Download data from VEuPathDB

https://VEuPathDB.org   help@VEuPathDB.org

VEuPathDB
Eukaryotic Pathogen, Vector & Host
Informatics Resources

DOWNLOAD GENOME DATA FILES

VEuPathDB genomes (assemblies and annotation) are easily downloaded. If your studies are based on a previous release, use >Data File Properties >Release to choose your desired version.

- **GFF** - gene structure in general feature format
- **CDSs.fasta** - multifasta
- **Proteins.fasta** - multifasta
- **Transcripts.fasta** - multifasta
- **CodonUsage.txt** - occurrence of each codon
- **Curated_GO.gaf** - Gene ontology terms assigned by curators
- **GO.gaf** - Gene ontology terms assigned by electronic prediction or transferred by orthology
- **GeneAliases.txt** - previous IDs for current genes
- **Genome.fasta** - chromosomes contigs or scaffolds
- **NCBILinkout_Nucleotide.xml**
- **NCBILinkout_Protein.xml**
- **Orf50.gff**
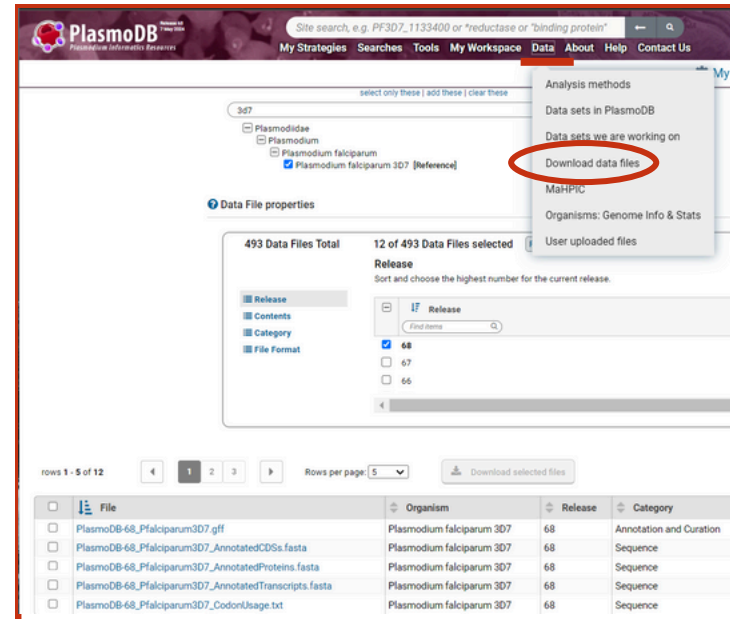
DOWNLOAD DATA FOR MULTIPLE GENES

Data such as expression values, gene metrics, domains, etc. are available for bulk download as custom or preconfigured tables for any list of genes, including a whole genome. This download tool works similarly for lists of other records such as SNPs or metabolic pathways.

