

## What is Galaxy?

Galaxy is an open, web-based platform for data analysis under the FAIR principles of data sharing and re-use. Galaxy is an open-source platform that allows you to perform, reproduce, and share complete analyses without the use of command line scripting. The VEuPathDB project developed its own Galaxy instance in collaboration with Globus.

The VEuPathDB Galaxy offers pre-loaded genomes, pre-configured workflows and other tools for private data analysis and display. A custom-built set of tools also allows the ability to export Galaxy results into private workspaces within VEuPathDB sites (My Workspace > My data sets section). The datasets within the “My data sets” workspace can be explored using the FungiDB interface and tools and cross-referenced with the public data integrated in FungiDB.

VEuPathDB Galaxy access requires an account with FungiDB/VEuPathDB. The account is free and can be used to sign-in into any VEuPathDB genomics site.

The Galaxy instance is not meant for long term data storage. Datasets are automatically deleted after 60 days. To save your data, download your analysis results locally and then *delete and purge* files to free up space for your next analysis.

The Galaxy project offers extensive learning materials that can be accessed here:

[https://wiki.galaxyproject.org/Learn#Galaxy\\_101](https://wiki.galaxyproject.org/Learn#Galaxy_101)

**Important:** The Galaxy module consists of RNA-Seq and SNP analysis modules. These are concurrent sessions. This exercise will be carried out in groups of 4 people using the workshop Galaxy instance. Please do not use live FungiDB.org for this exercise. The detailed tutorials for both modules are available to all course participants.

## Variant Calling analysis, Part I.

### Learning objectives:

- Become familiar with the VEuPathDB Galaxy workspace.
- Upload raw data into Galaxy workspace and run a pre-configured SNP workflow

For this exercise, we will retrieve raw sequence files from the “shared history” section in VEuPathDB Galaxy and then run files through a pre-configured RNA-Seq workflow that will align the data to a reference genome, calculate expression values and determine differential expression.

**Important:** We will be working in groups of four people but only one person in each group should download data and deploy the pre-configured workflow. The other members’ roles are to ensure that the correct datasets are used and that the correct workflow parameters are selected. In the Part 2 of this exercise, everyone will get a copy of the workflow output and practice how to perform data analysis.

- **Access the VEuPathDB Galaxy workshop instance.**

If you do not have an account with VEuPathDB/FungiDB, please create one now.

1. Click on the following URL to begin:  
<https://veupathdb1.globusgenomics.org/>
2. On the next page, you will be asked to define your organization. Choose the “VEuPathDB” option and click on the “Continue” button.
3. If you are not already logged into VEuPathDB, you will be prompted to do so.
4. Click on “Continue” on the next page (no need to link an existing account).
5. Select the “non-profit” option and agree to the Terms of Service. Click continue.
6. The next page will ask for permissions required to use this Galaxy instance. Click on “Allow”

**1** <https://veupathdb1.globusgenomics.org/>

Log in to use veupathdb1

Use your existing organizational login  
e.g., university, national lab, facility, project

**2** VEuPathDB

By selecting Continue, you agree to Globus [terms of service](#) and [privacy policy](#).

**Continue**

OR

Sign in with Google Sign in with ORCID iD

Didn't find your organization? Then use [Globus ID to sign in](#). ([What's this?](#))

**3**

**VEuPathDB**  
Eukaryotic Pathogens, Vector & Host  
Informatics Research

Please log in

Username or Email:   
Password:

**Login** **Cancel**

[Forgot Password?](#) [Register/Subscribe](#)

Visit our partner Bioinformatics Resource Center, [BV-BRC](#)

**Welcome – You've Successfully Logged In**

This is the first time you are accessing Globus with your **VEuPathDB** login.

If you have previously used Globus with another login you can link it to your **VEuPathDB** login. When linked, both logins will be able to access the same Globus account permissions and history.

**4** **Continue** **Link to an existing account** [Why should I link accounts?](#)

**Complete Your Sign Up For**  
**test@veupathdb.org**

Name   
Email   
Organization

Account will be used for

☒ non-profit research or educational purposes **5**  
☐ commercial purposes

☒ I have read and agree to the [Globus Terms of Service](#) and [Privacy Policy](#).

**Continue**

\* This field is specified by the identity provider, and cannot be modified by Globus. If you change it with your identity provider, it will propagate to Globus the next time you log in.

