

Hands-On Session - Introduction to Galaxy

Content

1. Uploading Data to Galaxy
2. Build a small Workflow
3. Galaxy's History system
4. Extracting Workflows from Histories

Uploading Data to Galaxy

- (A) File Upload “Standard”
- (B) File Upload “URL”
- (C) File Upload “FTP”
- (D) File Upload “Dropbox”

- Login with your user credentials -> <https://usegalaxy.eu>
- The starting point for all upload methods, as mentioned above, is called "Upload Data".

press this button

The screenshot shows the Galaxy Europe web interface. At the top, there is a navigation bar with links for Analyze Data, Workflow, Visualize, Shared Data, Help, User, and a search bar. Below the navigation bar, there is a blue header bar with the text "Galaxy Europe" and "Using 77.8 GB". The main content area has several sections:

- Tools:** A sidebar with a "search tools" input field and a "Upload Data" button, which is highlighted with a red box and a red arrow pointing to it.
- COVID-19 Research!**: A green box containing text about COVID-19 research, SARS-CoV-2 data, and training materials.
- Limited computing capacity on next April 14th and 15th, 2021**: A light blue box containing text about a shutdown due to infrastructure upgrade.
- History:** A section showing an "Unnamed history" with an "(empty)" message and a note that the history is empty and can be loaded from an external source.
- Bottom:** A quote by Prof. Stephen Hawking: "Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind in the spectrum of human understanding." – Prof. Stephen Hawking.

The screenshot shows the Galaxy Europe web interface. At the top, there's a navigation bar with links like 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'User', and a search bar. On the right, it says 'Using 77.8 GB'. Below the navigation is a sidebar with various tool categories: 'Tools', 'Get Data', 'Send Data', 'Collection C...', 'GENERAL TEXT...', 'Text Manip...', 'Filter and S...', 'Join, Subtra...', 'GENOMIC FILE...', 'Convert For...', 'FASTA/FAST...', 'FASTQ Qual...', and 'Quality Cont...'. A large central window is titled 'Download from web or upload from disk'. It has tabs for 'Regular', 'Composite', 'Collection', and 'Rule-based'. There's a 'Drop files here' area with a file icon. Below it are buttons for 'Choose local files', 'Choose remote files', 'Paste/Fetch data', 'Start', 'Pause', 'Reset', and 'Close'. A status bar at the bottom contains a quote by Prof. Stephen Hawking: '... my research, but to the research of every great and enquiring mind, is to explore the spectrum of human understanding.' — Prof. Stephen Hawking.

(A) File Upload “Standard”

Choose local file

Select a single file from the “data” folder

Start

Select the “start” button to start the upload procedure

This screenshot shows the same Galaxy Europe interface after a file has been uploaded. The 'Download from web or upload from disk' dialog box now displays a table of uploaded files. The table has columns for 'Name', 'Size', 'Type', 'Genome', 'Settings', and 'Status'. One file, 'med4_set1.fastq', is listed with a size of '644 b', type 'Auto-de...', genome 'Additional ...', settings icon, and a status of '0%'. Below the table are buttons for 'Choose local files', 'Choose remote files', 'Paste/Fetch data', 'Start', 'Pause', 'Reset', and 'Close'. The status bar at the bottom still contains the quote by Prof. Stephen Hawking.

(B) File Upload “URL”

Paste/Fetch data

Click on the “Paste/Fetch data” button

Copy and Paste

Select the URL (mentioned below) and paste the string into the box

<https://zenodo.org/api/files/c07c0fb...DRR000770.fastqsanger.gz>

Start

Select the “start” button to start the upload-procedure

The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with various tools categorized under 'GENERAL TEXT' and 'GENOMIC FILE'. The main area is titled 'Download from web or upload from disk' and shows a table of files being processed:

Name	Size	Type	Genome	Settings	Status
med4_set1.fastq	644 b	Auto-de...	----- Additional ...	⚙️	100% ✓
New File	90 b	Auto-de...	----- Additional ...	⚙️	100% ✓

Below the table, there's a text input field containing the URL: <https://zenodo.org/api/files/c07c0fb...DRR000770.fastqsanger.gz>. At the bottom of the interface, there's a toolbar with buttons for 'Choose local files', 'Choose remote files', 'Paste/Fetch data' (which has a red arrow pointing to it), 'Start', 'Pause', 'Reset', and 'Close'.

(C) File Upload “FTP”

Download a FTP-Client and install it on your local machine
<https://filezilla-project.org/download.php?type=client>

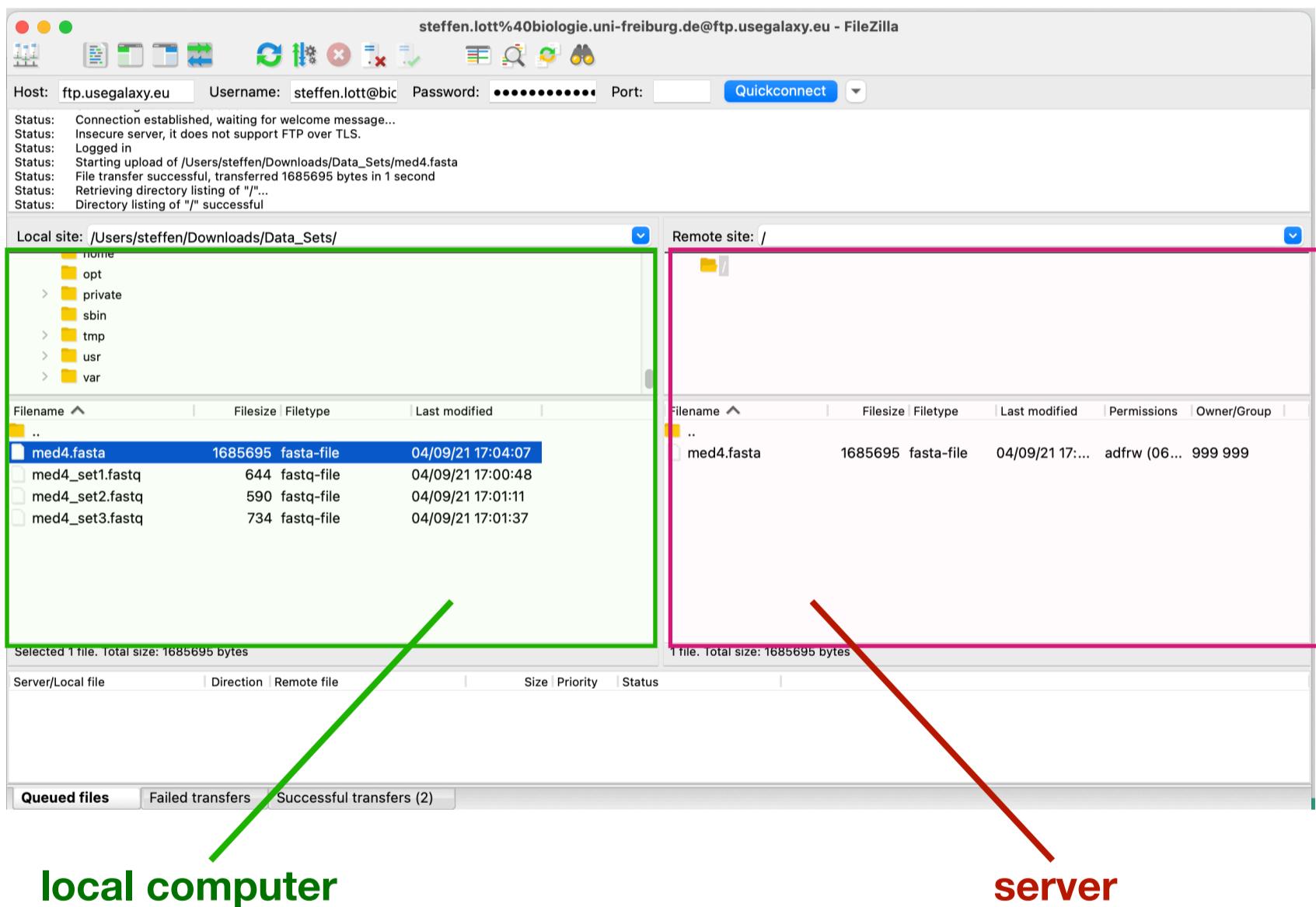
Start FileZilla

Host: ftp.usegalaxy.eu

Username: E-Mail (Galaxy user name)

Password: Password (Galaxy user password)

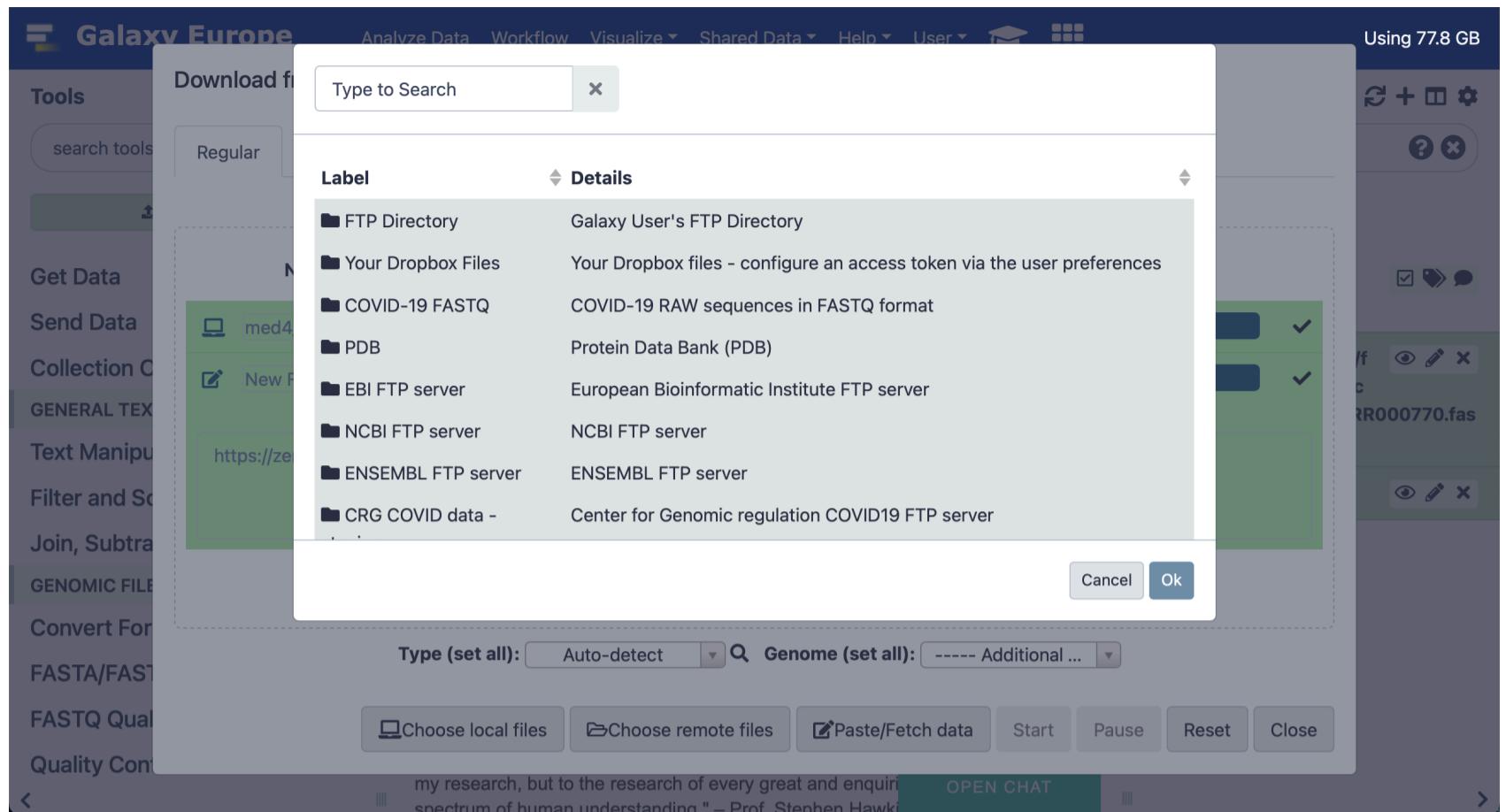
Press “Quickconnect”



File transfer

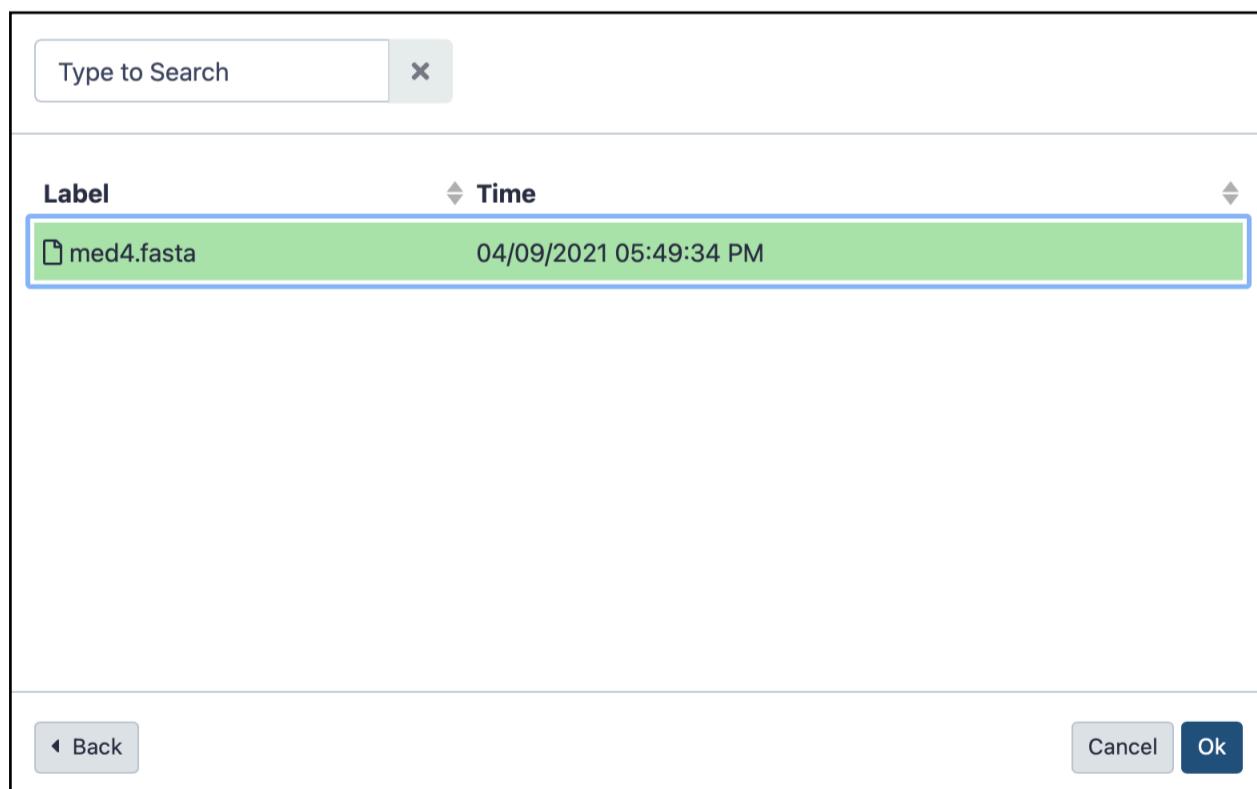
Drag-and-Drop a file from your local computer to the server
(e.g. med4.fasta)

Switch back to “[usegalaxy.eu](#)”



Choose remote files

Select “Choose remote files” and “FTP Directory” from the new dialog. Now, you can see your previously uploaded data. Before you can press “Ok,” you have to select the data first.



Final step

Finally, you have to press “start” to initiate the integration process into your history.

(D) File Upload “Dropbox”

Precondition

You should have a dropbox account for performing this type of upload method!

Dropbox

First of all, access your dropbox account and select the “grid” field.



- Now, select “App Center” and a new window appears.
- Scroll down until you see “Build an app”, and select those button
- To load the needed form, click on “Create apps”.
- Fill out the form as shown below and press “Create app”

Create a new app on the DBX Platform

1. Choose an API

A screenshot of the "Choose an API" step in the Dropbox App Center. It shows a "Scoped access" section with a "New" button and a note about selecting access level. A red arrow points to the "Learn more" link. On the right, there is a 3D cube icon representing the API.

2. Choose the type of access you need

[Learn more about access types](#)

A screenshot of the "Choose the type of access you need" step. It shows two options: "App folder" and "Full Dropbox". The "Full Dropbox" option is selected and highlighted with a red arrow. A second red arrow points to the "Learn more about access types" link.

3. Name your app

A screenshot of the "Name your app" step. A red arrow points to the input field where "GalaxyConnector" is typed.

A screenshot of the "Agree to Terms and Conditions" step. A red arrow points to the checkbox labeled "I agree to Dropbox API Terms and Conditions".

[Create app](#)

Permissions

Switch to the “Permissions” tab, select all options as shown below and press “Submit”.

GalaxyConnector

Individual Scopes Individual scopes include the ability to view and manage a user's files and folders. [View Documentation](#)

Account Info
Permissions that allow your app to view and manage Dropbox account info

account_info.write View and edit basic information about your Dropbox account such as your profile photo

account_info.read View basic information about your Dropbox account such as your username, email, and country

Files and folders
Permissions that allow your app to view and manage files and folders

files.metadata.write View and edit information about your Dropbox files and folders

files.metadata.read View information about your Dropbox files and folders

files.content.write Edit content of your Dropbox files and folders

files.content.read View content of your Dropbox files and folders

Collaboration
Permissions that allow your app to view and manage sharing and collaboration settings

sharing.write View and manage your Dropbox sharing settings and collaborators

sharing.read View your Dropbox sharing settings and collaborators

Click Submit when you are done making changes. (Existing access tokens will not be affected)

Settings

- Switch back to the “Settings” tab

Note - Later, we need this token to connect Galaxy with your Dropbox account!

Note - Do not share your token with anyone!

