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

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.

RNA sequence data analysis via VEuPathDB Galaxy, Part I

Learning objectives:

- Become familiar with the VEuPathDB Galaxy workspace.
- Create collections of datasets from the pre-loaded data.
- Run a pre-configured RNA-Seq workflow.

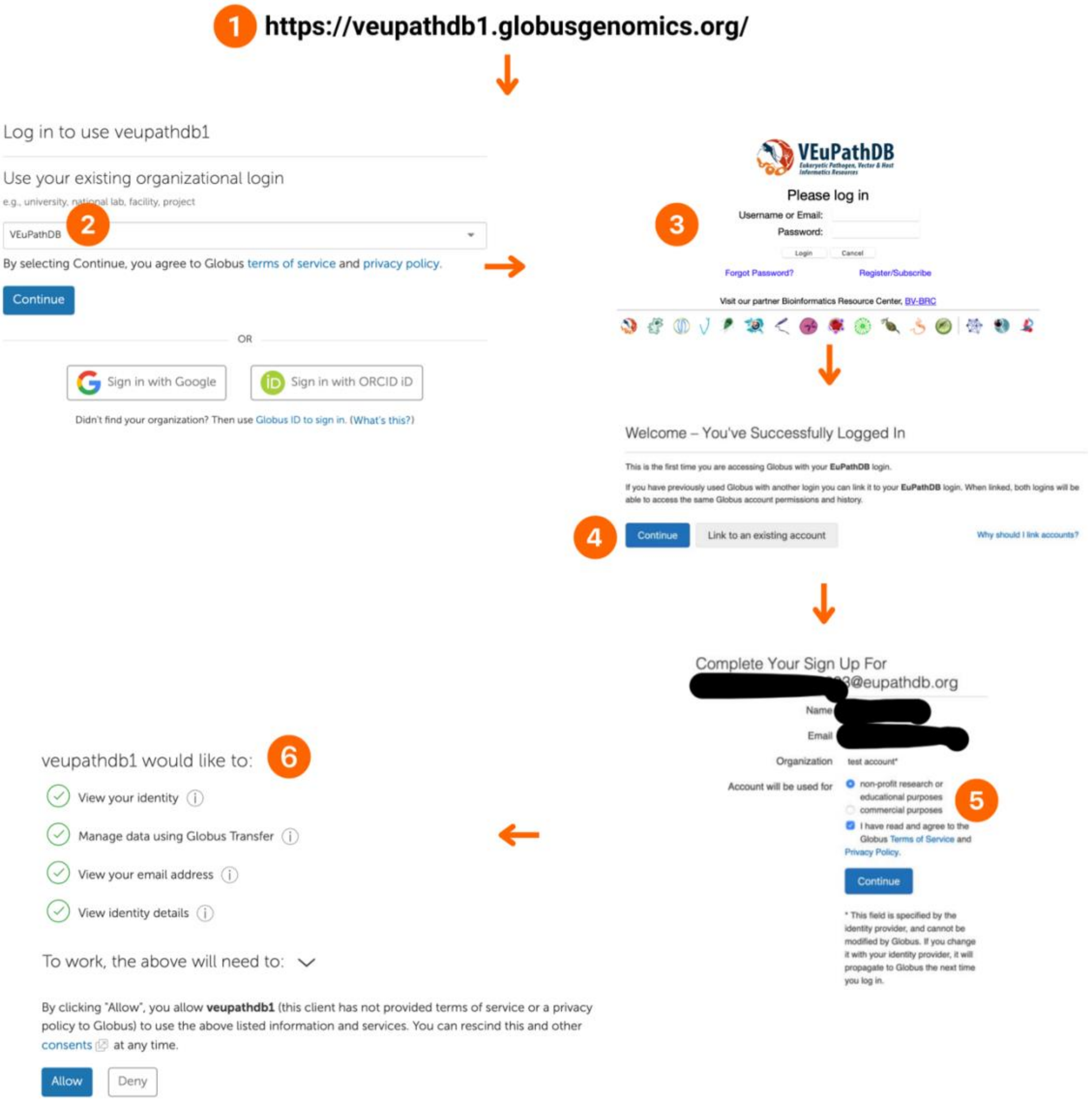
For this exercise, we will retrieve raw sequence files from the "shared history" section in VEuPathDB Galaxy and then run files through a pre-configured RNA-Seq workflow that will align the data to a reference genome, calculate expression values and determine differential expression.

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

VEuPathDB 2

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?)

VEuPathDB
Eukaryotic Pathogen, Vector & Host Informatics Resources

3 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 **EuPathDB** login.

If you have previously used Globus with another login you can link it to your **EuPathDB** 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
[redacted]@eupathdb.org

Name [redacted]

Email [redacted]

Organization test account*

Account will be used for ● non-profit research or educational purposes ○ commercial purposes 5

☑ 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.

veupathdb1 would like to: 6

✓ View your identity ⓘ

✓ Manage data using Globus Transfer ⓘ

✓ View your email address ⓘ

✓ View identity details ⓘ

To work, the above will need to:

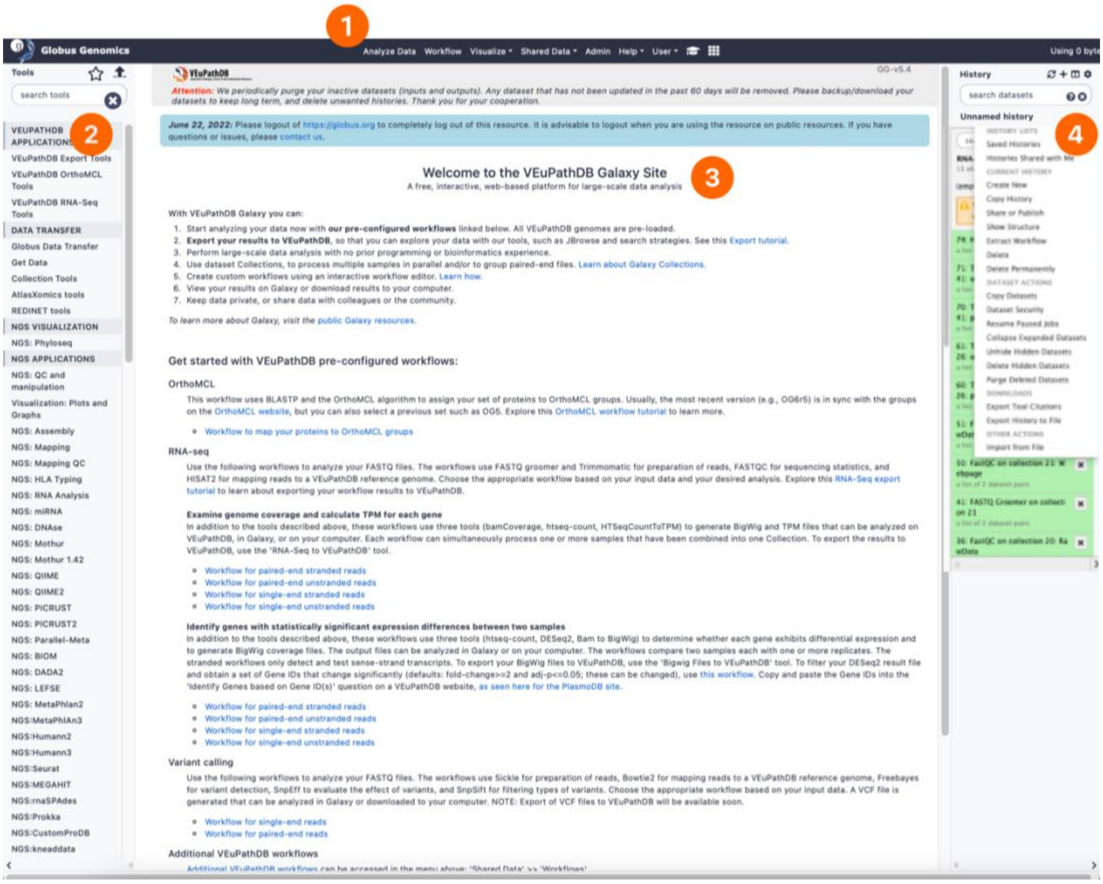
By clicking "Allow", you allow **veupathdb1** (this client has not provided terms of service or a privacy policy to Globus) to use the above listed information and services. You can rescind this and other consents at any time.