**6**

veupathdb1 would like to:

- ☒ View your identity ⓘ
- ☒ Manage data using Globus Transfer ⓘ
- ☒ View your email address ⓘ
- ☒ View identity details ⓘ

To work, the above will need to: ✓

By clicking "Allow", you allow **veupathdb1** (this client has not provided terms of service or a privacy policy to Globus) to use the above listed information and services. You can rescind this and other [consents](#) at any time.

**Allow** **Deny**

## The anatomy of the VEuPathDB Galaxy landing page.

The workspace has four major components:

1. The top menu controls the main interface, provides access to the landing page, shared data, public and private workflows & more.
2. The left panel has a list of available tools where the VEuPathDB export tools are listed at the top.
3. The main welcome (landing) page is the interactive interface that houses pre-configured workflows, workflows editor, etc.
4. The panel on the right provides access to histories, deleted datasets, and other useful functions, including options to delete and purge datasets.

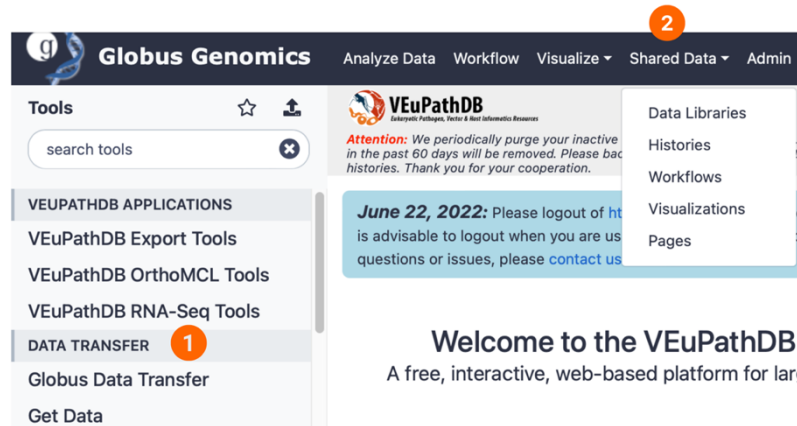
The screenshot shows the VEuPathDB Galaxy landing page. The interface is divided into four main sections, each highlighted with a red circle and a number:

- 1 (Top Menu):** The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Admin', 'Help', and 'User'. A 'Tools' button is also present.
- 2 (Left Panel):** A sidebar containing a search bar and a list of tools categorized under 'VEuPathDB APPLICATIONS', 'DATA TRANSFER', 'NGS VISUALIZATION', and 'NGS APPLICATIONS'.
- 3 (Main Content Area):** The central workspace displaying a welcome message: 'Welcome to the VEuPathDB Galaxy Site. A free, interactive, web-based platform for large-scale data analysis.' Below this, it lists 'Get started with VEuPathDB pre-configured workflows:' and provides links for 'OrthoMCL', 'RNA-seq', 'Examine genome coverage and calculate TPM for each gene', 'Identify genes with statistically significant expression differences between two samples', and 'Variant calling'.
- 4 (Right Panel):** A sidebar for 'History' and 'search datasets', showing a list of datasets and their associated actions (e.g., 'Delete', 'Purge', 'Export').

Don't see a tool you need for your research? – Let us know by sending an email to [help@fungidb.org](mailto:help@fungidb.org)

## Importing data for your workflow.

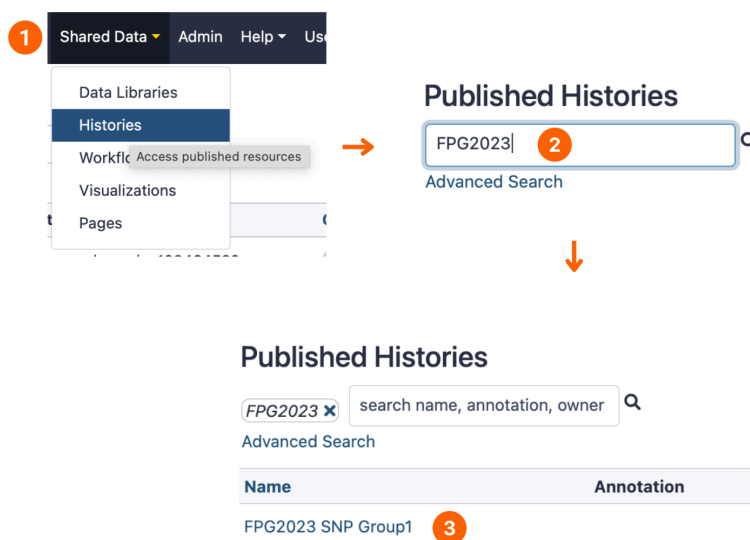
There are multiple ways to import data into your Galaxy workspace. You can transfer data via tools located under the “Data Transfer” section in menu on the left (1). You can also transfer data from the “Shared Data” section in the main menu (2). The latter provides access to pre-loaded raw data, publicly shared workflows, or workflow results (histories), etc.



For this exercise, pre-loaded raw files should be imported from the “Shared Data” > Histories.

Only one person per each group should import data files and deploy an SNP workflow. Everyone will practice data analysis in NGS Part 2 module. For group assignments, see below.

- **Import data for your SNP workflow via the Shared histories option.**
  1. From the top menu, select “Shared Data > Histories” option.
  2. Filter all public workflows on “FPG2023” .
  3. Click on the history link that correspond to your group number (e.g., FPG2023 SNP Group1) to import the data into your Galaxy workspace.



## Group assignments

**Groups 1** *Aspergillus fumigatus*. Paired-end data. A clinical isolate from pleural fluid of a patient. Isolate: AFIS2503.

History name for download (in Galaxy)	FPG2023 SNP Group1
Ref genome (in Galaxy)	FungiDB-29_AfumigatusAf293_Genome

**Groups 2** *Aspergillus fumigatus*. Paired-end data. A clinical isolate from pleural fluid of a patient. Isolate: AFIS1415.

History name for download (in Galaxy)	FPG2023 SNP Group2
Ref genome (in Galaxy)	FungiDB-29_AfumigatusAf293_Genome

**Group 3** *Zymoseptoria tritici*. Paired-end data. An isolate collected from common wheat (*Triticum aestivum*) in Switzerland: Eschikon. Isolate: ST16CH 1A27.

History name for download (in Galaxy)	FPG2023 SNP Group3
Ref genome (in Galaxy)	FungiDB-34_ZtriticiIP0323_Genome

**Group 4** *Zymoseptoria tritici*. Paired-end data. An isolate collected from common wheat (*Triticum aestivum*) in Oregon: USA. Isolate: ORE15 Mad G1.

History name for download (in Galaxy)	FPG2023 SNP Group4
Ref genome (in Galaxy)	FungiDB-34_ZtriticiIP0323_Genome

**Group 5** *Candida auris*. Paired-end data. An isolated collected from an apple surface in India. Isolate: VPCI-F37-B-2021.

History name for download (in Galaxy)	FPG2023 SNP Group5
Ref genome (in Galaxy)	FungiDB-37_CaurisB8441_Genome

**Group 6** *Candida auris*. Paired-end data. An isolated collected from an apple surface in India. Isolate: VPCI-F1-A-2020.

History name for download (in Galaxy)	FPG2023 SNP Group6
Ref genome (in Galaxy)	FungiDB-37_CaurisB8441_Genome

Once the data files have been transferred into your galaxy history you need to choose a workflow appropriate for your data (paired or single -read).