Allow public clients (Implicit Grant & PKCE) ⓘ

Generated access token ⓘ

Access token expiration ⓘ

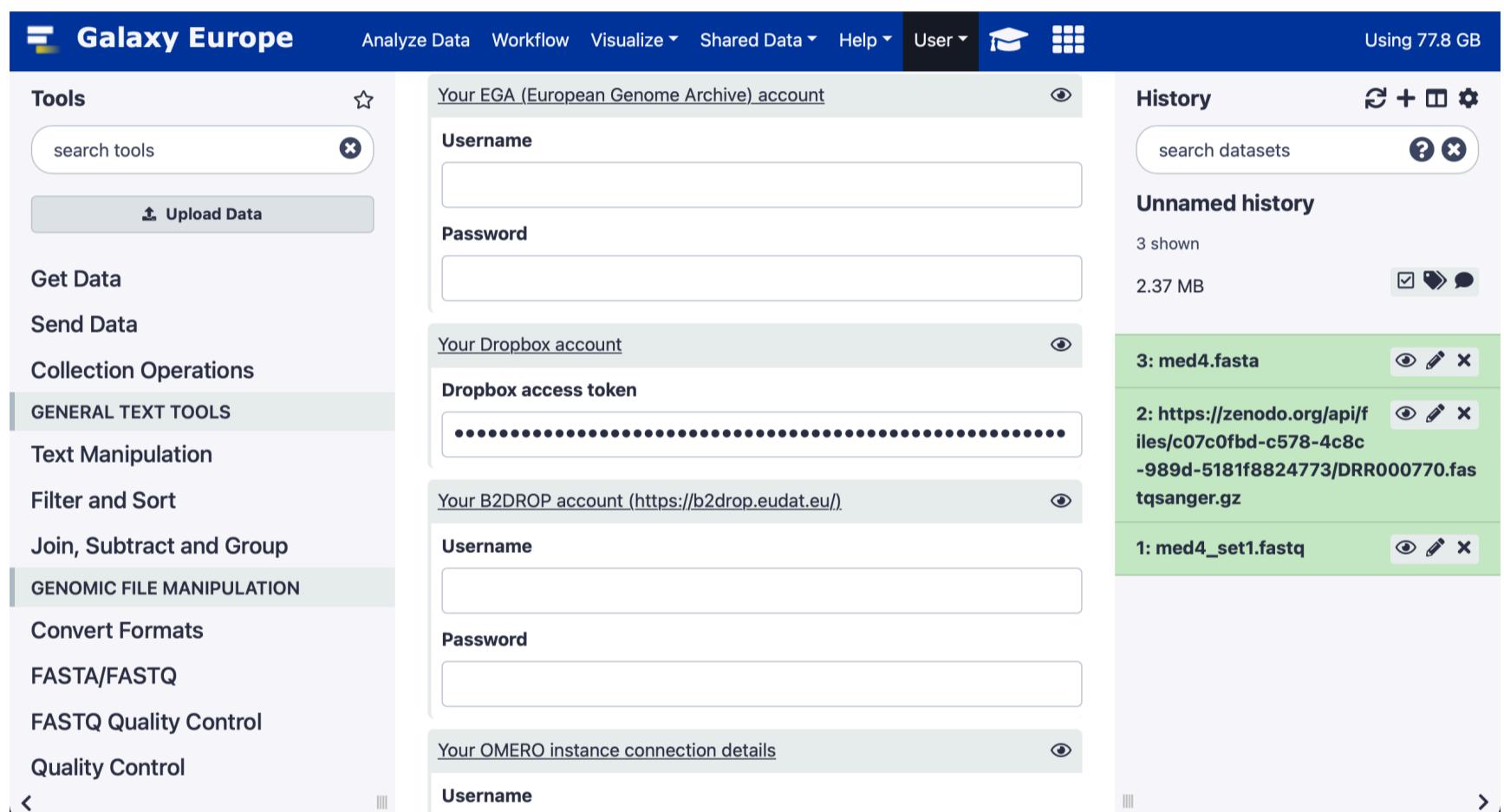
Short-lived

No expiration

- Change the “Access token expiration” from “Short-lived” to “No expiration”
- Now, click the button “Generate”

Galaxy

- Go back to "[usegalaxy.eu](#)" and after you have logged in with your user data, select "User" from the menu bar.
- First select "Preferences", followed by clicking on "Manage Information"
- Search for the field "Dropbox access token" and copy the previously created token into the field.
- Scroll down and press "Save" 



The screenshot shows the Galaxy Europe web interface. The top navigation bar includes links for Analyze Data, Workflow, Visualize, Shared Data, Help, User (selected), and a search bar. The main content area is titled "Manage Information". It contains several sections for connecting to external services:

- Your EGA (European Genome Archive) account:** Fields for Username and Password.
- Your Dropbox account:** A field for Dropbox access token containing a long string of characters.
- Your B2DROP account (<https://b2drop.eudat.eu/>):** Fields for Username and Password.
- Your OMERO instance connection details:** A field for Username.

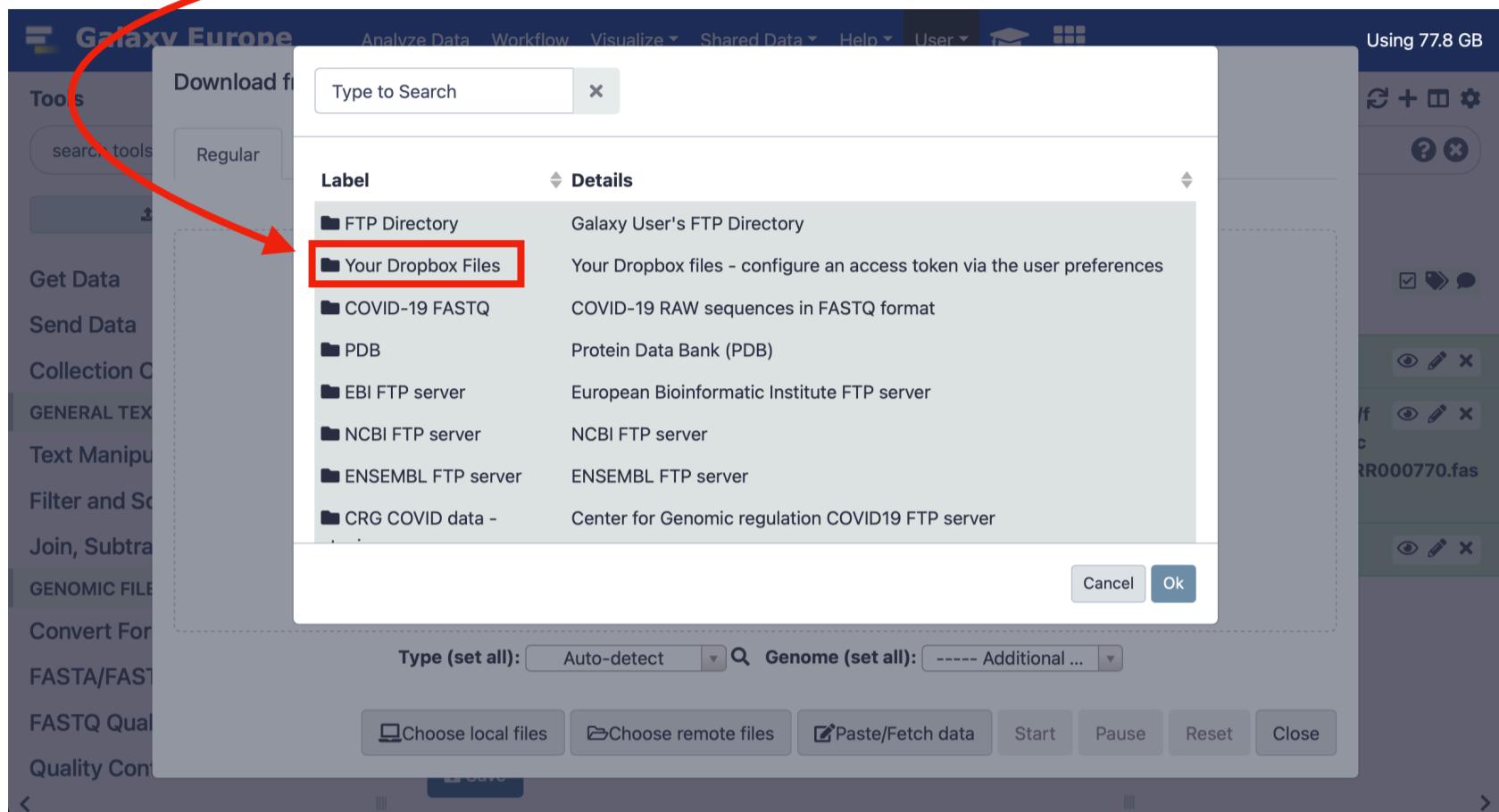
The right sidebar displays a history of datasets with three items listed:

- 3: med4.fasta
- 2: <https://zenodo.org/api/files/c07c0fbcd-c578-4c8c-989d-5181f8824773/DRR000770.fas> tqsanger.gz
- 1: med4_set1.fastq

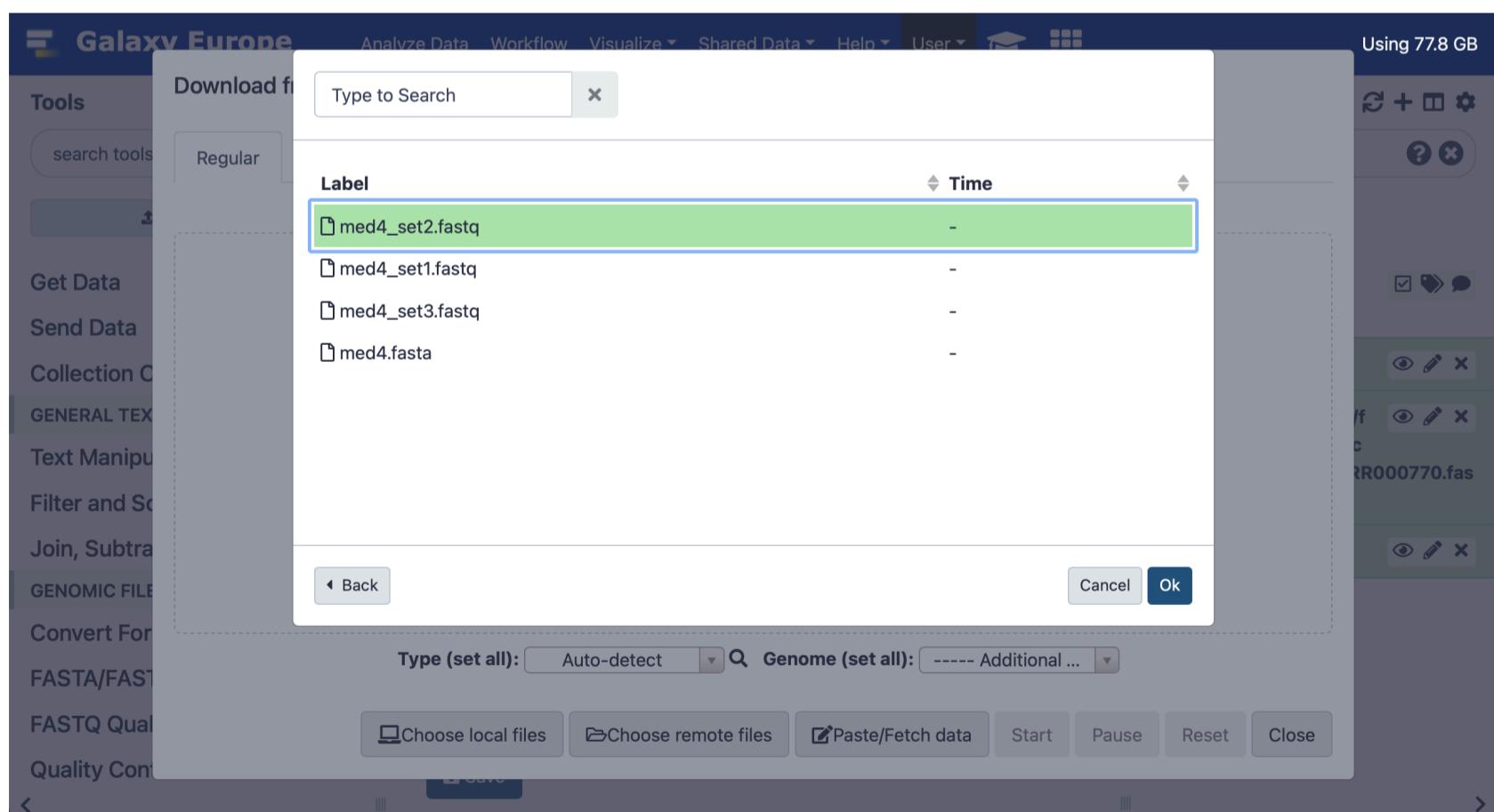
Congratulation, you have connected your Dropbox account with Galaxy.

Galaxy - Get data from your Dropbox account

- Press “Upload Data” and select “Choose remote files”
- Select “Your Dropbox Files”



- Choose the files to upload and press “Ok”



- Finally, press “Start” to complete the process.

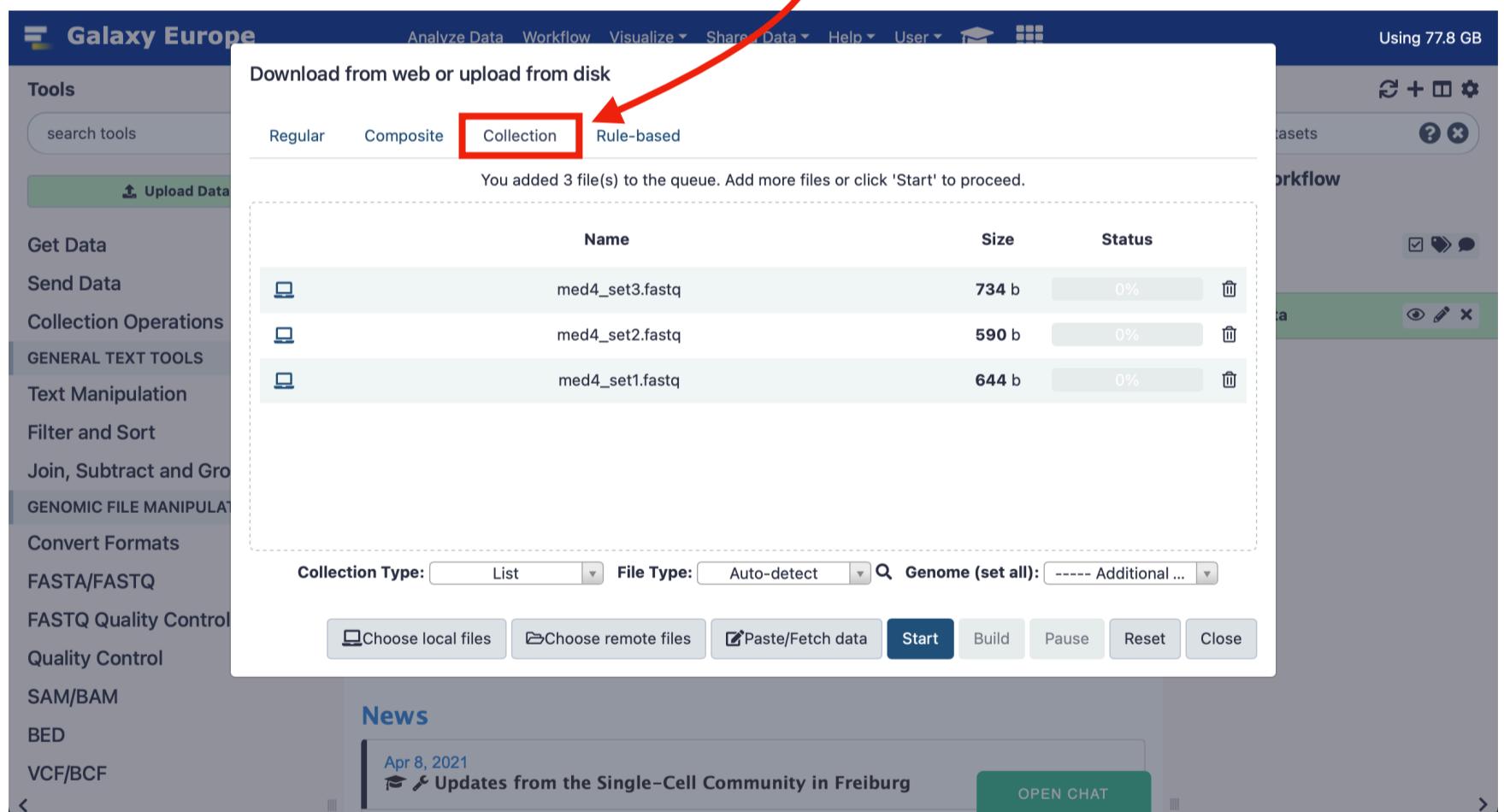
Build a small workflow

Note: For all tools we will use the default settings!

Note: First, create a new history and give a name like “My first workflow”

File Upload

- First, upload the reference genome “med4.fasta”
- Second, upload all three datasets (med4_set1.fastq, ..., med4_set3.fastq) as a collection



- Third, press “Start” to initiate the uploading procedure
- Fourth, press “Build” and name your new collection e.g., “RNA-Seq data”.
- Finally, press “Create list”

FastQC

- The first tool we want to use is called “FastQC”
- Enter the name (“FastQC”) in the search bar and select the tool

The screenshot shows the Galaxy Europe web interface. In the top left, the 'Tools' sidebar is open, showing various bioinformatics tools. The 'FastQC' tool is highlighted with a red box and an arrow pointing to its entry in the search results. The main panel displays the 'FastQC Read Quality reports' tool configuration. It includes fields for 'Short read data' (set to '5: RNA-Seq data'), 'Contaminant list' (set to 'Nothing selected'), 'Adapter list' (set to 'Nothing selected'), and 'Submodule and Limit specifying file' (set to 'Nothing selected'). The right side of the interface shows the user's history, which contains a workflow named 'My first workflow' and a dataset named '5: RNA-Seq data'.

- Before we can apply this algorithm to our data, we should select as input type “collection”
- Galaxy automatically recognizes the matching record from your history (RNA-Seq data)
- Scroll down and press **Execute**

Bowtie2

- The second tool we want to use is called “Bowtie2”
- Enter the name (“Bowtie2”) in the search bar and select the tool

The screenshot shows the Galaxy Europe interface. In the top left, the 'Tools' sidebar has 'Bowtie2' selected. The main workspace shows the 'Bowtie2 - map reads against reference genome' tool configuration. The 'Is this single or paired library?' dropdown is set to 'Single-end'. The 'FASTA/Q file' input field contains '1: med4.fasta'. In the 'History' panel on the right, '1: med4.fasta' is listed under 'My first workflow'. The 'Select reference genome' dropdown also shows '1: med4.fasta'. At the bottom right, there is an 'Execute' button.

- Before we can apply this algorithm to our data, we should select as input type “collection”
- Galaxy automatically recognizes the matching record from your history (RNA-Seq data)
- Select “Use a genome from the history and build index”
- Select our reference genome “med4.fasta”
- Scroll down and press **Execute**

Congratulation, you built your first workflow. Use this history for the following two tutorials.

Understanding Galaxy history system

When data is uploaded from your computer or analysis is done on existing data using Galaxy, each output from those steps generates a dataset. These datasets (and the output datasets from later analysis on them) are stored by Galaxy in **Histories**.

The Current History

All users have one 'current' history, which can be thought of as **a workspace** or **a current working directory** in bioinformatics terms. Your current history is displayed in the right hand side of the main 'Analyze Data' Galaxy page in what is called the history panel.

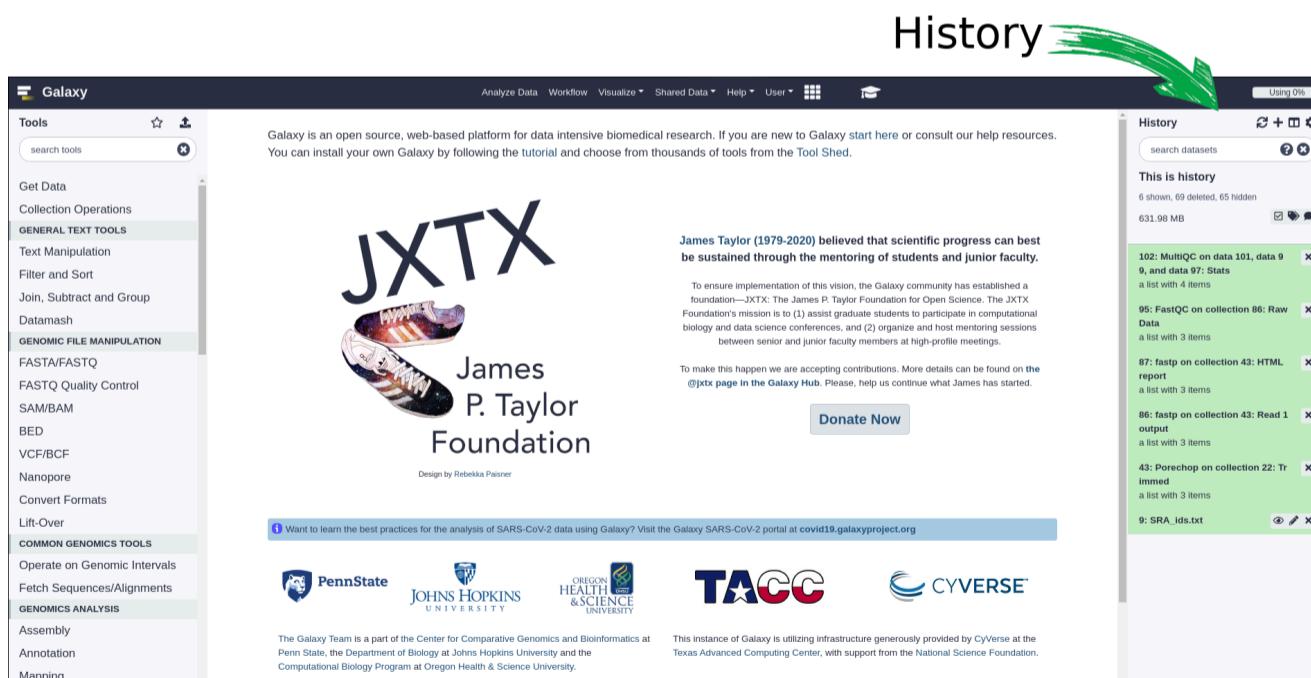


Figure 1: Galaxy History is simply the right panel of the interface

The history panel displays output datasets in the order in

which they were created, with the oldest/first shown at the bottom. As new analyses are done and new output datasets are generated, the newest datasets are added to the top of the history panel. In this way, the history panel displays the history of your **analysis over time**.