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.



Attention: We periodically purge your inactive datasets (inputs and outputs). Any dataset that has not been updated in the past 60 days will be removed. Please backup/download your datasets to keep long term, and delete unwanted histories. Thank you for your cooperation.

June 22, 2022: Please logout of https://globus.org to completely log out of this resource. It is advisable to logout when you are using the resource on public resources. If you have questions or issues, please contact us.

Welcome to the VEuPathDB Galaxy Site

A free, interactive, web-based platform for large-scale data analysis

With VEuPathDB Galaxy you can:

1. Start analyzing your data now with **our pre-configured workflows** linked below. All VEuPathDB genomes are pre-loaded.
2. **Export your results to VEuPathDB**, so that you can explore your data with our tools, such as JBrowse and search strategies. See this Export tutorial.
3. Perform large-scale data analysis with no prior programming or bioinformatics experience.
4. Use dataset Collections, to process multiple samples in parallel and/or to group paired-end files. Learn about Galaxy Collections.
5. Create custom workflows using an interactive workflow editor. Learn how.
6. View your results on Galaxy or download results to your computer.
7. Keep data private, or share data with colleagues or the community.

To learn more about Galaxy, visit the public Galaxy resources.

Get started with VEuPathDB pre-configured workflows:

OrthoMCL

This workflow uses BLASTP and the OrthoMCL algorithm to assign your set of proteins to OrthoMCL groups. Usually, the most recent version (e.g., OG6r5) is in sync with the groups on the OrthoMCL website, but you can also select a previous set such as OG5. Explore this OrthoMCL workflow tutorial to learn more.

- Workflow to map your proteins to OrthoMCL groups

RNA-seq

Use the following workflows to analyze your FASTQ files. The workflows use FASTQ groomer and Trimmomatic for preparation of reads, FASTQC for sequencing statistics, and HISAT2 for mapping reads to a VEuPathDB reference genome. Choose the appropriate workflow based on your input data and your desired analysis. Explore this RNA-Seq export tutorial to learn about exporting your workflow results to VEuPathDB.

Examine genome coverage and calculate TPM for each gene

In addition to the tools described above, these workflows use three tools (bamCoverage, htseq-count, HTSeqCountToTPM) to generate BigWig and TPM files that can be analyzed on VEuPathDB, in Galaxy, or on your computer. Each workflow can simultaneously process one or more samples that have been combined into one Collection. To export the results to VEuPathDB, use the 'RNA-Seq to VEuPathDB' tool.

- Workflow for paired-end stranded reads
- Workflow for paired-end unstranded reads
- Workflow for single-end stranded reads
- Workflow for single-end unstranded reads

Identify genes with statistically significant expression differences between two samples

In addition to the tools described above, these workflows use three tools (htseq-count, DESeq2, Bam to BigWig) to determine whether each gene exhibits differential expression and to generate BigWig coverage files. The output files can be analyzed in Galaxy or on your computer. The workflows compare two samples each with one or more replicates. The stranded workflows only detect and test sense-strand transcripts. To export your BigWig files to VEuPathDB, use the 'Bigwig Files to VEuPathDB' tool. To filter your DESeq2 result file and obtain a set of Gene IDs that change significantly (defaults: fold-change>=2 and adj-p<=0.05; these can be changed), use this workflow. Copy and paste the Gene IDs into the 'Identify Genes based on Gene ID(s)' question on a VEuPathDB website, as seen here for the PlasmoDB site.

- Workflow for paired-end stranded reads
- Workflow for paired-end unstranded reads
- Workflow for single-end stranded reads
- Workflow for single-end unstranded reads

Variant calling

Use the following workflows to analyze your FASTQ files. The workflows use Sickle for preparation of reads, Bowtie2 for mapping reads to a VEuPathDB reference genome, Freebayes for variant detection, SnpEff to evaluate the effect of variants, and SnpSift for filtering types of variants. Choose the appropriate workflow based on your input data. A VCF file is generated that can be analyzed in Galaxy or downloaded to your computer. NOTE: Export of VCF files to VEuPathDB will be available soon.

- Workflow for single-end reads
- Workflow for paired-end reads

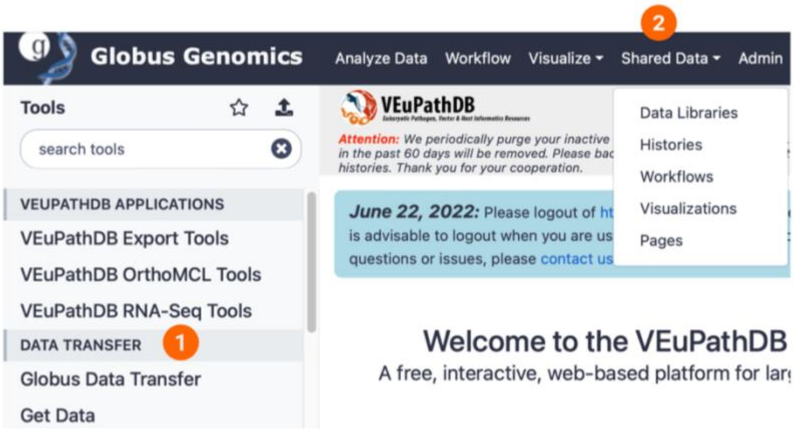
Additional VEuPathDB workflows

Additional VEuPathDB workflows can be accessed in the menu above: 'Shared Data' >> 'Workflows'

Don’t see a tool you need for your research? – Let us know by sending an email to help@fungidb.org

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.



Attention: We periodically purge your inactive in the past 60 days will be removed. Please bac histories. Thank you for your cooperation.

June 22, 2022: Please logout of ht is advisable to logout when you are us questions or issues, please contact us

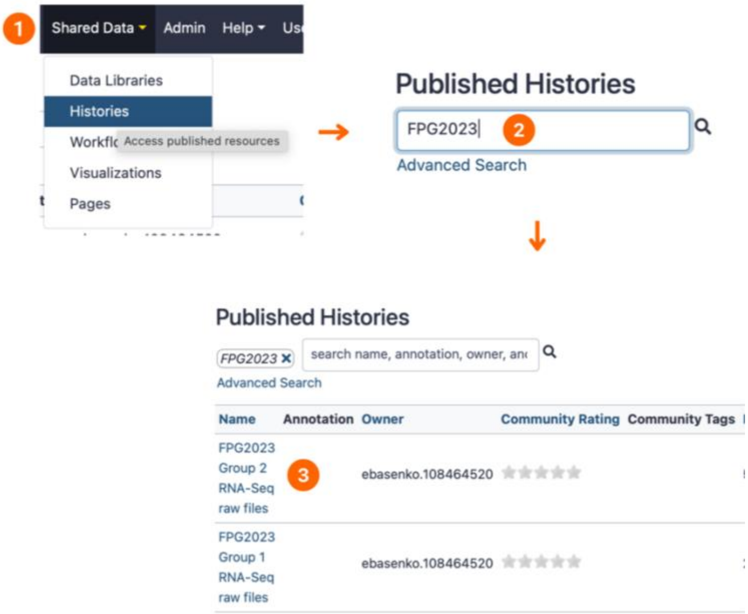
Welcome to the VEuPathDB

A free, interactive, web-based platform for lar

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 RNA-Seq workflow. Everyone will practice data analysis in NGS Part 2 module. For group assignments, see below.

- **Import data for your RNA-Seq 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 to import the data into your Galaxy workspace.



Name	Annotation	Owner	Community Rating	Community Tags
FPG2023 Group 2 RNA-Seq raw files		ebasenko.108464520	☆☆☆☆☆	
FPG2023 Group 1 RNA-Seq raw files		ebasenko.108464520	☆☆☆☆☆	

Group assignments (see more information about the files below)

Groups 1 & 2 *Aspergillus fumigatus*. Paired-end data. Analyze transcriptomes from cells incubated in human blood (B) and defined minimal media (M) for 30 and 180 min.

Group Number	**1**	**2**
Comparison	M30 vs B30	B30 vs B180
History name for download (in Galaxy)	FPG2023 Group 1 RNA-Seq raw files	FPG2023 Group 2 RNA-Seq raw files
Ref genome (in Galaxy)	FungiDB-29_AfumigatusAf293_Genome	

Reference: PMID: 26311470 BioProject: PRJNA287921

Group 3 *Candida parapsilopsis*. Paired-end data. Analyze transcriptomes from cells grown under planktonic and biofilm-inducing conditions. Control: planktonic.

Comparison	Planktonic vs Biofilm
History name for download (in Galaxy)	FPG2023 Group3 RNA-Seq raw files
Ref genome (in Galaxy)	FungiDB-42_CparapsilosisCDC317_Genome

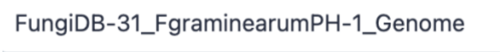
Reference: PMID: 25233198 BioProject: PRJNA246482

Group 4 *Coccidioides posadasii*. Single read data. Analyze transcriptomes from mycelia (non-pathogenic stage) and spherules (pathogenic stage).

Comparison	Mycelia vs Spherules
History name for download (in Galaxy)	FPG2023 Group 4 RNA-Seq raw files
Ref genome (in Galaxy)	FungiDB-61_CposadasiiSilveira2022_Genome

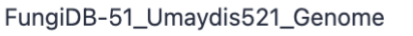
Reference: PMID: 22911737 BioProject: PRJNA169242

Group 5 *Fusarium graminearum*. Paired-end data. Analyze spore and mycelial transcriptomes.

Comparison	Spores vs Mycelia
History name for download (in Galaxy)	FPG2023 Group 5 RNA-Seq raw files
Ref genome (in Galaxy)	 FungiDB-31_FgraminearumPH-1_Genome

Reference: PMID: 24625133 BioProject: PRJNA239711

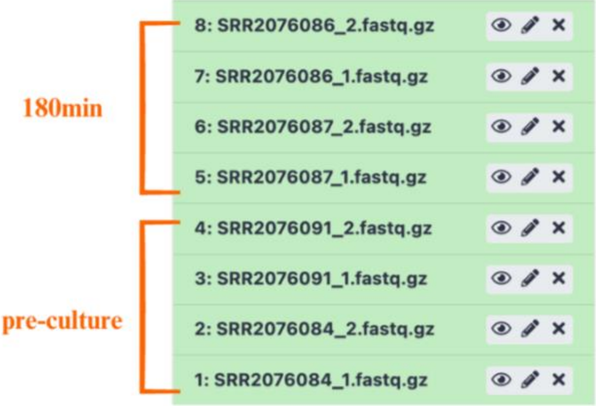
Group 6 *Ustilago maydis*. Paired-end data. Analyze transcriptomes from plant-associated development samples (axenic culture vs 12 days post infection (dpi)).

Comparison	0h vs 12 dpi
History name for download (in Galaxy)	FPG2023 Group 6 RNA-Seq raw files
Ref genome (in Galaxy)	 FungiDB-51_Umaydis521_Genome

Reference: PMID: 33653886 BioProject: PRJNA407369

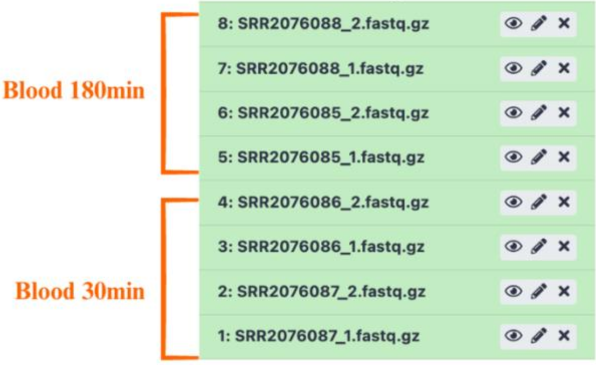
Guide to FPG2023 RNA-Seq histories and file organisation.

Group 1



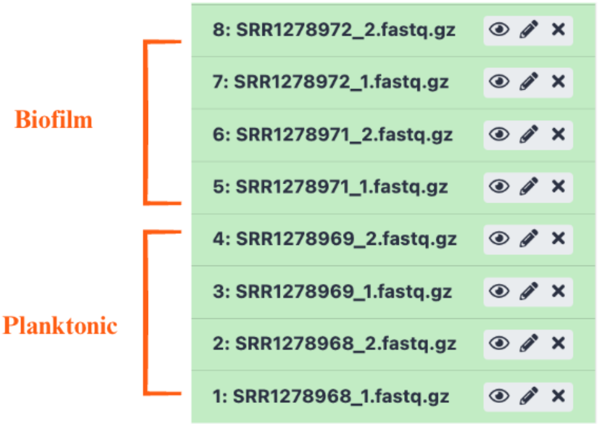
8: SRR2076086_2.fastq.gz
7: SRR2076086_1.fastq.gz
6: SRR2076087_2.fastq.gz
5: SRR2076087_1.fastq.gz
4: SRR2076091_2.fastq.gz
3: SRR2076091_1.fastq.gz
2: SRR2076084_2.fastq.gz
1: SRR2076084_1.fastq.gz

Group 2



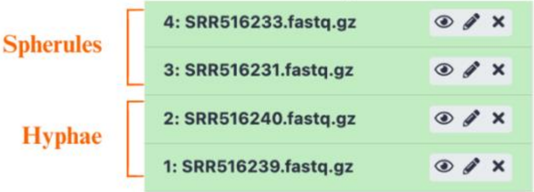
8: SRR2076088_2.fastq.gz
7: SRR2076088_1.fastq.gz
6: SRR2076085_2.fastq.gz
5: SRR2076085_1.fastq.gz
4: SRR2076086_2.fastq.gz
3: SRR2076086_1.fastq.gz
2: SRR2076087_2.fastq.gz
1: SRR2076087_1.fastq.gz

Group 3



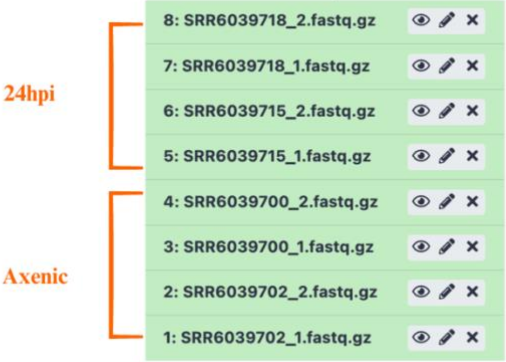
8: SRR1278972_2.fastq.gz
7: SRR1278972_1.fastq.gz
6: SRR1278971_2.fastq.gz
5: SRR1278971_1.fastq.gz
4: SRR1278969_2.fastq.gz
3: SRR1278969_1.fastq.gz
2: SRR1278968_2.fastq.gz
1: SRR1278968_1.fastq.gz

Group 4



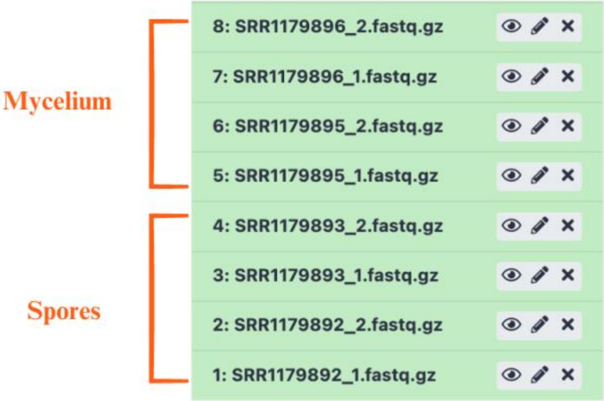
4: SRR516233.fastq.gz
3: SRR516231.fastq.gz
2: SRR516240.fastq.gz
1: SRR516239.fastq.gz

