

# Data Analysis using the Galaxy Australia Research Computing Platform

Dr Gayle Philip (Bioinformatician)

Dr Tristan Reynolds (Galaxy Outreach Specialist)





THE UNIVERSITY OF  
MELBOURNE

The University of Melbourne acknowledges the Traditional Owners of the unceded land on which we work, learn and live: the Wurundjeri Woi-wurrung and Bunurong peoples (Burnley, Fishermans Bend, Parkville, Southbank and Werribee campuses), the Yorta Yorta Nation (Dookie and Shepparton campuses), and the Dja Dja Wurrung people (Creswick campus).

The University also acknowledges and is grateful to the Traditional Owners, Elders and Knowledge Holders of all Indigenous nations and clans who have been instrumental in our reconciliation journey.

We recognise the unique place held by Aboriginal and Torres Strait Islander peoples as the original owners and custodians of the lands and waterways across the Australian continent, with histories of continuous connection dating back more than 60,000 years. We also acknowledge their enduring cultural practices of caring for Country.

We pay respect to Elders past, present and future, and acknowledge the importance of Indigenous knowledge in the Academy. As a community of researchers, teachers, professional staff and students we are privileged to work and learn every day with Indigenous colleagues and partners.

In making this Acknowledgment of Country we commit to respectful and responsible conduct towards all others according to the Traditional lores of this land, particularly at times of formal ceremony.

# What is Galaxy?



- A web interface to a complex mixture of computing resources
- A launching site for analysis tool(s) with pre-determined and scalable resources keyed to a user's dataset size
- A repository for reference data required to support analysis
- A platform for creating and running workflows (pipelines)
- A platform for sharing data and results

# The Galaxy Interface

A screenshot of the Galaxy Australia web interface. The interface is divided into three main panels: Tool Panel (left), View Panel (center), and Current History (right).

**Tool Panel:** