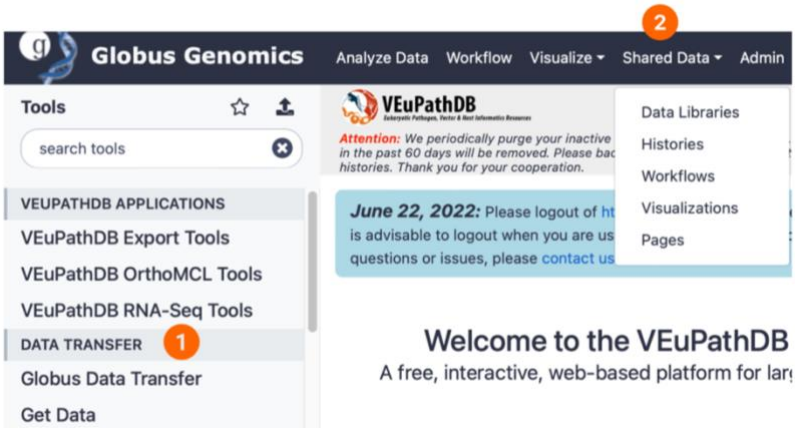
Variant Calling analysis, Part I.

Learning objectives:

- Upload raw data into Galaxy workspace and run a pre-configured SNP workflow

Importing data for your workflow.

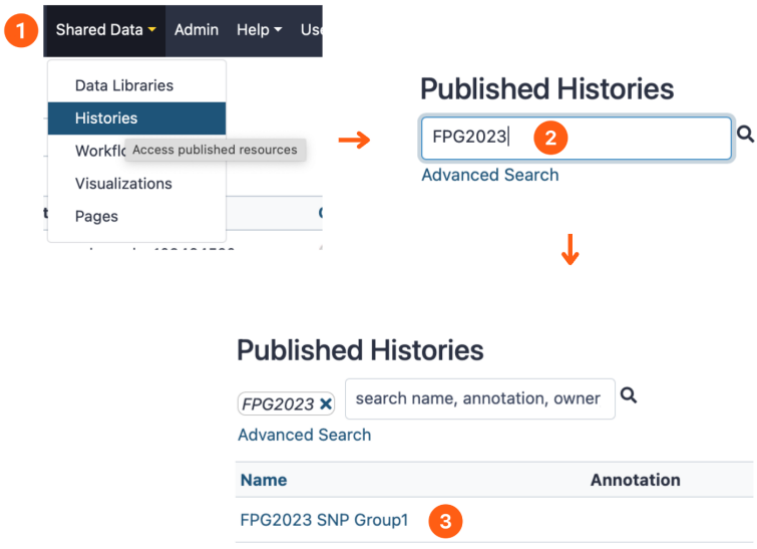
There are multiple ways to important data into your Galaxy workspace. You can transfer data via tools located under the "Data Transfer" section in menu on the left (1). You can also transfer data from the "Shared Data" section in the main menu (2). The latter prides access to pre-loaded raw data, publicly 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_ZtriticilPO323_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_ZtriticilPO323_Genome

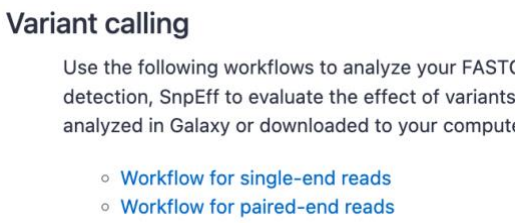
Group 5 *Candida auris*. Paired-end data. An isolated collected from an apple surface in India. Isolate: VPCI-F37-B-2021.

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.

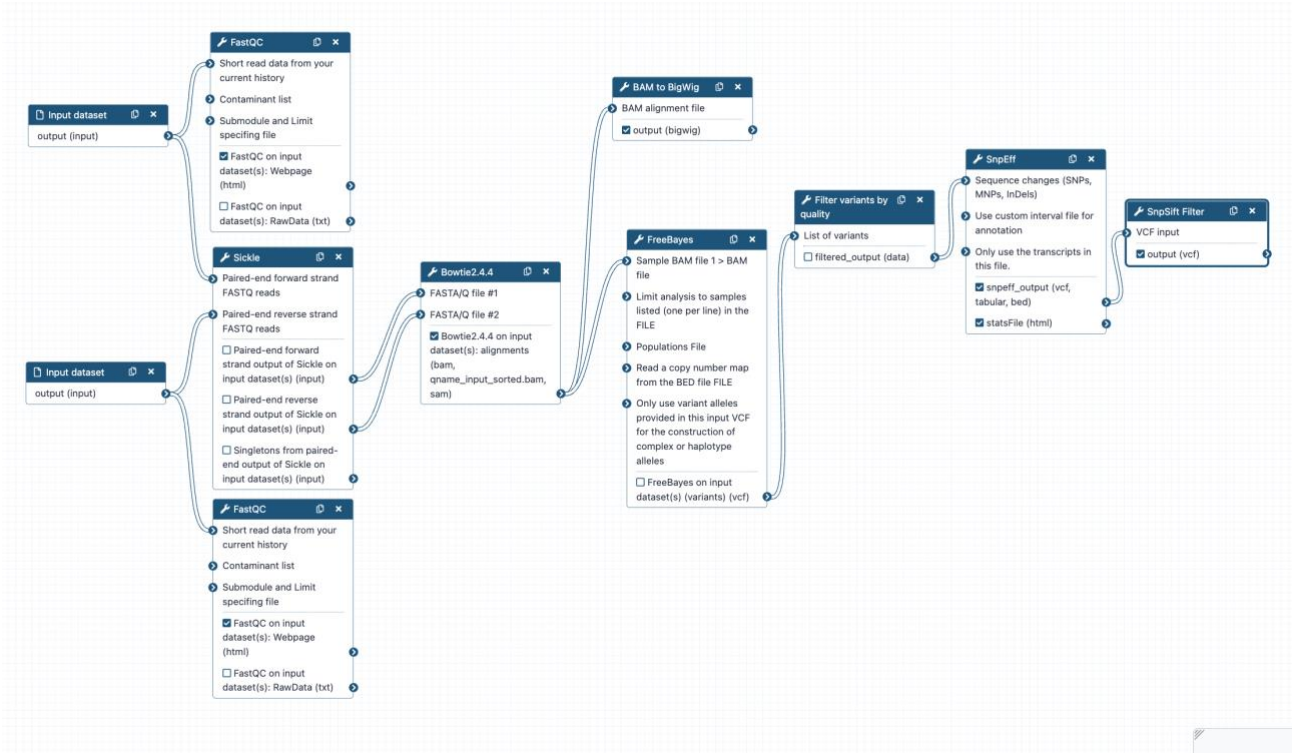
History name for download (in Galaxy)	FPG2023 SNP Group6
Ref genome (in Galaxy)	FungiDB-37_CaurisB8441_Genome

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).



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 of 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.