Group 6



8: SRR6039718_2.fastq.gz
7: SRR6039718_1.fastq.gz
6: SRR6039715_2.fastq.gz
5: SRR6039715_1.fastq.gz
4: SRR6039700_2.fastq.gz
3: SRR6039700_1.fastq.gz
2: SRR6039702_2.fastq.gz
1: SRR6039702_1.fastq.gz

Group 5



8: SRR1179896_2.fastq.gz
7: SRR1179896_1.fastq.gz
6: SRR1179895_2.fastq.gz
5: SRR1179895_1.fastq.gz
4: SRR1179893_2.fastq.gz
3: SRR1179893_1.fastq.gz
2: SRR1179892_2.fastq.gz
1: SRR1179892_1.fastq.gz

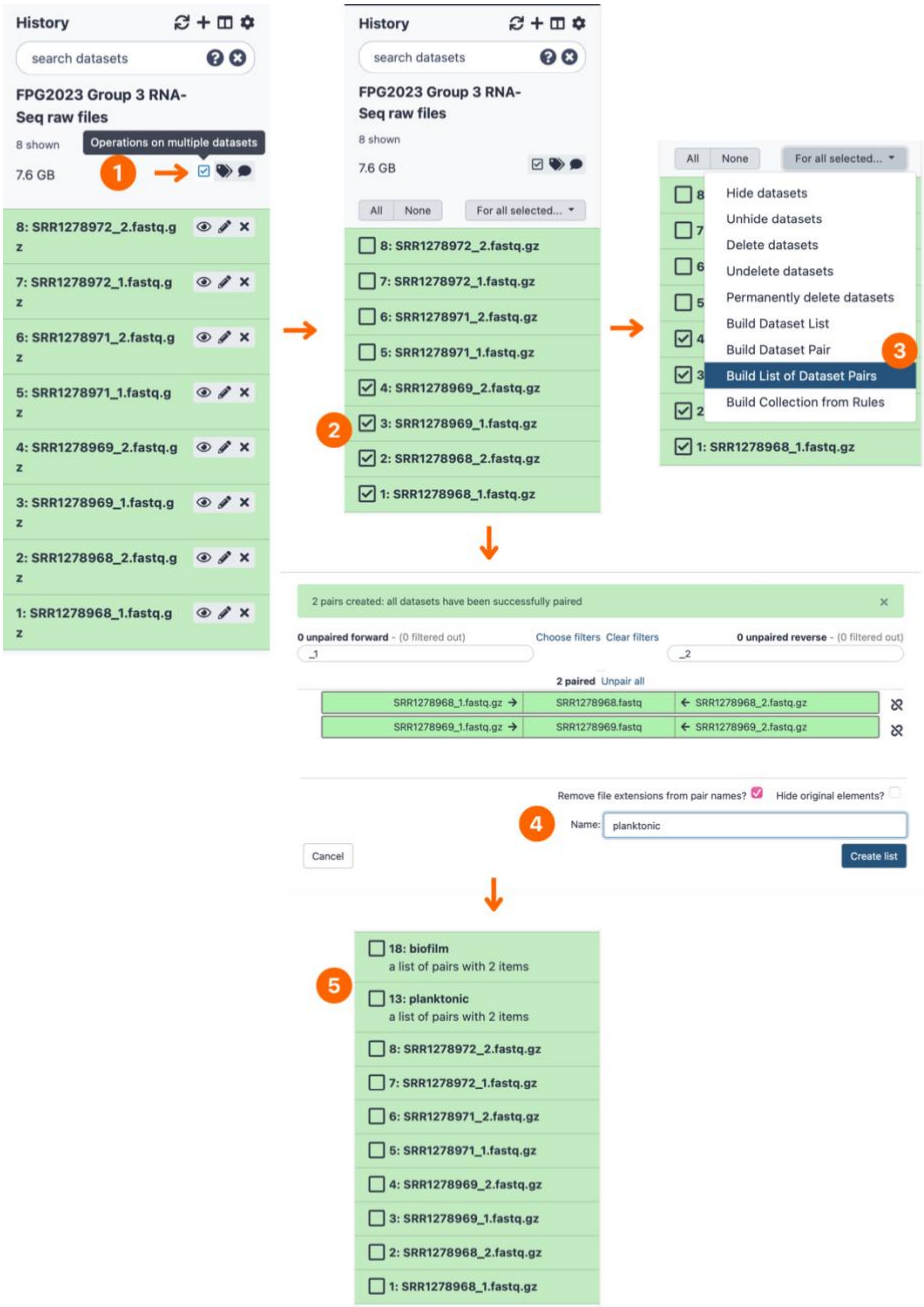
Each dataset contains two replicates. For datasets with multiple samples (e.g., containing biological replicates), it is useful to organize them into “Collections” (e.g., spore and mycelia). Organizing samples with replicates into collections also reduces the complexity Galaxy workflows.

- **Organize samples with replicates into collections:**

1. Click on the checkbox function "operation on multiple datasets".
2. Select samples that belong to the same condition (control samples will appear at the bottom, see file mapping notes for each group below).
3. Click on "For all selected" and choose "Build List of Dataset Pairs".

Note: for single read data, choose "Build Data List" option instead.

4. Name the sample (e.g. planktonic) and click "Create List". Note: Usually the correct pairs are auto selected.
5. Repeat for the comparator sample. You should end up with 2 datasets (e.g., planktonic and biofilm).



8: SRR1278972_2.fastq.gz
7: SRR1278972_1.fastq.gz
6: SRR1278971_2.fastq.gz
5: SRR1278971_1.fastq.gz
☑ 4: SRR1278969_2.fastq.gz
☑ 3: SRR1278969_1.fastq.gz
☑ 2: SRR1278968_2.fastq.gz
☑ 1: SRR1278968_1.fastq.gz

2 pairs created: all datasets have been successfully paired

0 unpaired forward - (0 filtered out) _1

0 unpaired reverse - (0 filtered out) _2

2 paired

SRR1278968_1.fastq.gz → SRR1278968.fastq ← SRR1278968_2.fastq.gz
SRR1278969_1.fastq.gz → SRR1278969.fastq ← SRR1278969_2.fastq.gz

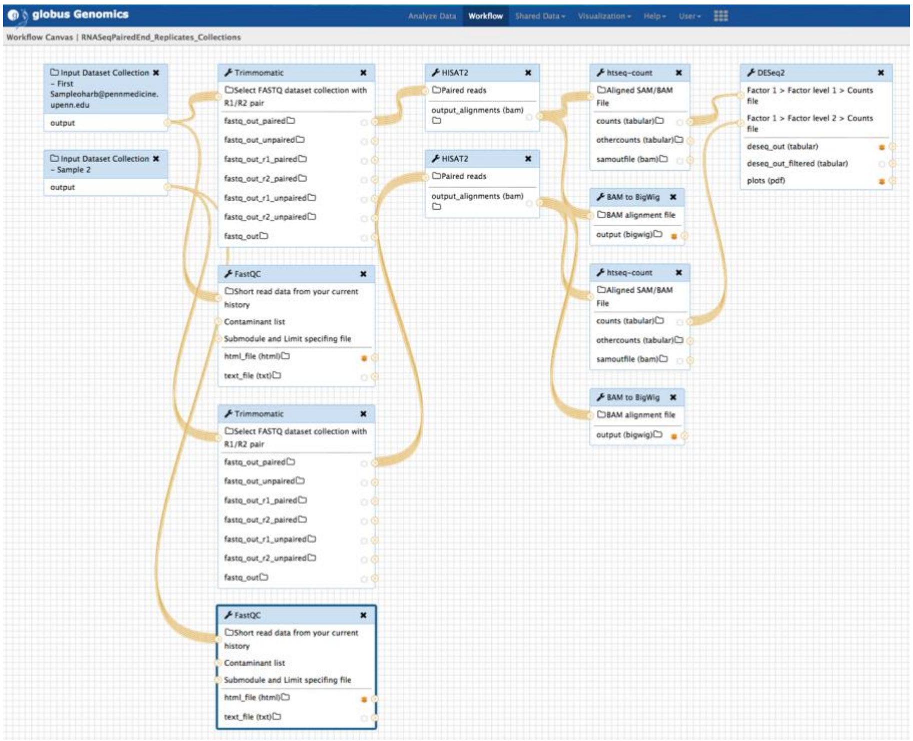
Name: planktonic

18: biofilm
a list of pairs with 2 items
13: planktonic
a list of pairs with 2 items
8: SRR1278972_2.fastq.gz
7: SRR1278972_1.fastq.gz
6: SRR1278971_2.fastq.gz
5: SRR1278971_1.fastq.gz
4: SRR1278969_2.fastq.gz
3: SRR1278969_1.fastq.gz
2: SRR1278968_2.fastq.gz
1: SRR1278968_1.fastq.gz

Running a workflow in Galaxy

You can create your own workflows in galaxy using the tools from the menu on the left. For this exercise we will use a preconfigured workflow that consists of the following steps:

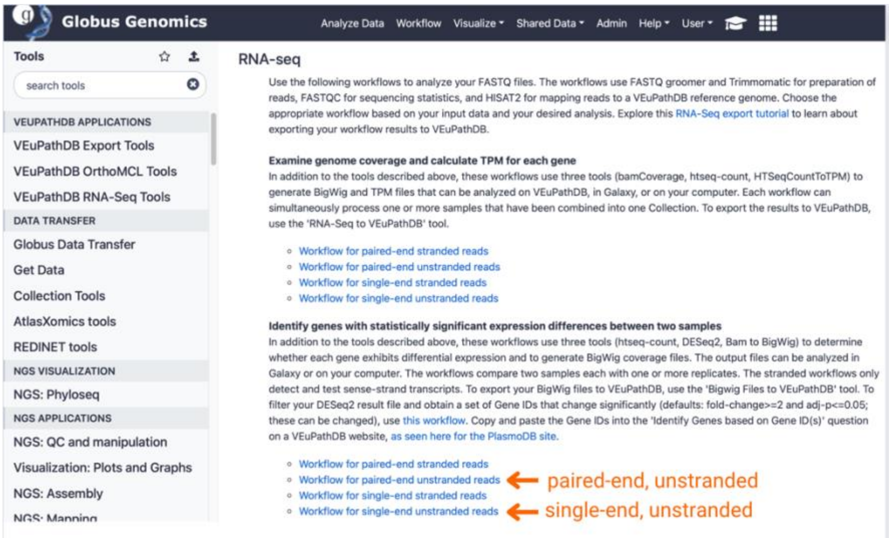
1. Input: raw data, dataset collections.
2. FASTQC: analyse for quality, generate read quality reports.
3. Trimmomatic: trims the reads based on their quality scores and adaptor sequences.
4. HISAT2: align reads to a reference and generate coverage plots.
5. HTSeq: estimate abundance (read counts per gene), generate coverage plots for JBrowse (BAM to BigWig).
6. DESeq2: differential expression of genes between samples.



- **Deploy a pre-configured workflow.**

To do this, navigate to the Galaxy home page and select the workflow appropriate for your dataset:

  - For paired-read datasets choose "Workflow for paired-end unstranded reads".
  - For single read data, choose "Workflow for single-end unstranded reads".



RNA-seq

Use the following workflows to analyze your FASTQ files. The workflows use FASTQ groomer and Trimmomatic for preparation of reads, FASTQC for sequencing statistics, and HISAT2 for mapping reads to a VEuPathDB reference genome. Choose the appropriate workflow based on your input data and your desired analysis. Explore this RNA-Seq export tutorial to learn about exporting your workflow results to VEuPathDB.

Examine genome coverage and calculate TPM for each gene

In addition to the tools described above, these workflows use three tools (bamCoverage, htseq-count, HTSeqCountToTPM) to generate BigWig and TPM files that can be analyzed on VEuPathDB, in Galaxy, or on your computer. Each workflow can simultaneously process one or more samples that have been combined into one Collection. To export the results to VEuPathDB, use the 'RNA-Seq to VEuPathDB' tool.

- Workflow for paired-end stranded reads
- Workflow for paired-end unstranded reads
- Workflow for single-end stranded reads
- Workflow for single-end unstranded reads

Identify genes with statistically significant expression differences between two samples

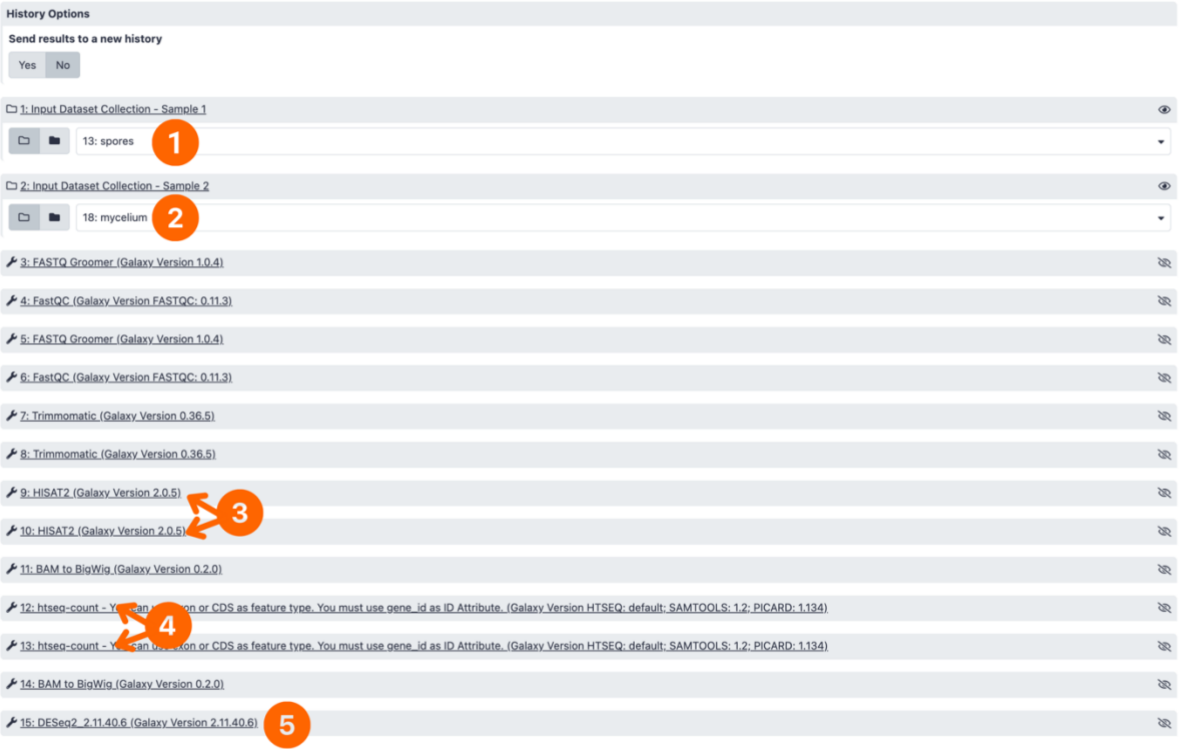
In addition to the tools described above, these workflows use three tools (htseq-count, DESeq2, Bam to BigWig) to determine whether each gene exhibits differential expression and to generate BigWig coverage files. The output files can be analyzed in Galaxy or on your computer. The workflows compare two samples each with one or more replicates. The stranded workflows only detect and test sense-strand transcripts. To export your BigWig files to VEuPathDB, use the 'Bigwig Files to VEuPathDB' tool. To filter your DESeq2 result file and obtain a set of Gene IDs that change significantly (defaults: fold-change>=2 and adj-p<=0.05; these can be changed), use this workflow. Copy and paste the Gene IDs into the 'Identify Genes based on Gene ID(s)' question on a VEuPathDB website, as seen here for the PlasmoDB site.