1. **Retrieve your list of genes as a search result**
   a. For all genes in a genome use
      >Genes >Taxonomy >Organism
   b. To enter a list of your favorite genes use
      >Genes >Annotation >List of IDs
2. **Open the Download tool** from the search result.
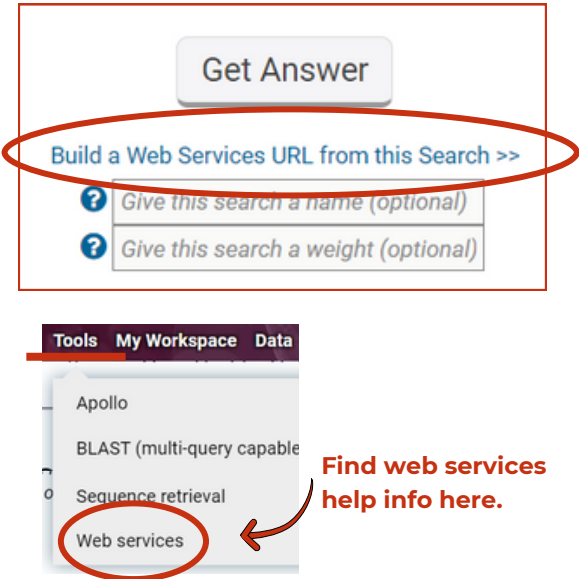3. **Configure the Download tool**
   a. custom table - one row per gene, choose from all data columns
   b. pre-configured tables - single data type data table that can contain multiple rows per gene
   c. BED - coordinates of sequences, configurable
   d. FASTA - sequence retrieval, configurable
   e. GFF3 - gene models
   f. Standard JSON - compatible with web services.
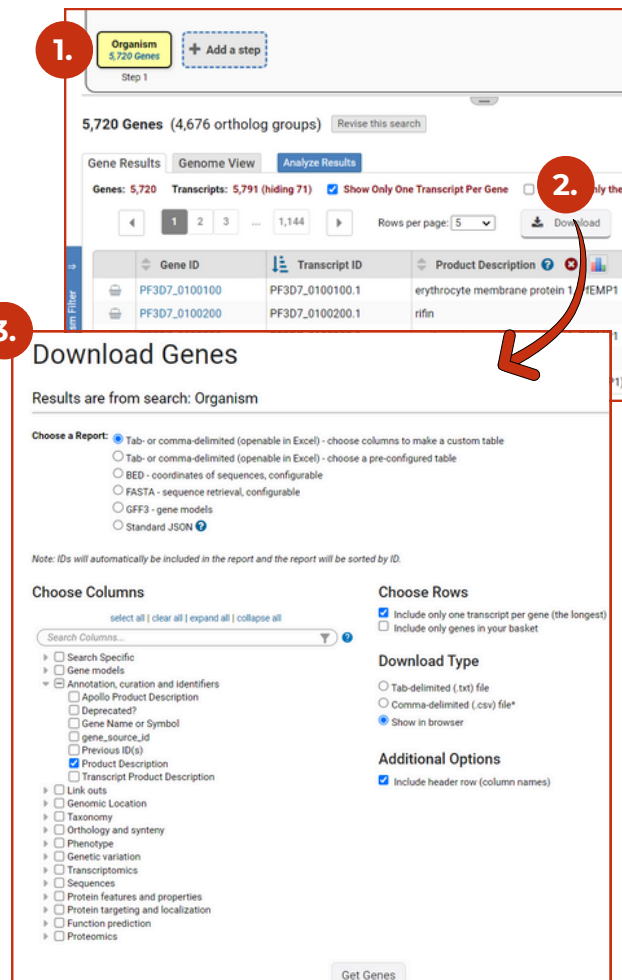
DOWNLOAD DATA VIA WEB SERVICES

Data download using Web Services is a programmatic alternative to the **DOWNLOAD DATA FOR MULTIPLE GENES** process described above. Web Services creates a URL that performs a search and retrieves the data to your browser or to a file.
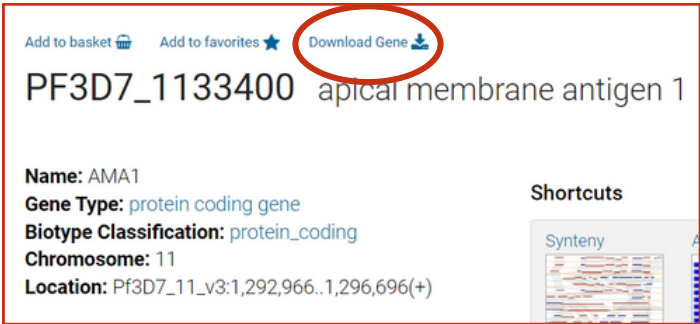
- Each search contains a link to a wizard that helps create the URL. URLs can be used in scripts to programmatically access VEuPathDB data.
- For FASTA and GFF genome files, the getAllGenomeFasta.py (repo: https://github.com/kathryncrouch/misc) is useful for obtaining files in bulk for all the organisms in one of our component sites. The Python requests library must be installed for the script to run properly.

Get Answer

Build a Web Services URL from this Search >>

Give this search a name (optional)

Give this search a weight (optional)

Tools   My Workspace   Data

Apollo

BLAST (multi-query capable

Sequence retrieval

Web services

Find web services help info here.

DOWNLOAD A SINGLE GENE

The download link at the top of gene record pages implements the download tool where text, bed, fasta and GFF3 files can be retrieved. Similar download options are available from other record pages such as SNP and metabolic pathway records.
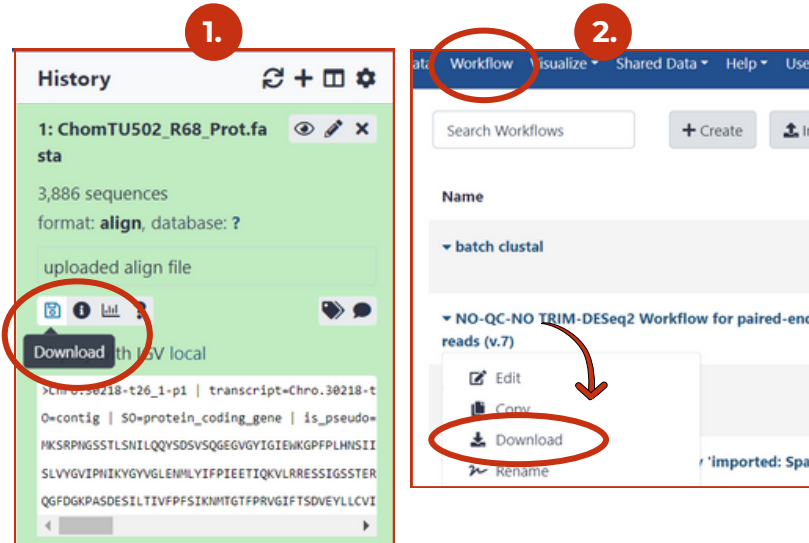
- Text - choose from columns and/or tables
- BED - coordinates of sequences, configurable
- FASTA - sequence retrieval, configurable
- GFF3: Gene models and optional sequences

