

Dec 03, 2025

Kraken2 - Taxonomic Classification Workflow on GalaxyTrakr

DOI

<https://dx.doi.org/10.17504/protocols.io.yxmvmbbw6g3p/v1>

Christopher Duda¹, Anna Brover¹, Padmini Ramachandran¹

¹FDA

GenomeTrakr

Tech. support email: genomeTrakr@fda.hhs.gov



Christopher Duda

FDA

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.yxmvmbbw6g3p/v1>

Protocol Citation: Christopher Duda, Anna Brover, Padmini Ramachandran 2025. Kraken2 - Taxonomic Classification Workflow on GalaxyTrakr. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.yxmvmbbw6g3p/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited



Protocol status: In development

We are still developing and optimizing this protocol

Created: October 24, 2025

Last Modified: December 03, 2025

Protocol Integer ID: 230684

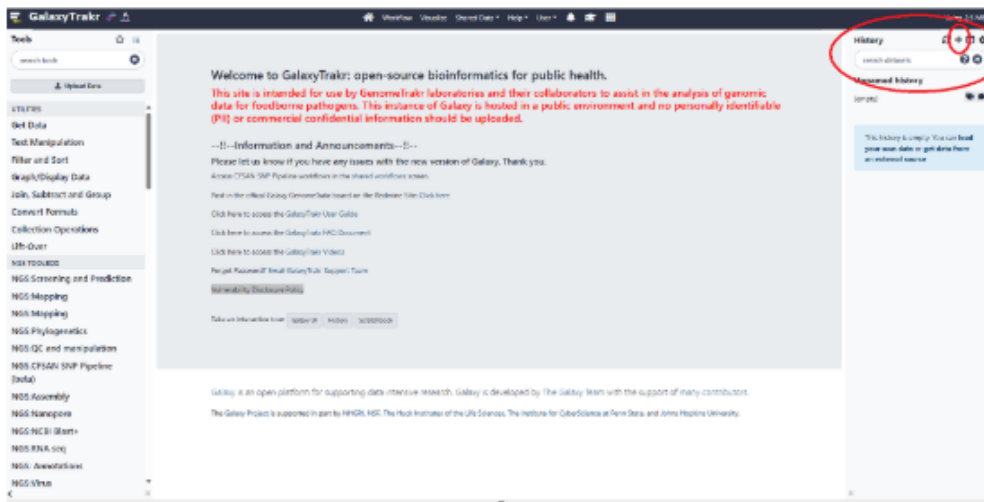
Keywords: taxonomic classification workflow on galaxytrkr kraken2, taxonomic sequence classifier, taxonomic classification workflow, taxonomic label, taxonomic tree, single taxonomic label for the read, taxonomic tree of all genome, nodes in the taxonomic tree, single taxonomic label, galaxytrkr kraken2, kraken2, kraken, genomic library, genomic library to the lowest common ancestor, mer in kraken, genome, microbe, lowest common ancestor, set of lca taxa, lca taxa

Abstract

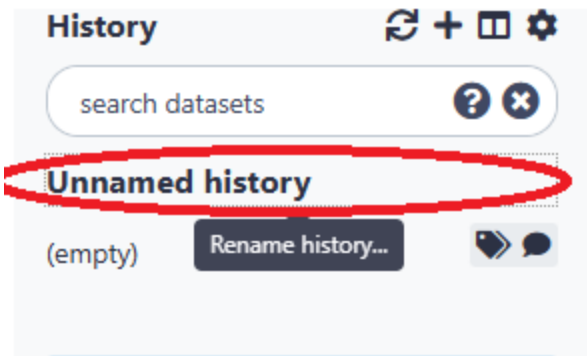
Kraken2 is a taxonomic sequence classifier that assigns taxonomic labels to sequence data. Essentially profiling the microbes that is present in a sample. It does this by examining the k-mers within a read and querying a database with those k-mers. This database contains a mapping of every k-mer in Kraken's genomic library to the lowest common ancestor (LCA) in a taxonomic tree of all genomes that contain that k-mer. The set of LCA taxa that correspond to the k-mers in a read are then analyzed to create a single taxonomic label for the read; this label can be any of the nodes in the taxonomic tree. Kraken is designed to be rapid, sensitive, and highly precise.

Pairing Fastq Datasets

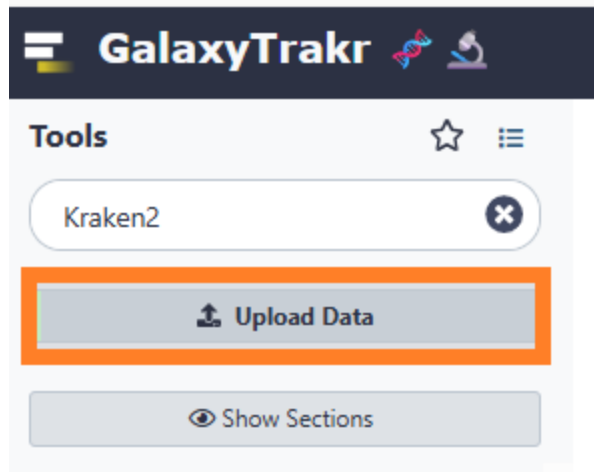
- 1 Log into your GalaxyTrakr account at [Galaxy](#).
- 2 On the right-hand side of the screen, click the plus next to the History tab to create a new history.



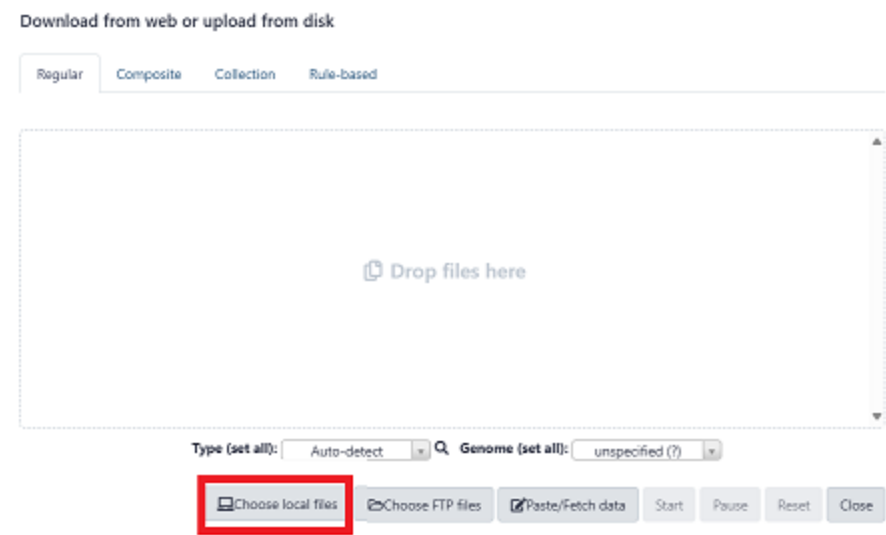
Click on "Unnamed History" and rename it according to your system.



- 3 On the left side of the screen, click the "Upload data" button.



An interface where we can find and select files to upload will appear, for most users, the files will be on local files so choose that option unless otherwise noted.



Select all files to be processed, It should resemble this:
Click "Start" to begin uploading.

Download from web or upload from disk

Regular Composite Collection Rule-based

You added 8 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
4RH_S57_R1_001.fast	186.5 MB	Auto-de...	unspecified (?)		0%
4RH_S57_R2_001.fast	211 MB	Auto-de...	unspecified (?)		0%
4RX_S56_R1_001.fast	186 MB	Auto-de...	unspecified (?)		0%
4RX_S56_R2_001.fast	210.4 MB	Auto-de...	unspecified (?)		0%
4TH_S59_R1_001.fast	148.6 MB	Auto-de...	unspecified (?)		0%
4TH_S59_R2_001.fast	169.8 MB	Auto-de...	unspecified (?)		0%

Type (set all): Auto-detect Genome (set all): unspecified (?)

Choose local files Choose FTP files Paste/Fetch data **Start** Pause Reset Close

Note* Depending on the number of files, it can take a while to upload.
This is what it looks like when uploading has started.

Download from web or upload from disk

Regular Composite Collection Rule-based

Please wait...8 out of 8 remaining.

Name	Size	Type	Genome	Settings	Status
4RH_S57_R1_001.fast	186.5 MB	Auto-de...	unspecified (?)		38%
4RH_S57_R2_001.fast	211 MB	Auto-de...	unspecified (?)		0%
4RX_S56_R1_001.fast	186 MB	Auto-de...	unspecified (?)		0%
4RX_S56_R2_001.fast	210.4 MB	Auto-de...	unspecified (?)		0%
4TH_S59_R1_001.fast	148.6 MB	Auto-de...	unspecified (?)		0%
4TH_S59_R2_001.fast	169.8 MB	Auto-de...	unspecified (?)		0%

Type (set all): Auto-detect Genome (set all): unspecified (?)

Choose local files Choose FTP files Paste/Fetch data Start Pause Reset Close



Download from web or upload from disk

Regular Composite Collection Rule-based

Name	Size	Type	Genome	Settings	Status
 4RH_S57_R1_001.fast	186.5 MB	Auto-de... 	 unspecified (?)		100% 
 4RH_S57_R2_001.fast	211 MB	Auto-de... 	 unspecified (?)		100% 
 4RX_S56_R1_001.fast	186 MB	Auto-de... 	 unspecified (?)		100% 
 4RX_S56_R2_001.fast	210.4 MB	Auto-de... 	 unspecified (?)		100% 
 4TH_S59_R1_001.fast	148.6 MB	Auto-de... 	 unspecified (?)		100% 
 4TH_S59_R2_001.fast	169.8 MB	Auto-de... 	 unspecified (?)		100% 

Type (set all): Auto-detect  Genome (set all): unspecified (?) Choose local files  Choose FTP files  Paste/Fetch data Start Pause Reset **Close**

- 4 Before running "Kraken2", the dataset needs to be paired.
Click on the square icon, or the "Operations on multiple datasets" on the right side of the tab. select boxes will appear on the green file list. This will allow you to select all files in your dataset.

History

search datasets

?

x

Practice Example Data

8 shown, 10 deleted, 8 hidden

2.72 GB

☒
☐
☐
☐
☐
☐
☐
☐

All

None

For all selected... ▼

☐ 18: 4TX_S58_R2_001.fastq.gz

☐ 17: 4TX_S58_R1_001.fastq.gz

☐ 16: 4TH_S59_R2_001.fastq.gz

☐ 15: 4TH_S59_R1_001.fastq.gz

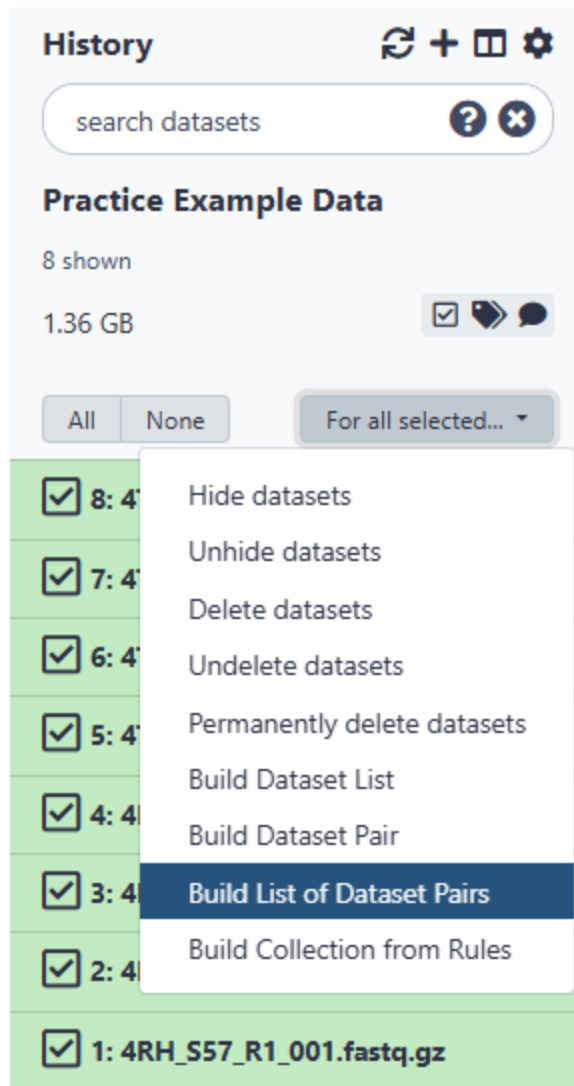
☐ 14: 4RX_S56_R2_001.fastq.gz

☐ 13: 4RX_S56_R1_001.fastq.gz

☐ 12: 4RH_S57_R2_001.fastq.gz

☐ 11: 4RH_S57_R1_001.fastq.gz

Select all files in the workspace, then click the dropdown that says for all selected. A menu will appear, select "Build List of Dataset Pairs".



5 A new interface will appear.

Create a collection of paired datasets

Could not automatically create any pairs from the given dataset names. You may want to choose or enter different filters and try auto-pairing again. [cancel](#) and reselect new elements.

Collections of paired datasets are ordered lists of dataset pairs (often forward and reverse reads). These collections can be passed to tools and workflows in order to have analyses done on each member of the entire group. This interface allows you to create...

_R1

0 unpaired forward - 0 filtered out

Clear Filters
Auto-pair

_R2

0 unpaired reverse - 0 filtered out

No datasets were found matching the current filters.

0 pairs Unpair all

Hide original elements ☒ Remove file extensions ☒

Name:

[Cancel](#) [Create collection](#)

In the dropdown on the left, select `_R1`. This will auto select `_R2` for the other dropdown.

Create a collection of paired datasets

Could not automatically create any pairs from the given dataset names. You may want to choose or enter different filters and try auto-pairing again. [cancel](#) and reselect new elements.

Collections of paired datasets are ordered lists of dataset pairs (often forward and reverse reads). These collections can be passed to tools and workflows in order to have analyses done on each member of the entire group. This interface allows you to create...

_R1

0 unpaired forward - 0 filtered out

Clear Filters
Auto-pair

_R2

0 unpaired reverse - 0 filtered out

No datasets were found matching the current filters.

0 pairs Unpair all

Hide original elements ☒ Remove file extensions ☒

Name:

[Cancel](#) [Create collection](#)

Create a collection of paired datasets

Could not automatically create any pairs from the given dataset names. You may want to choose or enter different filters and try auto-pairing again. [cancel](#) and reselect new elements.

Collections of paired datasets are ordered lists of dataset pairs (often forward and reverse reads). These collections can be passed to tools and workflows in order to have analyses done on each member of the entire group. This interface allows you to create...

_R1

4 unpaired forward - 4 filtered out

Clear Filters
Auto-pair

_R2

4 unpaired reverse - 4 filtered out

43M_15K_R1_001.fastq.gz	Pair these datasets	43M_15K_R2_001.fastq.gz
43M_15K_R1_002.fastq.gz	Pair these datasets	43M_15K_R2_002.fastq.gz
43M_15K_R1_003.fastq.gz	Pair these datasets	43M_15K_R2_003.fastq.gz
43M_15K_R1_004.fastq.gz	Pair these datasets	43M_15K_R2_004.fastq.gz

4 pairs Unpair all

You will see both columns populate with filenames. Double check that the files have the same name other than "R1" or "R2". Once you know the program has properly paired the

files, Click the "Pair these datasets" for all entries. There two columns will be empty if you do it correctly and they appear in green below the black bar.

