbox"/>	dengue virus type 1 137/SH/2019	137/SH/2019	OM281597	10736	1	2019	Philippines	Human
<input checked="" type="checkbox"/>	dengue virus type 1 144/SH/2019	144/SH/2019	OM281598	10735	1	2019	Thailand	Human
<input checked="" type="checkbox"/>	dengue virus type 1 030/SH/2019	030/SH/2019	OM281579	10735	1	2019	Cambodia	Human

1 - 130 of 130 results

Download selected Genomes (130) as...  
Text DNA FASTA  
CSV

GENOMES GROUP

Uploads Jobs 0 0 3