

## Sourcing of Annotation Files

The following method was used to download the files for the reference genomes. This method was chosen as it contained direct download links to the gff files which can be used to load annotation tracks into the browsers using just a wget command rather than having to save the file locally.

Steps to access the annotation links

1. Go to [NCBI Genome](https://www.ncbi.nlm.nih.gov/genome/) and type in desired virus strain, here Dengue Virus Type 2 is used as an example

The screenshot shows the NCBI Genome search interface. The search bar contains 'dengue virus type 2'. Below the search bar, there are filters and a table of results. The table has columns for Assembly, GenBank, RefSeq, Scientific name, Modifier, Annotation, and Action. The first result is 'ViralProj20183' with GenBank ID GCA\_000871845.1 and RefSeq ID GCF\_000871845.1. The scientific name is 'dengue virus type 2' and the modifier is '16681 (strain)'. The annotation is 'NCBI RefSeq'.

2. Click on the available result and it should take you to this page. If you have multiple results, use the entry that matches your needs.

The screenshot shows the NCBI Genome assembly page for 'ViralProj20183'. The page has a header with the NCBI logo and a search bar. Below the header, there is a table with columns for Download, datasets, URL, FTP, and Actions. The table contains information about the assembly, including the NCBI RefSeq assembly ID GCF\_000871845.1, the Submitted GenBank assembly ID GCA\_000871845.1, the Taxon 'dengue virus type 2', the Strain '16681', the Relation to type material 'ICTV species exemplar', the Submitter 'Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Services', and the Date 'Nov 1, 2007'. To the right of the table, there is a section titled 'Additional genomes' with a link to 'Browse all Orthoflavivirus dengue genomes (10)'. Below this, there is a section titled 'Publications' showing two publications: 'Virology 1997 Construction of infectious cDNA clones for dengue 2 virus: strain 16681 and its attenuated vaccine derivative, strain PDK-53 RM Kinney, et al.' and 'Viruses 2023 Subgenomic Flaviviral RNAs of Dengue Viruses Y Liu, et al.'. There is a 'View in PubMed' button for the second publication. At the bottom, there is a section titled 'Assembly statistics'.

3. Now click on FTP and it will take you to this page

### **Index of /genomes/all/GCF/000/871/845/GCF\_000871845.1\_ViralProj20183**

Name	Last modified	Size
<a href="#">Parent Directory</a>		-
<a href="#">GCF_000871845.1_ViralProj20183_assembly_report.txt</a>	2024-09-03 06:32	1.1K
<a href="#">GCF_000871845.1_ViralProj20183_assembly_stats.txt</a>	2024-09-03 06:32	5.0K
<a href="#">GCF_000871845.1_ViralProj20183_cds_from_genomic.fna.gz</a>	2019-12-18 09:52	3.5K
<a href="#">GCF_000871845.1_ViralProj20183_feature_count.txt</a>	2023-12-13 19:15	373
<a href="#">GCF_000871845.1_ViralProj20183_feature_table.txt.gz</a>	2019-12-18 09:52	727
<a href="#">GCF_000871845.1_ViralProj20183_genomic.fna.gz</a>	2015-05-15 01:55	3.5K
<a href="#">GCF_000871845.1_ViralProj20183_genomic.gbff.gz</a>	2023-07-23 01:22	9.7K
<a href="#">GCF_000871845.1_ViralProj20183_genomic.gff.gz</a>	2023-05-04 18:03	1.7K
<a href="#">GCF_000871845.1_ViralProj20183_genomic.gtf.gz</a>	2022-08-16 01:30	567
<a href="#">GCF_000871845.1_ViralProj20183_protein.faa.gz</a>	2023-07-23 01:22	2.1K
<a href="#">GCF_000871845.1_ViralProj20183_protein.gpff.gz</a>	2023-07-23 01:22	4.6K
<a href="#">GCF_000871845.1_ViralProj20183_rna_from_genomic.fna.gz</a>	2019-12-18 09:52	407
<a href="#">GCF_000871845.1_ViralProj20183_translated_cds.faa.gz</a>	2019-12-18 09:52	2.2K
<a href="#">README.txt</a>	2024-08-27 13:56	55K
<a href="#">annotation_hashes.txt</a>	2024-02-25 03:26	410
<a href="#">assembly_status.txt</a>	2024-11-24 08:00	14
<a href="#">md5checksums.txt</a>	2024-09-03 06:32	1.1K
<a href="#">uncompressed_checksums.txt</a>	2024-08-26 16:58	789