Users that have registered an account and logged in can have many histories and the history panel allows switching between them and creating new ones. This can be useful to organize different analyses.

Anonymous users (if your Galaxy allows them) are users that have not registered an account. Anonymous users are only allowed one history. On our main, public Galaxy server, users are encouraged to register and log in with the benefit that they can work on many histories and switch between them.

The histories of anonymous users are only associated through your browser's session. **If you close the browser or clear your sessions - that history will be lost!** We can not recover it for you if it is.

Current history controls



Figure 2: History Controls

Above the current history panel are three buttons: the refresh, history options, and 'view all histories' button.

- The 'refresh' button will entirely reload the history being viewed. This can be helpful if you believe the history interface needs to be updated or isn't updating properly.
- The 'create new history' button will create an empty history.
- The 'view all histories' button sends you to the interface for managing multiple histories.
- The 'history options' button opens the history options menu which allows you to perform history-related tasks.

History Information

Histories also store information in addition to the datasets they contain. They can be named/re-named, tagged, and annotated.

Renaming a history

All histories begin with the name 'Unnamed history'. Non-anonymous users can rename the history as they see fit:

1. Click the existing name. A text input field will appear with the current name.
2. Enter a new name or edit the existing one.

3. Press `Enter` to save the new name. The input field will disappear and the new history name will display.
4. To cancel renaming, press `Esc` or click outside the input field.

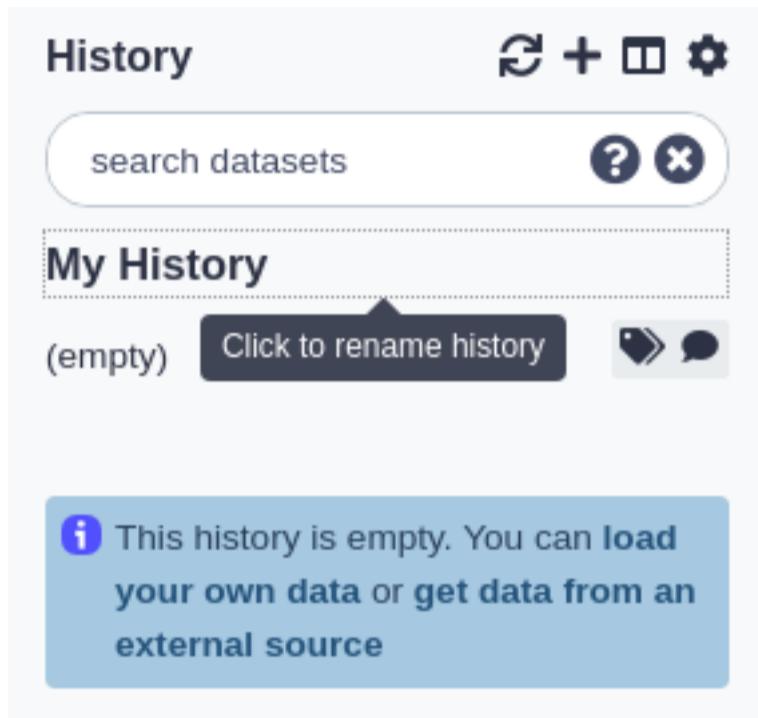


Figure 3: Renaming a history by clicking on its name and entering a new one

Tagging a history

Tags are short pieces of text used to describe the thing they're attached to and many things in Galaxy can be tagged. Each item can have many tags and you can add new tags or remove them at any time. Tags can be another useful way to organize and search your data. For instance, you might tag a history with the type of analysis you did in it: 'assembly' or 'variants'. Or you may tag them according to data sources or some other metadata: 'long-term-care-facility' or 'yellowstone park:2014'.

Note: It is recommended to replace spaces in tags with _

or -. Although spaces are allowed in tags for histories, they are removed from the tags for datasets.

To tag a history:

1. Click the tag button at the top of the history panel. An input field showing existing tags (if any) will appear.
2. Begin typing your new tag in the field. Any tags that you've used previously will show below your partial entry - allowing you to use this 'autocomplete' data to re-use your previous tags without typing them in full.
3. Press enter or select one of the previous tags with your arrow keys or mouse.
4. To remove an existing tag, click the small 'X' on the tag or use the backspace key while in the input field.

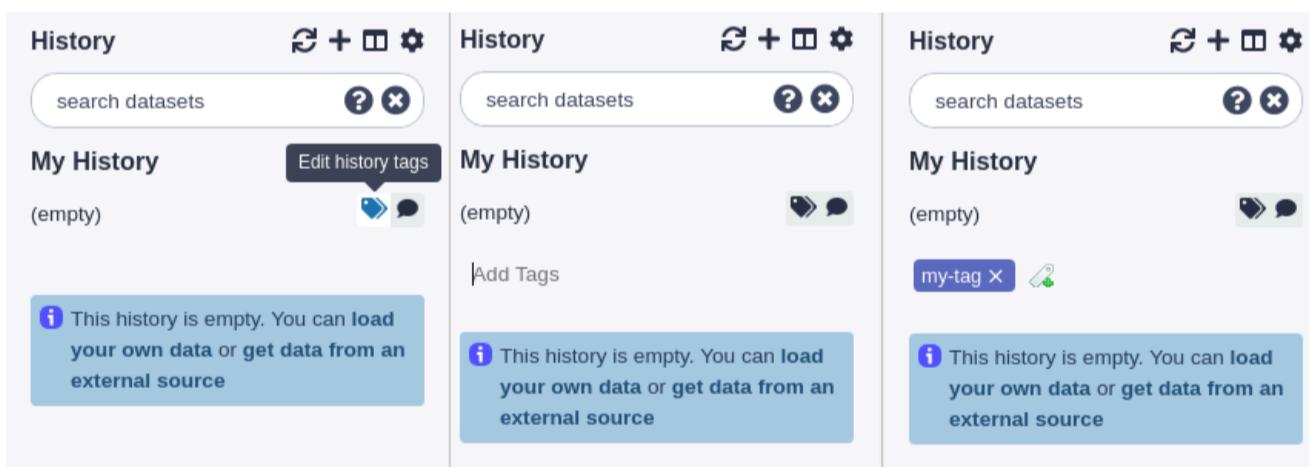


Figure 4: Tagging a history will help searching for it later on.

Annotating a history

Sometimes tags and names are not enough to describe the work done within a history. Galaxy allows you to create history annotations: longer text entries that allow for more

formatting options. The formatting of the text is preserved. Later, if you publish or share the history, the annotation will be displayed automatically - allowing you to share additional notes about the analysis.

To annotate a history:

1. Click the annotation button at the top of the history panel. A larger text section will appear displaying any existing annotation (or, if there's none, italic text saying you can click on the control to create an annotation).
2. Click the annotation section. A larger input field will appear.
3. Add your annotations. Enter will move the cursor to the next line. (Tabs cannot be entered since the 'Tab' button is used to switch between controls on the page - tabs can be pasted in however).
4. To save the annotation, click the 'Done' button.

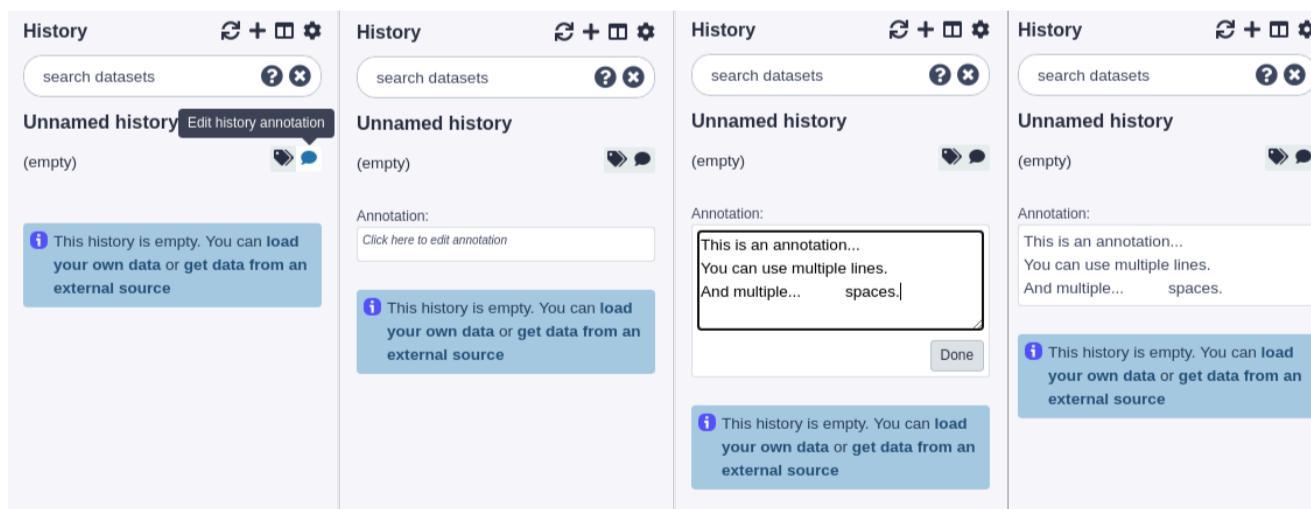


Figure 5: Annotating a history allows entering more information such as, for example, experimental details related to the analysis

History size

As datasets are added to a history, Galaxy will store them on the server. The total size of these files, for all the datasets in a history, is displayed underneath the history name. For example, if a history has 200 megabytes of dataset data on Galaxy's filesystem, '200 MB' will be displayed underneath the history name.

If your Galaxy server uses quotas, the total combined size of all your histories will be compared to your quota. If you're using more than the quota allows, Galaxy will prevent you from running any new jobs until you've deleted some datasets and brought that total below the quota.

History Panel Datasets

Datasets in the history panel show the state of the job that has generated or will generate the data.

There are several different 'states' a dataset can be in:

1. When you first upload a file or run a tool, the dataset will be in the **queued** state. This indicates that the job that will create this dataset has not yet started and is in line to begin.
2. When the job starts, the dataset will be in the **running** state. The job that created these datasets is now running on Galaxy's cluster.

3. When the job has completed successfully, the datasets it generated will be in the **ok** state.
4. If there's been an error while running the tool, the datasets will be in the **error** state.
5. If a previously running or queued job has been paused by Galaxy, the dataset will be in the **paused** state. You can re-start/resume paused jobs using the options menu above the history panel and selecting 'Resume Paused Jobs'.

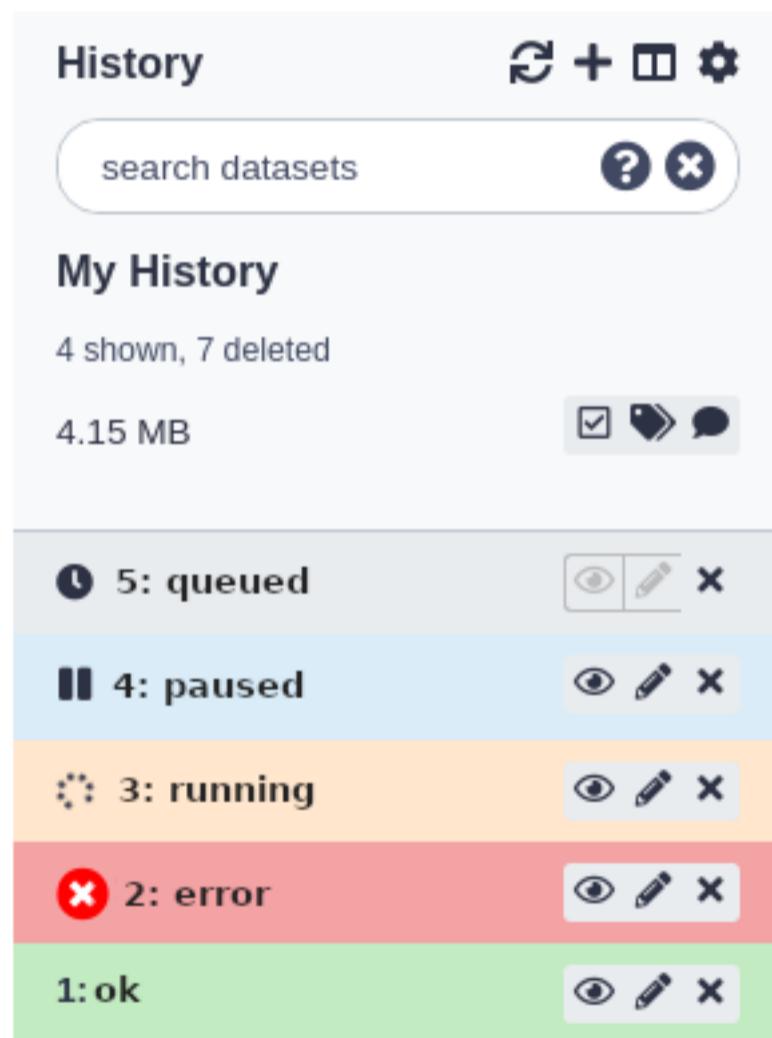


Figure 6: Dataset states

Datasets in the panel are initially shown in a 'summary' view, that only displays:

1. A **number** indicating in what order (or what step) this

- dataset was created,
2. The dataset **name**.
 3. A **view** button: click this to view the dataset contents in raw format in the browser.
 4. An **edit** button: click this to edit dataset properties.
 5. A **delete** button: click this to delete the dataset from the history (*don't worry*, you can undo this action).

Note: some of the buttons above may be disabled if the dataset is in a state that doesn't allow the action. For example, the 'edit' button is disabled for datasets that are still queued or running.

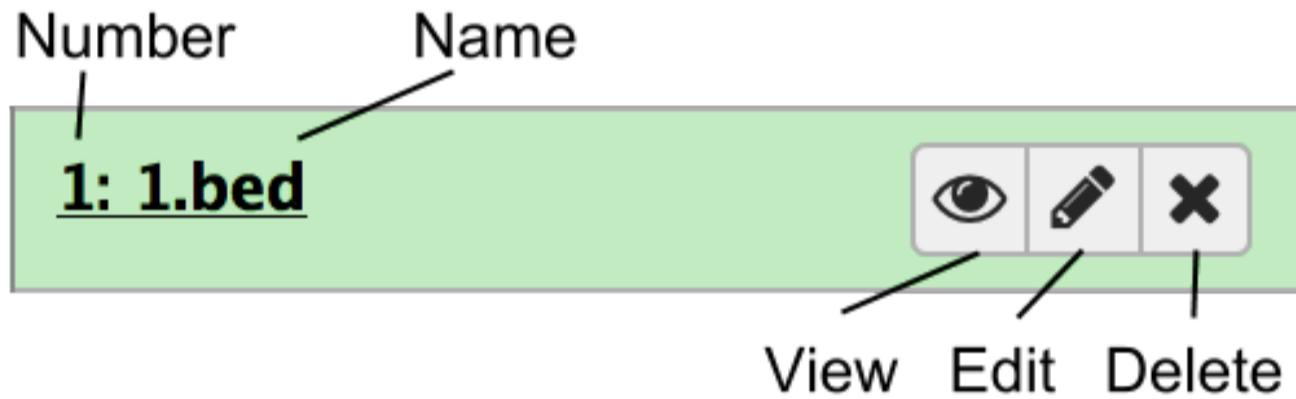


Figure 7: Controls for 'summary' (or collapsed) dataset view

You can click the dataset name and the view will expand to show more details:

1. A short description of the data.
2. The file **format**.
3. The reference sequence (or **database**) for the data.
4. (Optionally) some information/output from the job that produced this dataset.
5. A row of buttons that allow further actions on the

- dataset.
6. A **peek** of the data: a couple of rows of data with the column headers (if available).

The screenshot shows the expanded view of a dataset named "1: 1.bed". The top bar includes an "eye" icon, a pencil icon, and a delete icon. Below the title, there are sections for "Description" (65 regions), "File Format" (format: bed, database: ?), "More Info" (uploaded bed file), "Actions" (with icons for save, info, refresh, and chart), and "External Display Apps" (display in IGB View). A "Data peek" section is highlighted with a green border, showing a table with columns: 1.Chrom, 2.Start, 3.End, and 4.Name. The data rows are:

1.Chrom	2.Start	3.End	4.Name
chr1	147962192	147962580	CCDS989.1
chr1	147984545	147984630	CCDS990.1
chr1	148078400	148078582	CCDS993.1
chr1	148185136	148185276	CCDS996.1
chr10	55251623	55253124	CCDS7248.1
chr11	116124407	116124501	CCDS8374.1

Figure 8: Controls for expanded dataset view.

Note: many of these details are only displayed if the dataset has finished running, is in the 'ok' state, and is not deleted. Otherwise, you may only see a shorter message describing the dataset's state (e.g. 'this dataset is waiting to run')

Managing Datasets Individually

Hiding and unhiding datasets

Some procedures in Galaxy such as workflows will often

hide history datasets in order to simplify the history and hide intermediate steps of an automated analysis. These hidden datasets won't normally appear in the history panel but they're still mentioned in the history subtitle (the smaller, grey text that appears below the history name). If your history has hidden datasets, the number will appear there (e.g. '3 hidden') as a clickable link. If you click this link, the hidden datasets are shown. Each hidden dataset has a link in the top of the summary view that allows you to unhide it. You can click that link again (which will now be 'hide hidden') to make them not shown again.

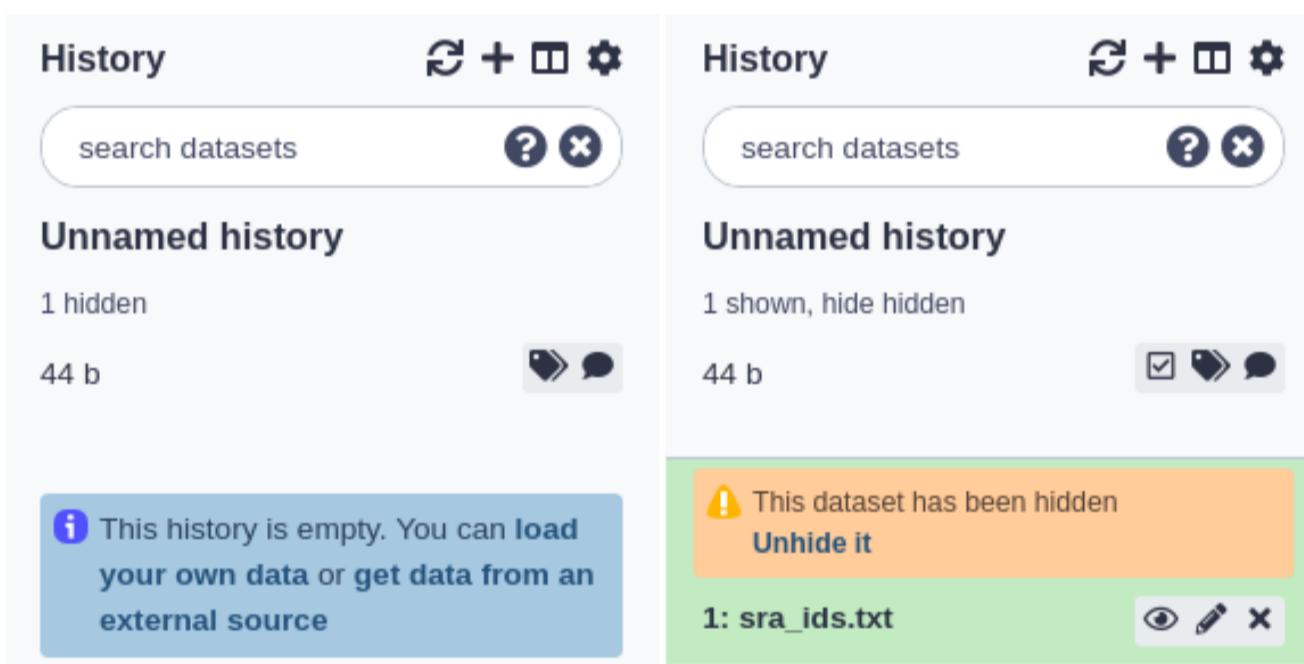


Figure 9: Hiding and unhiding datasets. Left side shows a history with one hidden dataset. We know this because a link "1 hidden" appears under the history's name. Clicking this link will reveal the hidden dataset as shown on the right side of the figure.