The screenshot displays the Protocols.io interface for creating a collection of paired datasets. At the top, there are two search bars, one for '4' and one for '2', both showing '0 unpaired forward' and '0 filtered out'. Below the search bars, a dropdown menu is open, showing 'No datasets were found' and a search input field. The middle section shows a list of datasets with 'Pair these datasets' buttons. The bottom section shows a collection of paired datasets with a 'Create collection' button.

Below the search bars, there is a section titled 'Collections of paired datasets are ordered lists of dataset pairs (often forward and reverse reads). These collections can be passed to tools and workflows in order to have analyses done on each member of the entire group. This interface allows you to create...'.

The interface shows two columns of datasets, each with a 'Pair these datasets' button. The datasets are listed as follows:

- Column 1: 409L_1557_R1_001.fastq.gz, 409L_1557_R2_001.fastq.gz, 47H_1559_R1_001.fastq.gz, 47H_1559_R2_001.fastq.gz
- Column 2: 409L_1557_R1_001.fastq.gz, 409L_1557_R2_001.fastq.gz, 47H_1559_R1_001.fastq.gz, 47H_1559_R2_001.fastq.gz

Below the datasets, there is a section titled 'Create a collection of paired datasets'. It contains a message: 'Could not automatically create any pairs from the given dataset names. You may want to choose or enter different files and try auto-pairing again, cancel and loaded new elements.' Below this message, there is a section titled 'Collections of paired datasets are ordered lists of dataset pairs (often forward and reverse reads). These collections can be passed to tools and workflows in order to have analyses done on each member of the entire group. This interface allows you to create...'. Below this section, there is a search bar for '4' and a dropdown menu showing '0 unpaired forward' and '0 filtered out'. Below the search bar, there is a message: 'No datasets were found matching the current filters.' Below this message, there is a section titled '4 pairs (unpaired)' showing a list of paired datasets. The datasets are listed as follows:

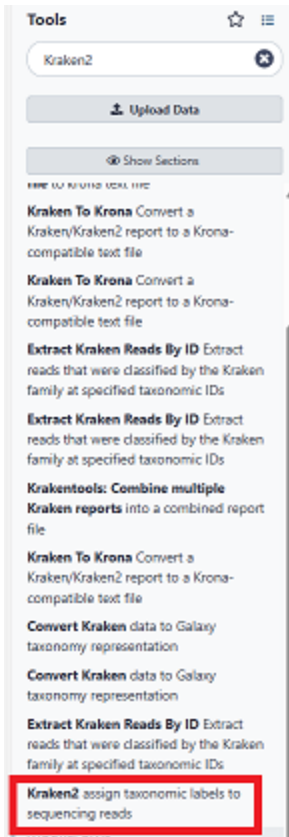
- 409L_1557_R1_001.fastq.gz → 409L_1557_R2_001.fastq.gz
- 409L_1557_R2_001.fastq.gz → 409L_1557_R1_001.fastq.gz
- 47H_1559_R1_001.fastq.gz → 47H_1559_R2_001.fastq.gz
- 47H_1559_R2_001.fastq.gz → 47H_1559_R1_001.fastq.gz

Below the list of paired datasets, there is a section titled 'Hide original elements' and 'Remove this auto-pairing'. Below this section, there is a 'Name' field with the text 'Enter a name for your new collection' and a 'Create collection' button.

6 Create a name for your collection and click the "Create collection" button.

Creating Graphical Reports

- 6.1 On the left-hand side of the screen, click the search bar in the "tools" section and search for **"Kraken2"**.



6.2 An interface will appear looking like this.



Kraken2 assign taxonomic labels to sequencing reads (Galaxy Version 2.1.3+galaxy1)

Single or paired reads

Paired Collection

--paired

! Please provide a value for this option.
Collection of paired reads
No fasta, fastq, fasta.gz, fasta.bz2, fastq.gz or fastq.bz2 dataset collection available.

Print scientific names instead of just taxids
☒ No
(--use-names)

Confidence
0

Confidence score threshold. Must be in [0, 1] (--confidence)

Minimum Base Quality
0

Minimum base quality used in classification (only effective with FASTQ input) (--minimum-base-quality)

Minimum hit groups
2

Number of overlapping k-mers sharing the same minimizer needed to make a call (--minimum-hit-groups)

Enable quick operation
☒ No
Quick operation (use first hit) (--quick)

Split classified and unclassified outputs?
☒ No
Sets --unclassified-out and --classified-out

[Create Report](#)

Select a Kraken2 database

Change the "Select to read collection type" to "paired collection".

Single or paired reads

Paired Collection

--paired

! Please provide a value for this option.
Collection of paired reads
No fasta, fastq, fasta.gz, fasta.bz2, fastq.gz or fastq.bz2 dataset collection available.

- 6.3 Click on the slider under "Print scientific name instead of just taxids" so it says yes and is a solid dark blue color.



Kraken2 assign taxonomic labels to sequencing reads (Galaxy Version 2.1.3+galaxy1)

Single or paired reads

Single

--paired

! Please provide a value for this option.

Input sequences

No fasta, fastq, fasta.gz, fasta.bz2, fastq.gz or fastq.bz2 dataset available.

Print scientific names instead of just taxids

☒ Yes

(--use-names)

Confidence

0

Confidence score threshold. Must be in [0, 1] (--confidence)

Minimum Base Quality

0

Minimum base quality used in classification (only effective with FASTQ input) (--minimum-base-quality)

Minimum hit groups

2

Number of overlapping k-mers sharing the same minimizer needed to make a call (--minimum-hit-groups)

Enable quick operation

☐ No

Quick operation (use first hit) (--quick)

Split classified and unclassified outputs?

☐ No

6.4 Click on "Create Report"

Kraken2 assign taxonomic labels to sequencing reads (Galaxy Version 2.1.3+galaxy1)

Single or paired reads

Single

--paired

! Please provide a value for this option.

Input sequences

No fasta, fastq, fasta.gz, fasta.bz2, fastq.gz or fastq.bz2 dataset available.

Print scientific names instead of just taxids

☒ Yes

(--use-names)

Confidence

0

Confidence score threshold. Must be in [0, 1] (--confidence)

Minimum Base Quality

0

Minimum base quality used in classification (only effective with FASTQ input) (--minimum-base-quality)

Minimum hit groups

2

Number of overlapping k-mers sharing the same minimizer needed to make a call (--minimum-hit-groups)

Enable quick operation

☐ No

Quick operation (use first hit) (--quick)

Split classified and unclassified outputs?

☐ No

Sets --unclassified-out and --classified-out

Create Report

Click on the slider under "Print a report with aggregate counts/clades to a file" so it says yes and is a solid dark blue color.



Create Report

Print a report with aggregate counts/clade to file

☒ Yes

--report

Format report output like Kraken 1's kraken-mpa-report

☐ No

(--use-mpa-style)

Report counts for ALL taxa, even if counts are zero

☐ No

(--report-zero-counts)

Report minimizer data

☐ No

Report minimizer and distinct minimizer count information in addition to normal Kraken report (--report-minimizer-data)

6.5 Next you need to pick a kraken database. click on the dropdown under **Select a Kraken2 database** and select "Minikraken2 v2".

Select a Kraken2 database *

Minikraken2 v2 (Created: 2025-11-16T125646Z)

Additional Options

Email notification

☐ No

Send an email notification when the job completes.

Attempt to re-use jobs with identical parameters?

☐ No

This may skip executing jobs that you have already run.

[Run Tool](#)

Select a Kraken2 database *

Select Value

minikraken2_v1 (Created: 2025-05-12T234311Z)	☐
Minikraken2 v2 (Created: 2025-05-10T013145Z)	☐
Prebuilt Refseq Indexes: Viral (Version: 2024-12-28 - Downloaded: 2025-06-03T141450Z)	☐
Prebuilt Refseq Indexes: Standard-Full (archaea, bacteria, viral, plasmid, human, UniVec_Core) (Version: 2024-12-26 - Downloaded: 2025-06-03T143959Z)	☐
Prebuilt Refseq Indexes: Standard-Full (archaea, bacteria, viral, plasmid, human, UniVec_Core) (Version: 2025-07-14 - Downloaded: 2025-11-17T191603Z)	☐
Minikraken2 v2 (Created: 2025-11-16T125646Z)	☒

6.6 Click on the "Run Tool" button.

Select a Kraken2 database *

Minikraken2 v2 (Created: 2025-11-16T125648Z)

Additional Options

Email notification

☐ No

Send an email notification when the job completes.

Attempt to re-use jobs with identical parameters?

☐ No

This may skip executing jobs that you have already run.

[▶ Run Tool](#)

The report is generating once a green box appears. It should look similar to this.



Started tool **Kraken2** and successfully added 4 jobs to the queue.

It produces 6 outputs:

- 23: Kraken2 on data 12 and data 11: Report
- 24: Kraken2 on data 12 and data 11: Classification
- 25: Kraken2 on data 16 and data 15: Report
- 26: Kraken2 on data 16 and data 15: Classification
- 27: Kraken2 on data 16 and data 17: Report
- 28: Kraken2 on data 16 and data 17: Classification
- 29: Kraken2 on data 14 and data 13: Report
- 30: Kraken2 on data 14 and data 13: Classification

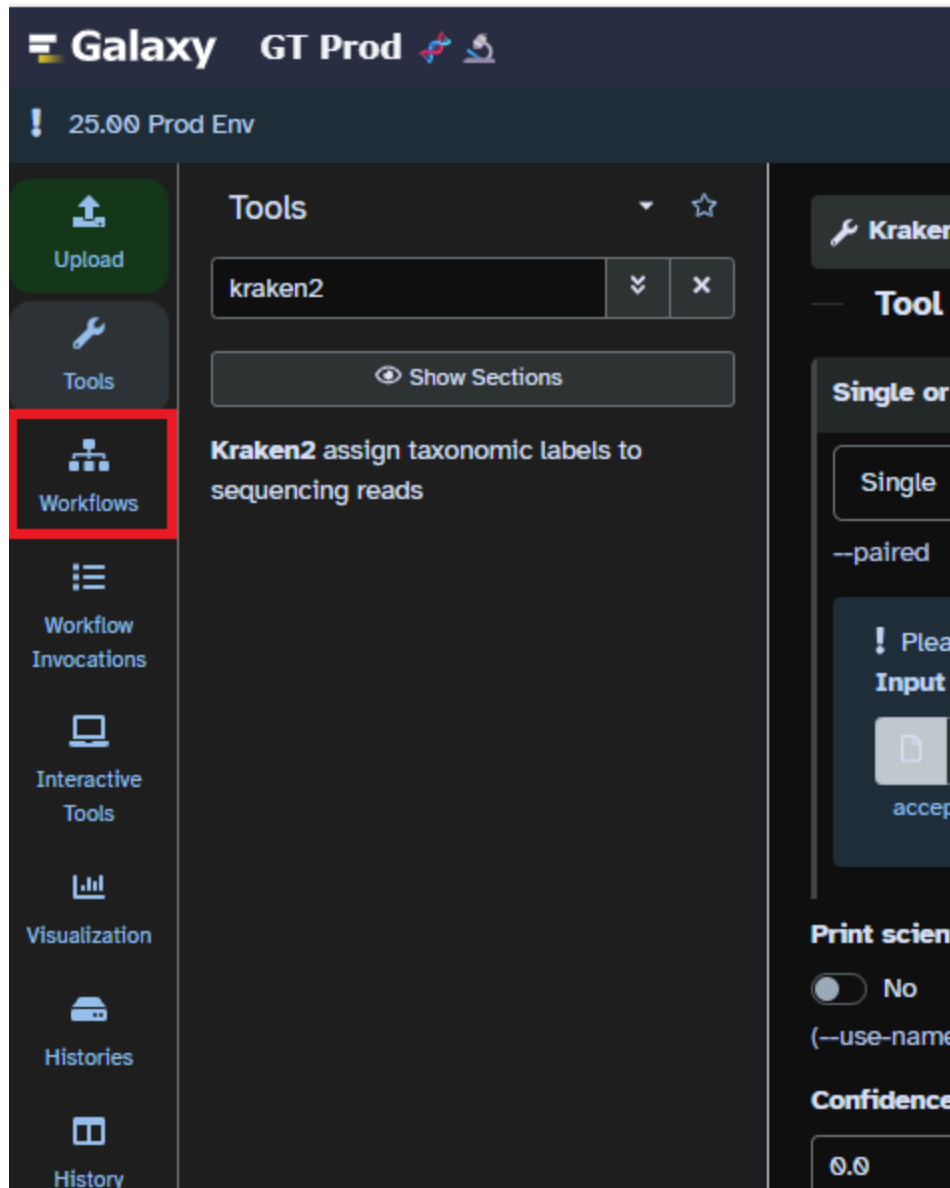
You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

Here is a link to each job: [d47783ffed26ef0](#) [16466d8aaa2868bf](#) [338a631b552dc5ea](#) [b1b55241016776cb](#)

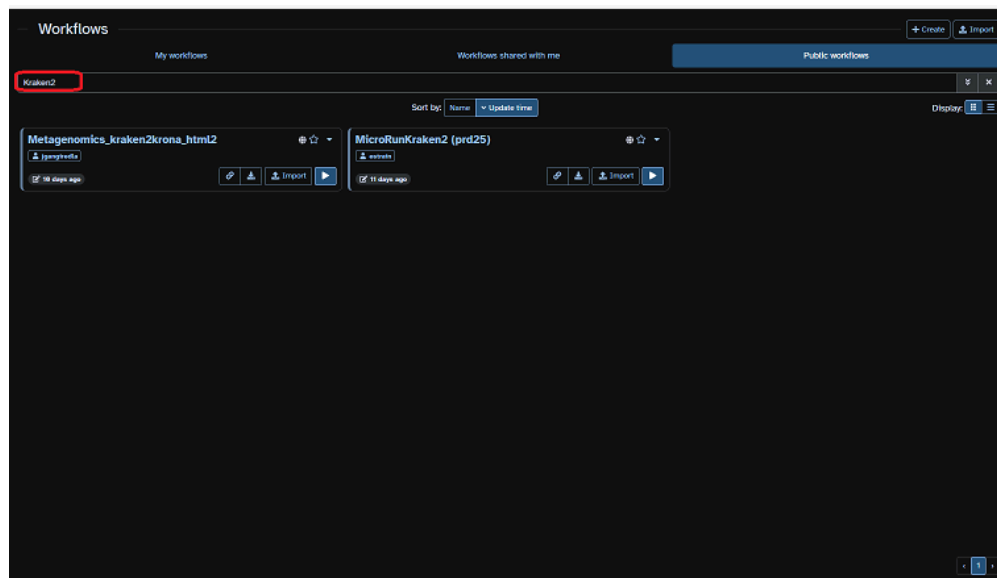
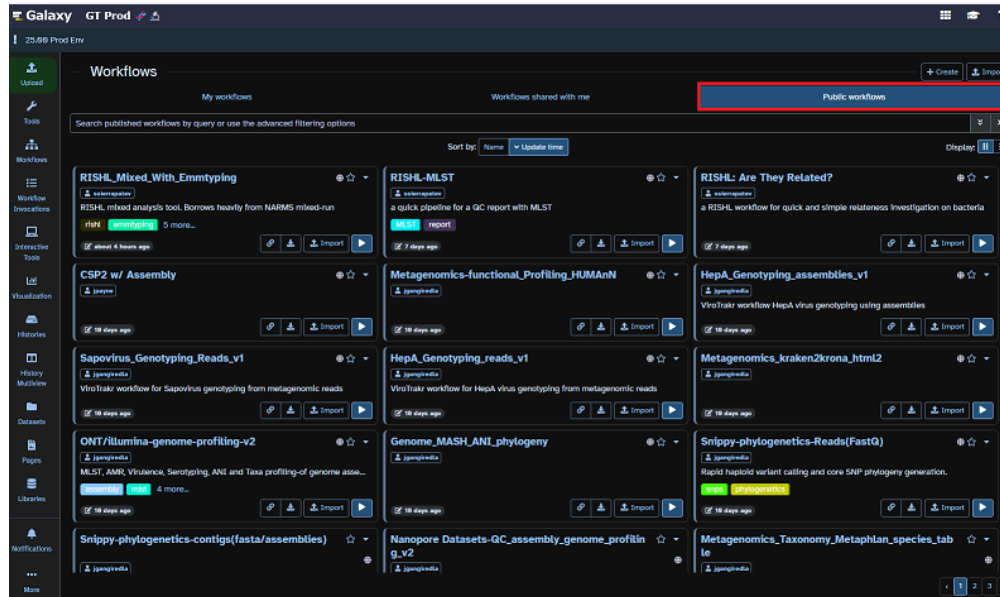
This process usually takes over an hour to generate the report.

6.7 Generating graphical reports.

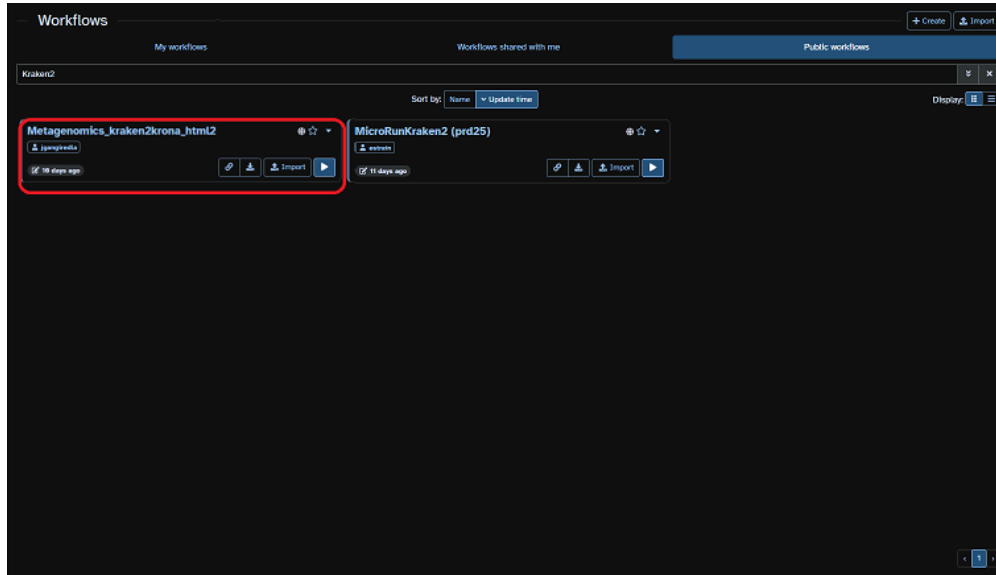
On the left-hand side of the screen, click the tab titled "workflows".



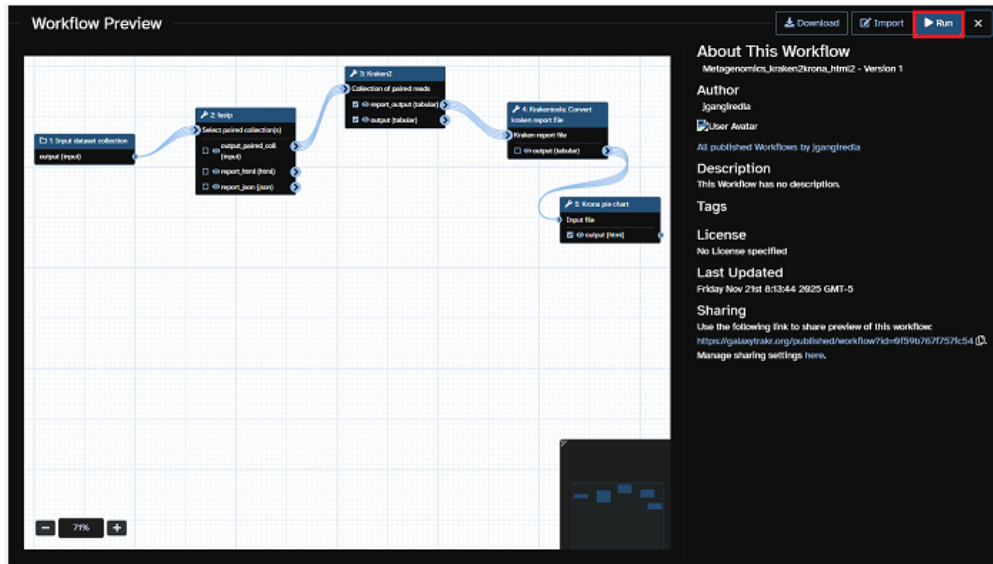
6.8 Click on the "Public workflows" tab on the top right and input kraken2 in the search bar.