- Workflow for paired-end stranded reads
- Workflow for paired-end unstranded reads ← paired-end, unstranded
- Workflow for single-end stranded reads
- Workflow for single-end unstranded reads ← single-end, unstranded

- **Configure an RNA-Seq workflow.**

There are multiple steps in the workflow, but you do not need to configure all of them. For this exercise, you will need to configure the following:

1. Input dataset collection 1 (e.g., planktonic).
2. Input dataset collection 2 (e.g., biofilm).
3. Both HISAT2 steps (requires reference genome – refer to the group assignments section above for this info).
4. Both htseq-count steps (requires reference genome – refer to the group assignments section above for this info).
5. DESeq2 (requires reference genome – refer to the group assignments section above for this info).



History Options

Send results to a new history

Yes No

1: Input Dataset Collection - Sample 1

13: spores

2: Input Dataset Collection - Sample 2

18: mycelium

3: FASTQ Groomer (Galaxy Version 1.0.4)

4: FastQC (Galaxy Version FASTQC: 0.11.3)

5: FASTQ Groomer (Galaxy Version 1.0.4)

6: FastQC (Galaxy Version FASTQC: 0.11.3)

7: Trimmomatic (Galaxy Version 0.36.5)

8: Trimmomatic (Galaxy Version 0.36.5)

9: HISAT2 (Galaxy Version 2.0.5)

10: HISAT2 (Galaxy Version 2.0.5)

11: BAM to BigWig (Galaxy Version 0.2.0)

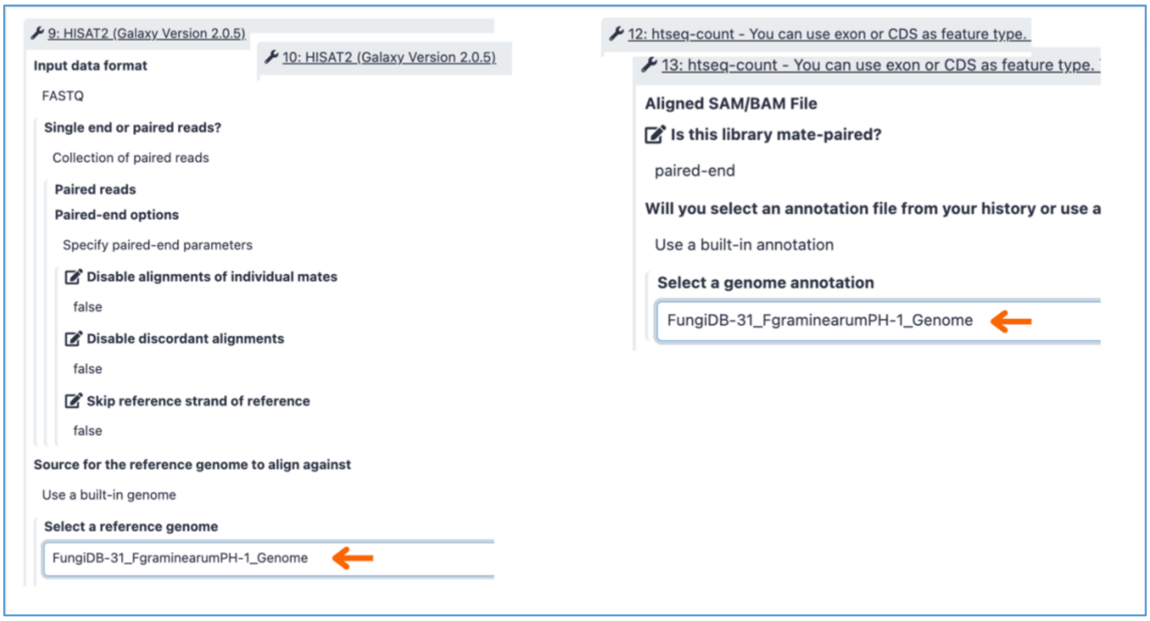
12: htseq-count - You can use exon or CDS as feature type. You must use gene_id as ID Attribute. (Galaxy Version HTSEQ: default; SAMTOOLS: 1.2; PICARD: 1.134)

13: htseq-count - You can use exon or CDS as feature type. You must use gene_id as ID Attribute. (Galaxy Version HTSEQ: default; SAMTOOLS: 1.2; PICARD: 1.134)

14: BAM to BigWig (Galaxy Version 0.2.0)

15: DESeq2_2.11.40.6 (Galaxy Version 2.11.40.6)

Make sure to set the correct reference genomes for HISAT2, htseq-count, and DESeq2 steps. It is critical that you select the correct genome that matches the experimental organism for your samples:



9: HISAT2 (Galaxy Version 2.0.5)

10: HISAT2 (Galaxy Version 2.0.5)

Input data format

FASTQ

Single end or paired reads?

Collection of paired reads

Paired reads

Paired-end options

Specify paired-end parameters

Disable alignments of individual mates

false

Disable discordant alignments

false

Skip reference strand of reference

false

Source for the reference genome to align against

Use a built-in genome

Select a reference genome

FungiDB-31_FgraminearumPH-1_Genome

12: htseq-count - You can use exon or CDS as feature type.

13: htseq-count - You can use exon or CDS as feature type.

Aligned SAM/BAM File

Is this library mate-paired?

paired-end

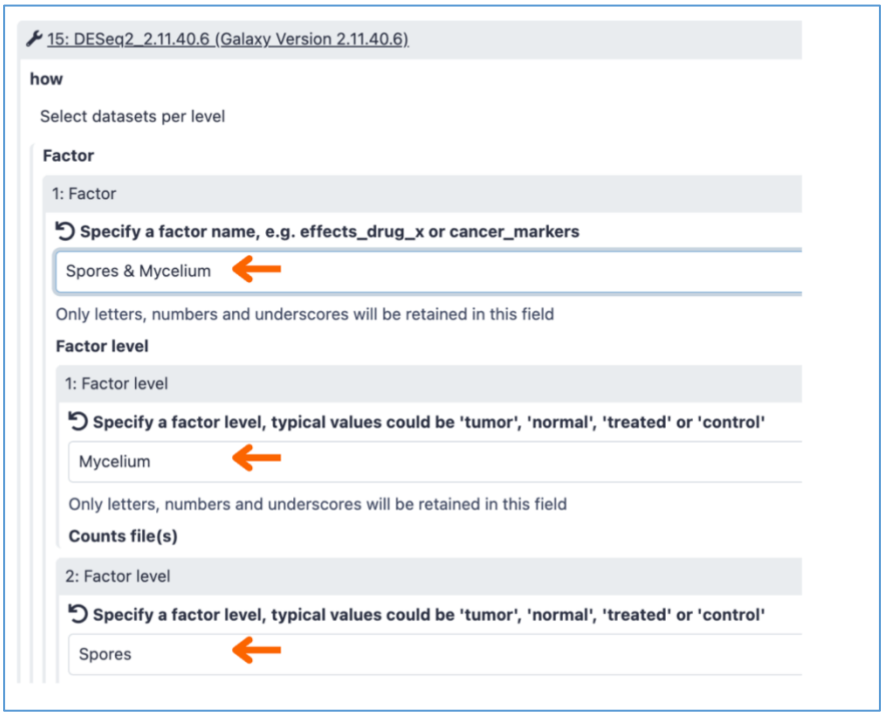
Will you select an annotation file from your history or use a

Use a built-in annotation

Select a genome annotation

FungiDB-31_FgraminearumPH-1_Genome

Name you factor levels. This helps keep everything organized and name properly in the workflow. Each factor level is typically the name of the condition, like "mycelia" or "spore".