[HHS Vulnerability Disclosure](#)

4. From this list we need the genomic fna file, the genomic gff file, so hover over the hyperlinks and copy the link addresses. The links will look something like this
- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/871/845/GCF\\_000871845.1\\_ViralProj20183/GCF\\_000871845.1\\_ViralProj20183\\_genomic.fna.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/871/845/GCF_000871845.1_ViralProj20183/GCF_000871845.1_ViralProj20183_genomic.fna.gz)
  - [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/871/845/GCF\\_000871845.1\\_ViralProj20183/GCF\\_000871845.1\\_ViralProj20183\\_genomic.gff.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/871/845/GCF_000871845.1_ViralProj20183/GCF_000871845.1_ViralProj20183_genomic.gff.gz)

**These are the exact links we used:**

#### Dengue Type 1

- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/862/125/GCF\\_000862125.1\\_ViralProj15306/GCF\\_000862125.1\\_ViralProj15306\\_genomic.fna.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/862/125/GCF_000862125.1_ViralProj15306/GCF_000862125.1_ViralProj15306_genomic.fna.gz)
- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/862/125/GCF\\_000862125.1\\_ViralProj15306/GCF\\_000862125.1\\_ViralProj15306\\_genomic.gff.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/862/125/GCF_000862125.1_ViralProj15306/GCF_000862125.1_ViralProj15306_genomic.gff.gz)

#### Dengue Type 2

- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/871/845/GCF\\_000871845.1\\_ViralProj20183/GCF\\_000871845.1\\_ViralProj20183\\_genomic.fna.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/871/845/GCF_000871845.1_ViralProj20183/GCF_000871845.1_ViralProj20183_genomic.fna.gz)
- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/871/845/GCF\\_000871845.1\\_ViralProj20183/GCF\\_000871845.1\\_ViralProj20183\\_genomic.gff.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/871/845/GCF_000871845.1_ViralProj20183/GCF_000871845.1_ViralProj20183_genomic.gff.gz)

#### Dengue Type 3

- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/031/099/465/GCA\\_031099465.1\\_ASM3109946v1/GCA\\_031099465.1\\_ASM3109946v1\\_genomic.fna.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/031/099/465/GCA_031099465.1_ASM3109946v1/GCA_031099465.1_ASM3109946v1_genomic.fna.gz)
- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/031/099/465/GCA\\_031099465.1\\_ASM3109946v1/GCA\\_031099465.1\\_ASM3109946v1\\_genomic.gff.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/031/099/465/GCA_031099465.1_ASM3109946v1/GCA_031099465.1_ASM3109946v1_genomic.gff.gz)

#### Dengue Type 4

- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/004/786/575/GCF\\_004786575.1\\_ASM478657v1/GCF\\_004786575.1\\_ASM478657v1\\_genomic.fna.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/004/786/575/GCF_004786575.1_ASM478657v1/GCF_004786575.1_ASM478657v1_genomic.fna.gz)
- [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/004/786/575/GCF\\_004786575.1\\_ASM478657v1/GCF\\_004786575.1\\_ASM478657v1\\_genomic.gff.gz](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/004/786/575/GCF_004786575.1_ASM478657v1/GCF_004786575.1_ASM478657v1_genomic.gff.gz)