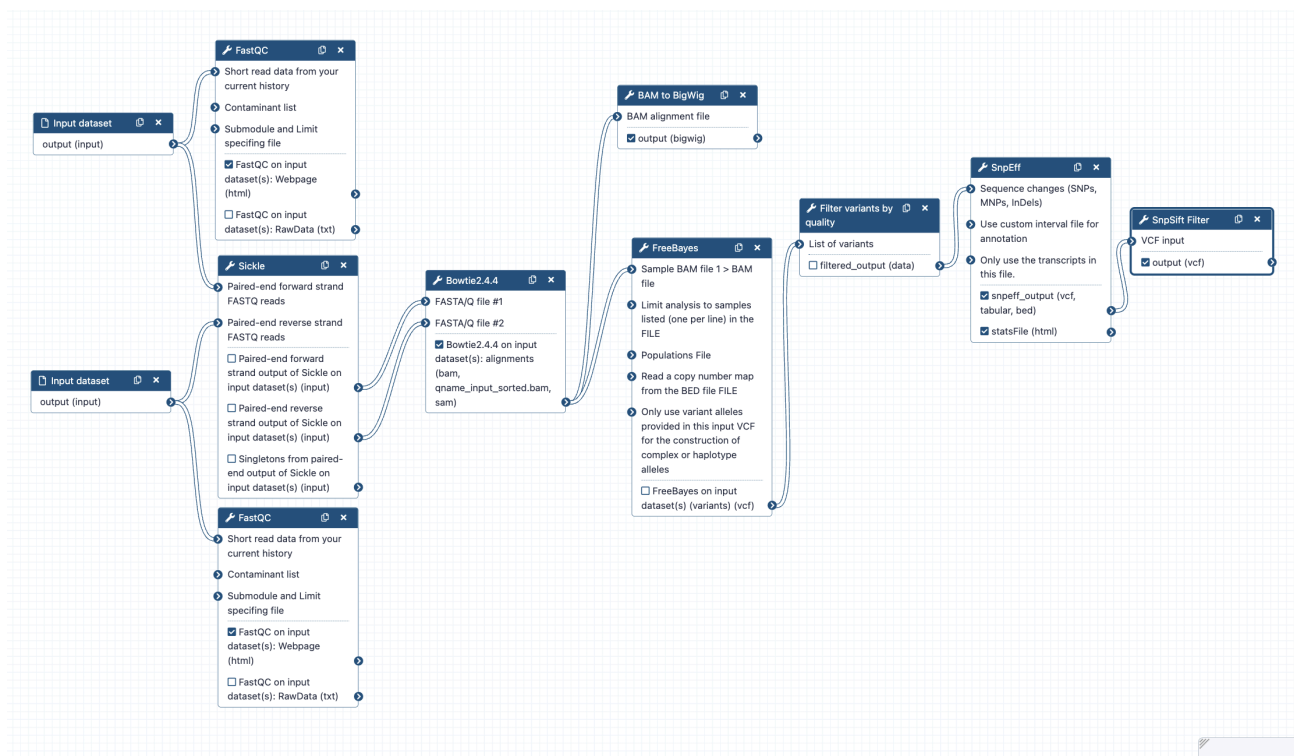
## Variant calling

Use the following workflows to analyze your FASTQ detection, SnpEff to evaluate the effect of variants analyzed in Galaxy or downloaded to your computer

- [Workflow for single-end reads](#)
- [Workflow for paired-end reads](#)

The pre-configured workflows follow these steps:

- Determine quality of the reads in your files and generates FASTQC reports.
- Trim reads based on their quality scores.
- Align reads to a reference genome using Bowtie2 and generating coverage plots.
- Sort alignments with respect to their chromosomal positions.
- Detect variants using FreeBayes.
- Filter SNP candidates.
- Analyze and annotate variants, and calculation of the effects via SnpEff.



- **Set workflow parameters.**

1. For paired-end data, make sure that the input steps are set to the xxxx **1**.fastq.gz and xxxx **2**.fastq.gz as by default both have the same one selected. Here is an example (disregards “7” and “8” and it simply refers to the ordered file number).

*Note: for single read data, you will have only one file.*

2. Select the correct reference genome for Bowtie2 (see group assignment above).
3. Select the correct reference genome for FreeBayes (see group assignment above).
4. Select the correct reference genome for SnpEff (see group assignment above).
5. Click Run Workflow.

Workflow: imported: Variant Calling Workflow for paired-end reads (v.7)

**5** ▶ Run Workflow

History Options

Send results to a new history

Yes No

1: Input dataset - 1

1: SRR11785185\_1.fastq.gz

2: Input dataset - 8

2: SRR11785185\_2.fastq.gz

3: FastQC - 2 (Galaxy Version FASTQC: 0.11.3)

4: Sickle (Galaxy Version 1.33.2)

5: FastQC - 9 (Galaxy Version FASTQC: 0.11.3)

6: Bowtie2.4.4 (Galaxy Version 2.4.4+galaxy0)

7: BAM to BigWig - 11 (Galaxy Version 0.2.0)

8: FreeBayes - 6 (Galaxy Version FREEBAYES: v0.9.21-19-gc003c1e; SAMTOOLS: 0.1.18)

9: Filter variants by quality - 7 (Galaxy Version 1.0.0)

10: SnpEff - 8 (Galaxy Version SNPEFF: snpEff\_3.6; JAVA: 1.8.0)

11: SnpSift Filter - 10 (Galaxy Version latest)

Select reference genome

FungiDB-29\_AfumigatusAf293\_Genome

If your genome of interest is not listed, contact the Galaxy team