15: DESeq2_2.11.40.6 (Galaxy Version 2.11.40.6)

how

Select datasets per level

Factor

1: Factor

Specify a factor name, e.g. effects_drug_x or cancer_markers

Spores & Mycelium

Only letters, numbers and underscores will be retained in this field

Factor level

1: Factor level

Specify a factor level, typical values could be 'tumor', 'normal', 'treated' or 'control'

Mycelium

Only letters, numbers and underscores will be retained in this field

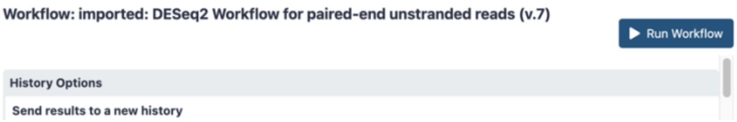
Counts file(s)

2: Factor level

Specify a factor level, typical values could be 'tumor', 'normal', 'treated' or 'control'

Spores

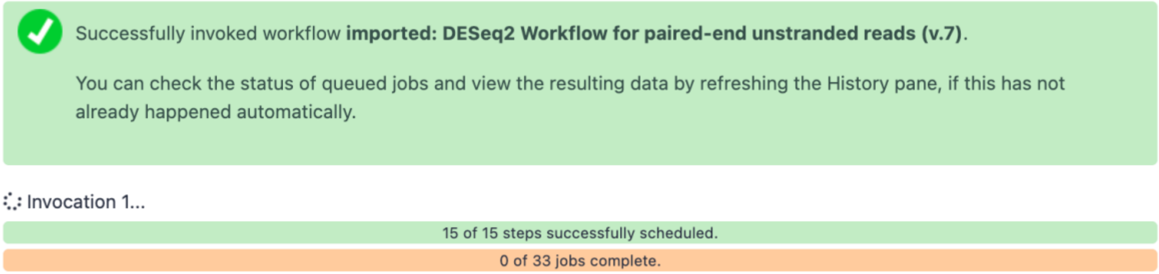
a. Once you are sure everything is configured correctly, click on "Run Workflow" at the top.



Workflow: imported: DESeq2 Workflow for paired-end unstranded reads (v.7)

History Options

Send results to a new history



Successfully invoked workflow **imported: DESeq2 Workflow for paired-end unstranded reads (v.7)**.

You can check the status of queued jobs and view the resulting data by refreshing the History pane, if this has not already happened automatically.

Invocation 1...

15 of 15 steps successfully scheduled.

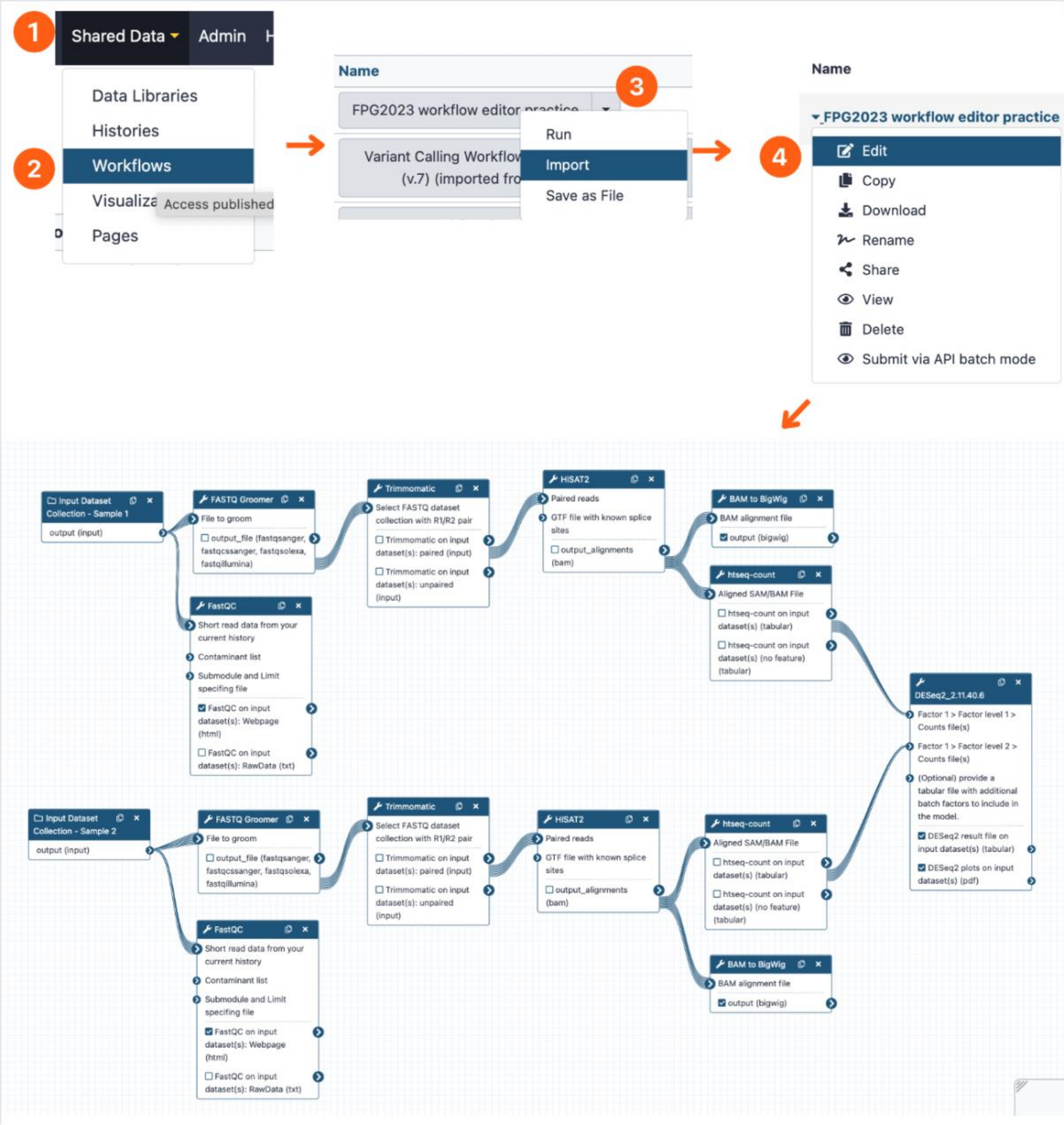
0 of 33 jobs complete.

The steps will start running in the history section on the right. Grey means they are waiting to start. Yellow means they are running. Green means they have completed. Red means there was an error in the step.

How to work with Galaxy editor (optional)

You can create your own workflows. The tools can all be added and configured in an interactive workflow editor.

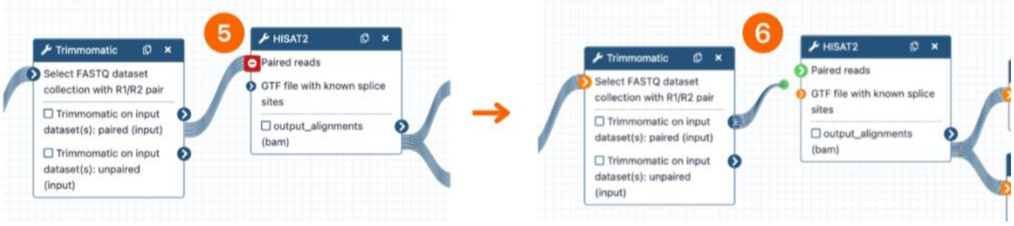
1. Navigate to the “Shared Data” menu.
2. Click on the “Workflows”.
3. Left-click on the “FPG2023 workflow editor practice” work to “import”
4. Once the workflow is important into your workspace, left-click and select “edit.