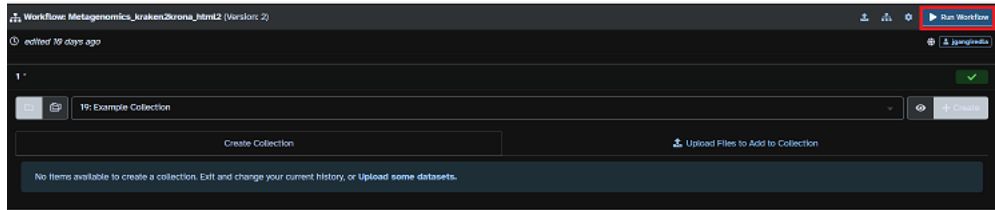
6.9 Select the workflow titled "**Metagenomics_kraken2krona_html2**".



6.10 Click the "Run" button on the top right of the window that appears.

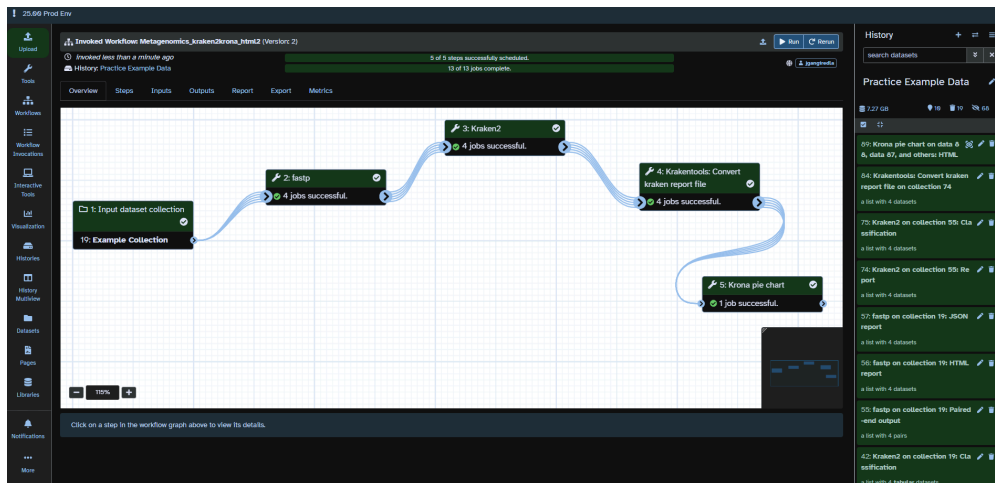


6.11 Make sure your paired fastq datasets are selected and click the "Run Workflow" button.

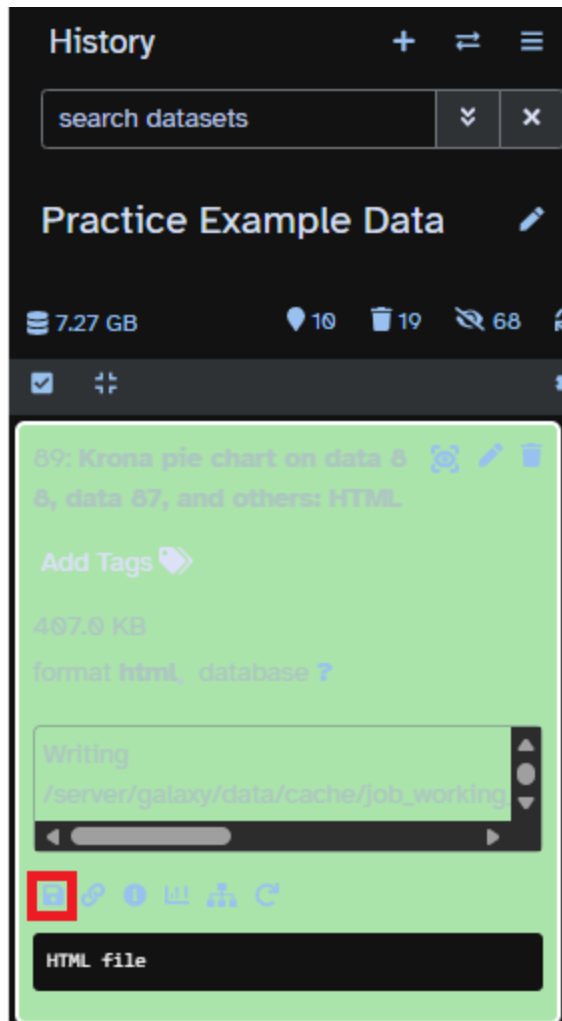


This will start the process for generating graphical reports. This can take over an hour to generate depending on size and traffic.

It will look like this when it is finished:



- 7 To access the figures, click on the desired file under the history and click download.



This will download a .zip file containing the figure for your data.



Protocol references

```
@article{Wood_2014,  
  journal = {Genome Biology},  
  number = {3},  
  publisher = {Springer Science and Business Media LLC},  
  title = {Kraken: ultrafast metagenomic sequence classification using exact alignments},  
  url = {http://dx.doi.org/10.1186/gb-2014-15-3-r46},  
  volume = {15},  
  author = {Wood, Derrick E and Salzberg, Steven L},  
  date = {2014-03},  
  year = {2014},  
  month = {3},  
}
```