Deleting and undeleting datasets

You can **delete** any dataset in your history by clicking the delete button. This does not immediately remove the dataset's data from Galaxy and **it is reversible**. When you

delete a dataset from the history, it will be removed from the panel but (like hidden datasets) the total number of deleted datasets is shown in the history subtitle as a link. Clicking this link (e.g. '3 deleted') will make the deleted datasets visible and each deleted dataset will have a link for manually undeleting it, above its title. You can click that link again (which will now be 'hide deleted') to make them not shown again.

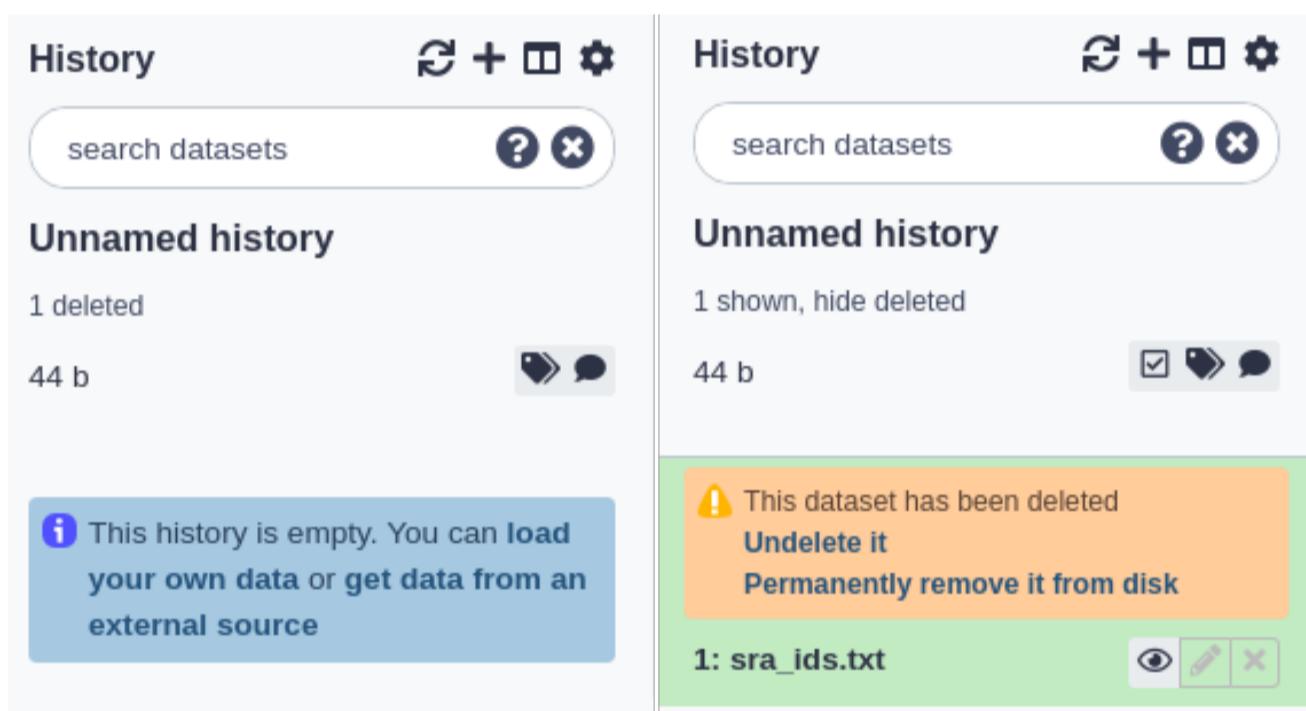


Figure 10: Deleting and undeleting datasets. The left side shows a history with one deleted dataset. We know this because a link '1 deleted' appears under the history's name. Clicking this link will reveal the deleted dataset as shown on the right side of the figure. From here it can be undeleted or deleted permanently.

Purging datasets and removing them permanently from Galaxy

If you are showing deleted datasets and *your Galaxy allows users to purge datasets*, you will see an additional link in the top of each deleted dataset titled '**Permanently remove it from disk**'. Clicking this will remove the file that

contains that dataset's data and will decrease the disk space used by the history. **This action is not reversible and cannot be undone.**

If your Galaxy doesn't allow users to purge their datasets, you will not see that link.

Admins may purge your deleted datasets

Depending on the policy of your Galaxy server, administrators will often run scripts that search for and purge the datasets you've marked as deleted. Often, deleted datasets and histories are purged based on the age of the deletion (e.g. datasets that have been marked as deleted for 90 days or more). Check with the administrators of your Galaxy instance to find out the policy used.

Tagging datasets

There are two types of tags that can be used as an additional level of labeling for datasets: **standard tags** and **hashtags** (also known as **name tags** or **propagating tags**). The standard tags work similarly to history tags described above - they add another level of description to datasets making them easier to find:

A.

The screenshots show a dataset viewer interface. In the first panel, a fasta file named 'patient_2.fasta' is displayed with 8,000 sequences in FASTA format. In the second panel, a tag 'mic' is added to the dataset, and the word 'microbiome' appears in a blue-highlighted box below the sequence. In the third panel, the tag 'microbiome' is confirmed and displayed prominently.

B.

The screenshots show a history panel. In the first panel, a search bar contains 'microbiome'. The second panel shows the search results, listing datasets 11: patient_8.fasta, 10: patient_7.fasta, 9: patient_6.fasta, 8: patient_5.fasta, 7: patient_4.fasta, 6: patient_3.fasta, 5: patient_2.fasta, and 4: patient_1.fasta. In the third panel, the tag 'microbiome' is highlighted in blue across all listed datasets.

Figure 11: Standard tags provide an additional level of annotation for individual datasets. A. Tags are added by clicking on the tags icon and entering a name. B. Here the tag is used to search the history. Entering 'microbiome' in the search box and pressing `Enter` shows the only dataset containing that tag.

Hashtags are much more powerful as they are **displayed** in the history panel and **propagate** through the analysis:

A.

9: patient_6.fasta

8,000 sequences
format: **fasta**, database: ?

uploaded fasta file

display with IGV local

>pir||R5NT28 ribosomal protein L28 precurs
MATMVAGISLRGPVMSSHRTFSVTKRASLPQSKLSSELSFV
QPVARRICPFTGKSNRANKVSHSNHKTKLQFVNQYKRIW
AKKAGIDLSKK

microbiome

9: patient_6.fasta

8,000 sequences
format: **fasta**, database: ?

uploaded fasta file

display with IGV local

>pir||R5NT28 ribosomal protein L28 precurs
MATMVAGISLRGPVMSSHRTFSVTKRASLPQSKLSSELSFV
QPVARRICPFTGKSNRANKVSHSNHKTKLQFVNQYKRIW
AKKAGIDLSKK

microbiome

B.

History

search datasets

Disease analysis

8 shown, 8 deleted

37.29 MB

16: Fasta Statistics on dat
a 15: Fasta summary stats
microbiome

15: NormalizeFasta on dat
a 9: Normalized FASTA dat
asset
microbiome

13: Fasta Statistics on dat
a 8: Fasta summary stats
disease

11: patient_8.fasta

10: patient_7.fasta

9: patient_6.fasta
microbiome

8: patient_5.fasta
disease

7: patient_4.fasta

Figure 12: Hashtags allow you to more easily track datasets through the analysis. A. Hashtags are added similarly to standard tags but with one important difference: they are prepended with a hash '#' symbol. B. Here you see a history where four starting datasets were given name tags. As the analysis progresses hashtags stay with all datasets that are derived from the initial ones. For example, you can easily see which of the `bwa` and `MarkDuplicates` outputs are derived from, say, mother data.

Extracting Workflows from Histories

This practical shows how to create a reusable analysis pipeline, called a *workflow* in Galaxy, from an analysis that you have already run in Galaxy, called a *history*.

Agenda

In this tutorial, we will cover:

1. Pretreatments
 1. Motivation
 2. Set your current history
 3. Extract the recipe from your history
 4. Final thoughts

Pretreatments

This tutorial is a good second step after running your first analysis on Galaxy.

Motivation

Galaxy excels at helping researchers figure out how to do analyses. It's easy to try different tools and settings, and if your first result isn't quite what you want, you can just keep trying until you get what you want. What you may end up with is a slightly (or very) messy analysis *history*

under your account. If this is the only time you'll run that particular analysis then it's fine to just hang on to that history. It might be handy to look at the next time you want to do something in Galaxy.

But what if you want to run that analysis again, maybe on updated datasets, or maybe even on entirely different datasets? You could manually rerun all the steps in your first analysis, using the new datasets instead of the old. But that is tedious, and error-prone. What you want is the ability to easily run exactly the same tools, with the same settings, in the same order as the first time you ran the analysis.

Galaxy *workflows* enable this, and this tutorial shows how you can create one from the analysis you have already done, and then run the analysis exactly as you did before, but on the new datasets.

Tip: Confused about *Histories* and *Workflows*?

Set your current history

By this time, you may have multiple histories under your Galaxy account. You'll need to make the history that you want create a recipe/workflow for be your *current* history. Your current history is the one shown in the History panel on the right.

Hands-on: Managing your histories

1. Make sure you are *logged in*.

- If you haven't yet created an account, now is an ideal time to that. Your current history will be saved as your first history under your new account.
2. *Click on the **table icon** at the top-right of the history panel to switch to your *histories* view.*
 - This lists your histories, from left to right in reverse chronological order, based on the last time each history was your current history.
 3. *Switch to the history you want to extract a workflow from.*
 - If the history that you want to create a repeatable workflow for is *not* your *current* history (the left-most one), then find the history you want and then *click* the **Switch to** button at the top of the history.
 - This makes that history the *current history* and moves it to the very left.
 4. *Click the **Done** button at the top upper left of your histories view.*
 - This returns you to the Galaxy home page with the selected history as your *current history*.

Tip: Always name your histories

Extract the recipe from your history

Now that we have the history we want, let's use Galaxy to create a reusable workflow from it. To do this we'll use the

history's **gear** / **history options** pull-down list.

Hands-on: Extract workflow

1. Click on the gear

icon at the top of your history.

- This opens a pull-down menu showing lots of actions that you can perform on this history.