1
Shared Data
Admin
Data Libraries
Histories
2
Workflows
Visualiza
Access published
Pages
Name
FPG2023 workflow editor practice
3
Variant Calling Workflow (v.7) (imported fro
Run
Import
Save as File
4
Name
FPG2023 workflow editor practice
Edit
Copy
Download
Rename
Share
View
Delete
Submit via API batch mode
Input Dataset Collection - Sample 1
output (input)
FASTQ Groomer
File to groom
☐ output_file (fastqsanger, fastqcssanger, fastqsolexa, fastqillumina)
FastQC
Short read data from your current history
Contaminant list
Submodule and Limit specifing file
☑ FastQC on input dataset(s): Webpage (html)
☐ FastQC on input dataset(s): RawData (txt)
Trimmomatic
Select FASTQ dataset collection with R1/R2 pair
☐ Trimmomatic on input dataset(s): paired (input)
☐ Trimmomatic on input dataset(s): unpaired (input)
HISAT2
Paired reads
GTF file with known splice sites
☐ output_alignments (bam)
BAM to BigWig
BAM alignment file
☑ output (bigwig)
htseq-count
Aligned SAM/BAM File
☐ htseq-count on input dataset(s) (tabular)
☐ htseq-count on input dataset(s) (no feature) (tabular)
DESeq2_2.11.40.6
Factor 1 > Factor level 1 > Counts file(s)
Factor 1 > Factor level 2 > Counts file(s)
(Optional) provide a tabular file with additional batch factors to include in the model.
☑ DESeq2 result file on input dataset(s) (tabular)
☑ DESeq2 plots on input dataset(s) (pdf)
Input Dataset Collection - Sample 2
output (input)
FASTQ Groomer
File to groom
☐ output_file (fastqsanger, fastqcssanger, fastqsolexa, fastqillumina)
FastQC
Short read data from your current history
Contaminant list
Submodule and Limit specifing file
☑ FastQC on input dataset(s): Webpage (html)
☐ FastQC on input dataset(s): RawData (txt)
Trimmomatic
Select FASTQ dataset collection with R1/R2 pair
☐ Trimmomatic on input dataset(s): paired (input)
☐ Trimmomatic on input dataset(s): unpaired (input)
HISAT2
Paired reads
GTF file with known splice sites
☐ output_alignments (bam)
htseq-count
Aligned SAM/BAM File
☐ htseq-count on input dataset(s) (tabular)
☐ htseq-count on input dataset(s) (no feature) (tabular)
BAM to BigWig
BAM alignment file
☑ output (bigwig)

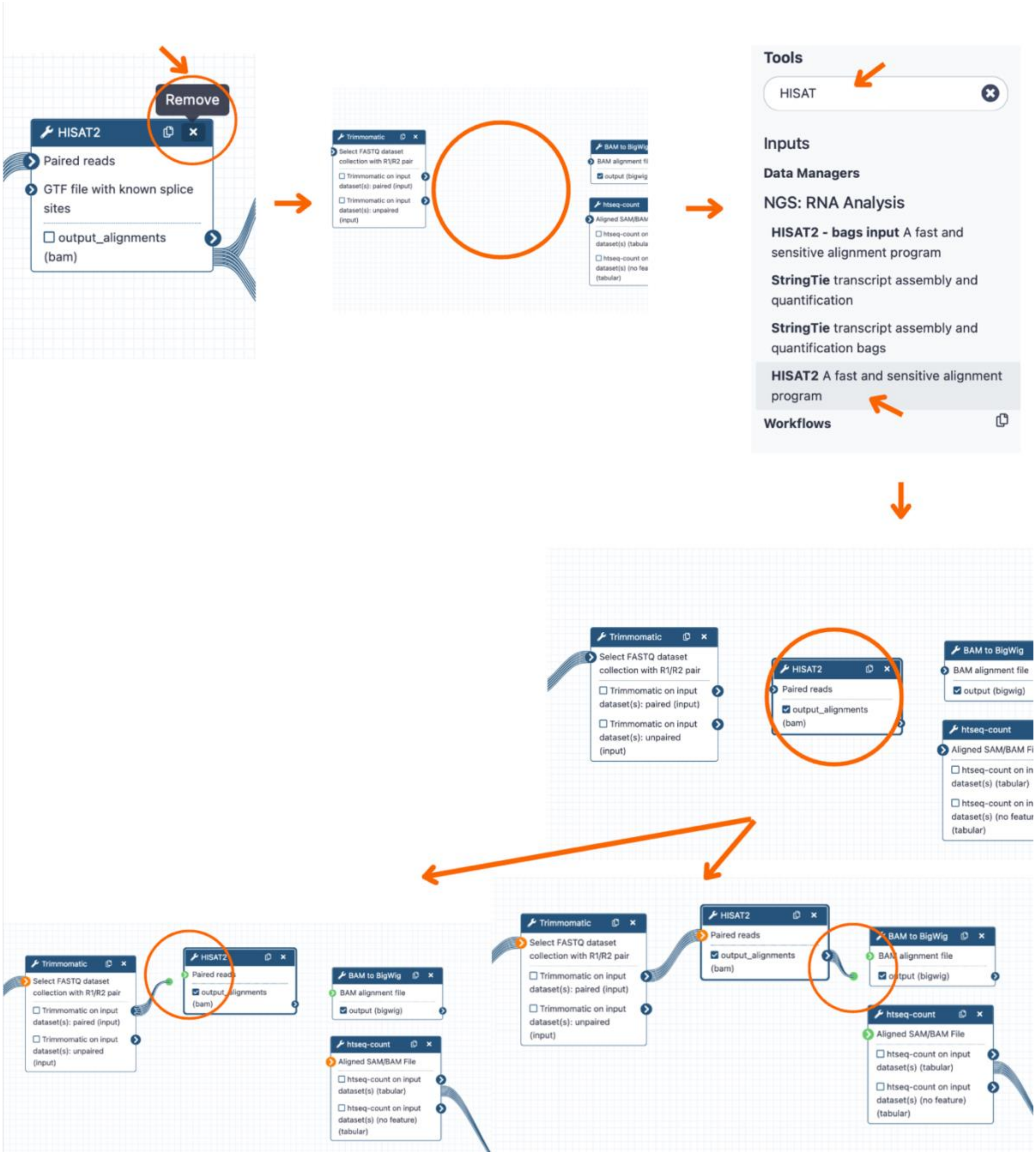
Once you are in the workflow editor:

5. Delete the Trimmomatic - HISAT2 connection.
6. Re-establish the connection by linking the “Trimmomatic on input dataset(s): paired (input) step to the “Paired reads” option in the HISTAS2.



5
Trimmomatic
Select FASTQ dataset collection with R1/R2 pair
Trimmomatic on input dataset(s): paired (input)
Trimmomatic on input dataset(s): unpaired (input)
HISAT2
Paired reads
GTF file with known splice sites
output_alignments (bam)
6

7. Delete HISAT2 step completely by clicking on the “x” in the top right corner and use the tools menu on the left to insert it back.



Remove
HISAT2
Paired reads
GTF file with known splice sites
output_alignments (bam)
Tools
HISAT
Inputs
Data Managers
NGS: RNA Analysis
HISAT2 - bags input A fast and sensitive alignment program
StringTie transcript assembly and quantification
StringTie transcript assembly and quantification bags
HISAT2 A fast and sensitive alignment program
Workflows
Trimmomatic
Select FASTQ dataset collection with R1/R2 pair
Trimmomatic on input dataset(s): paired (input)
Trimmomatic on input dataset(s): unpaired (input)
BAM to BigWig
BAM alignment file
output (bigwig)
htseq-count
Aligned SAM/BAM File
htseq-count on input dataset(s) (tabular)
htseq-count on input dataset(s) (no feature) (tabular)

Note: Sometimes you may be unable to re-establish connection. When this happens, take a look at the tool documentation notes in the right panel, check you r selection for single-read or paired-end setting in particular (paired-end setting must be selected if you are dealing with reverse and forward reads).

Now that you have learned the principals of workflow editing, you can either practice saving the workflow by clicking on the wheel at the far top corner or simply existing the workflow editor without saving.