Add to basket 🛒   Add to favorites ⭐   Download Gene ⬇

PF3D7_1133400   apical membrane antigen 1

Name: AMA1
Gene Type: protein coding gene
Biotype Classification: protein_coding
Chromosome: 11
Location: Pf3D7_11_v3:1,292,966..1,296,696(+)

Shortcuts

Synteny

DOWNLOAD GALAXY DATA

Files and workflows can be downloaded from VEuPathDB Galaxy and subsequently uploaded to a different server.

1. To download files (1.), open the tile in the history panel and choose Download. Histories need be be downloaded one file at a time.
2. To download a workflow (2.), go to the Workflow tab, choose the workflow name and click Download

1.

History   🔄 ➕ ▢ ⚙

1: ChomTU502_R68_Prot.fasta   👁 ✏ ✖

3,886 sequences
format: **align**, database: **?**

uploaded align file

📄 ℹ 📊 ⚓   🏷 💬
Download   th IGV local

>ChM 0.30218-t26_1-p1 | transcript=Chro.30218-t
0=contig | SO=protein_coding_gene | is_pseudo=
MKSRPNGSSTLSNILQQYSDSVSQGEGVGYIGIEWKGPFPLHNSII
SLVYGVIPNIKYGYVGLENMLYIFPIEETIQKVLRRESSIGSSTER
QGFDGKPASDESILTIVFPFSIKNMTGTFPRVGIFTSDVEYLLCVI

2.

Workflow   Visualize ▾   Shared Data ▾   Help ▾   User

Search Workflows   ➕ Create   ⬆ In

Name

▾ batch clustal

▾ NO-QC-NO TRIM-DESeq2 Workflow for paired-end reads (v.7)

✏ Edit
📋 Copy
⬇ Download
✂ Rename

'imported: Spa

DOWNLOAD SAVED STRATEGIES

Your VEuPathDB saved strategies are stored in the **All tab** of the **My Strategies** page. Saved searches need to be rerun before the results can be downloaded. Click the strategy name to rerun and then follow the instructions for **DOWNLOAD DATA FOR MULTIPLE GENES**. To download only IDs, choose to download a custom table with no columns checked (Clear All columns).
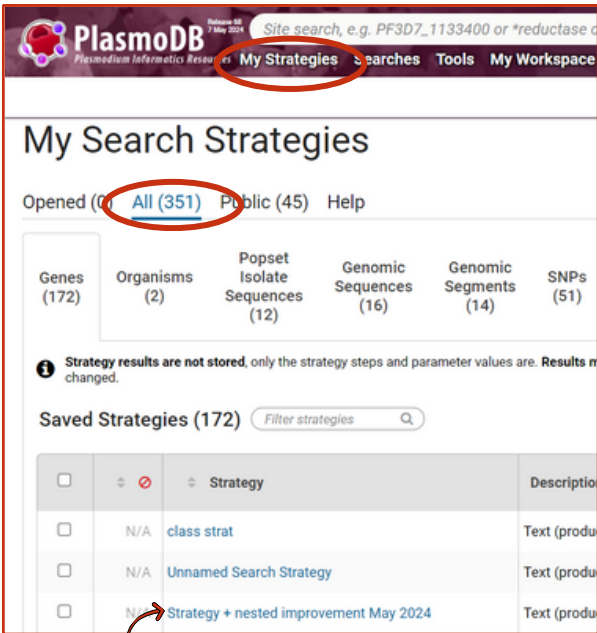
Click the name to rerun the saved strategy

DOWNLOAD JBROWSE TRACK DATA

JBrowse track data can be downloaded using the Save Track Data feature. Open the dropdown menu associated with the track name and choose the Save Track Data tool. Options to configure the Regions and data format are offered.