Right now, we are interested in only one of them:

2. Click on Extract Workflow. It's about half-way down the menu.

- This launches a form to create a workflow.

3. Give your workflow a meaningful name.

- Think about giving it a more general name than what your history has. The example history is about looking for overlapping genes on opposite strands in human on chromosome 22. That's pretty specific. We want to create a workflow that can be used to detect any type of overlapping features, not just genes, and in any species and on any chromosome (or on all chromosomes). The name should reflect this goal.

- For this example we'll name it **overlapping features on opposite strands**

4. Remove unwanted steps

- If there were any missteps or dead-ends in your history, uncheck those steps here.
- Don't worry if you miss one - we can drop it in a subsequent step.

5. Once you have named your workflow, click the **Create Workflow** button.

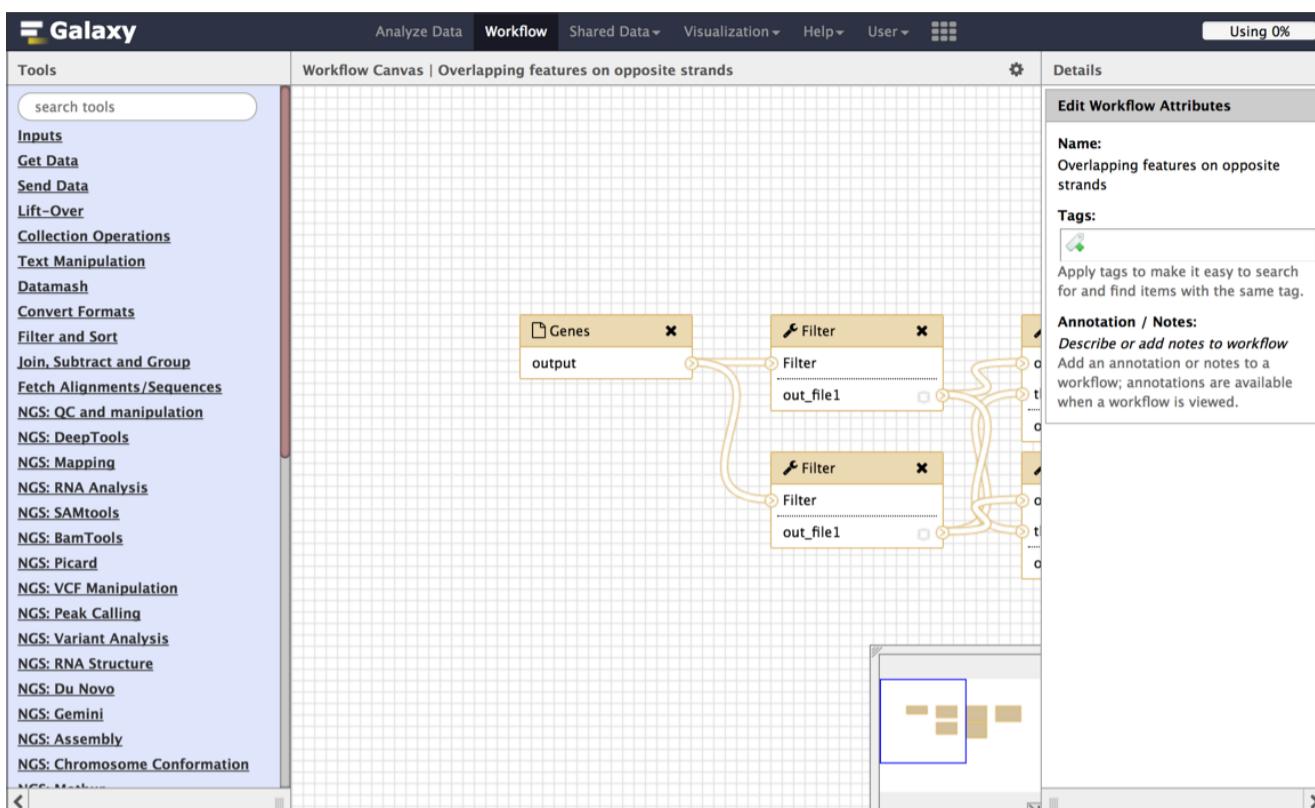
Tool	History items created
UCSC Main <i>This tool cannot be used in workflows</i>	1: Genes 2: Genes, forward strand 3: Genes, reverse strand 4: Overlapping forward genes 5: Overlapping reverse genes 6: Overlapping genes
Filter <input checked="" type="checkbox"/> Include "Filter" in workflow	2: Genes, forward strand
Filter <input checked="" type="checkbox"/> Include "Filter" in workflow	3: Genes, reverse strand
Intersect <input checked="" type="checkbox"/> Include "Intersect" in workflow	4: Overlapping forward genes
Intersect <input checked="" type="checkbox"/> Include "Intersect" in workflow	5: Overlapping reverse genes
Concatenate datasets <input checked="" type="checkbox"/> Include "Concatenate datasets" in workflow	6: Overlapping genes

- This replaces the workflow creation form with a box that says:
 - Workflow *whatever name you entered* created from current history. You can edit or run the workflow.

6. Click on the **edit** link in the box.

This launches the workflow editor and shows a graphical representation of the inputs and steps taken during the analysis. Each box is either an input dataset, or a tool that was run. Boxes are connected by lines that show the flow of data from input datasets, through intermediate tools, to the tools that produce the final output(s). Each tool box show a list of inputs (at the top) and a list of outputs (at the bottom). The flow of data is generally from left to

right.



While we could run this workflow right now, here are a few cleanup items we should do first.

Rename input datasets

The *extract workflow* step assigns the name of input datasets in your history to the corresponding inputs in your workflow. This is very helpful in the short term because it makes it clear which of the boxes are which datasets. But, for a workflow, we want a more general (but still helpful) name for input datasets.

Hands-on: Rename inputs

dataset.

- This changes the right panel to show information about the selected dataset. In

1. Click on the first input

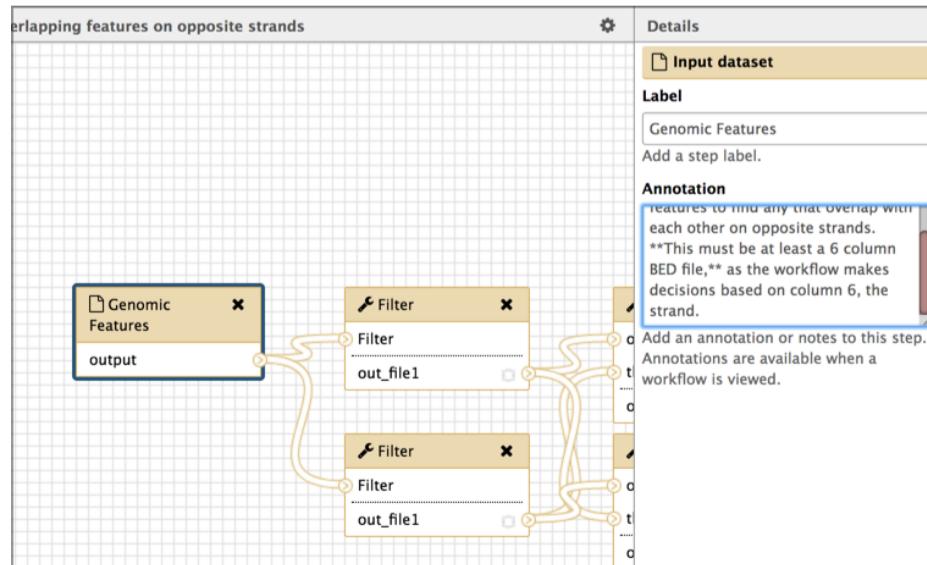
this example, this dataset is a set of genomic features (like genes or exons or repeats) that exist in an organism.

2. Set the **Label** field to something more general, yet still informative.

- In the example workflow, we'll name this Genome features about as general as you can get.

3. Add a *description* for your input datasets

- This helps make your general input dataset name more informative.
- add a description of the dataset in the **Annotation** box.
- For the example, we'll use:
 - Workflow checks this set of genomic features to find any that overlap with each other on opposite strands.
This must be at least a 6 column BED file, as the workflow makes decisions based on column 6, the strand.



There are several other things we could do in the workflow editor, but let's focus on just two of them.

First, your history might contain several false starts or tool runs where the parameters weren't quite right. Lets get rid of those now.

Hands-on: Drop unused input datasets and tool boxes

1. Delete unwanted steps

- If your history contained any false starts or tool runs that didn't contribute to your final result, then *delete* them from the workflow by clicking the black **x** in the corner of those datasets/tools' boxes.
- This will remove them from the workflow.

Tip: Removing unnecessary steps before creating workflow

Second, you might want to give your output datasets meaningful names too.

Hands-on: Name output datasets

1. Click on a step that produces an output file.

- This brings up information about that step in the right panel.

2. Scroll down in the right panel until you see a

Configure Output link. *Click* on it.

3. Enter a meaningful name in the **Rename dataset field.**

- In the example, we'll enter **Features** that

overlap another feature on opposite strand.

4. Repeat this for all your output datasets.

We could add annotation to each step in the process as well, and if this workflow is going to be published in a paper, or shared with others, then that is definitely worth doing.

Now that our edits are done, let's save our work and run the workflow.

Hands-on: Save the workflow edits

1. Click on the **gear icon** at the top right of the central panel, select **Save** from the pull-down menu.

Test the workflow

Now that we have finished creating our workflow, it is time to test it.

Hands-on: Run the workflow

1. Click on the **gear icon** and this time select **Run**.

2. Examine the workflow run form.

This form lists all the inputs and steps in the workflow also asks if you want to run this workflow in a new history, or just add it to your existing history.

In this form you can also change the run-time parameters of any of the steps. We aren't going to do that, but it can be useful when are experimenting with different parameters in different steps. Let's test the workflow by running it on the same input datasets we used in the current history. If we created the workflow correctly, we should get the same results as in the history.

3. Select **Yes** under **Send results to a new history**.
 - Sometimes you'll be building analyses out of component workflows and you will want to run them all in the same history. Here we want a new history because the results will be cleaner and easier to understand.
4. Give the new history a meaningful name.
5. Set your first input dataset using the pull-down menu.
 - Repeat until all input datasets are set.
6. Click the **Run workflow** button.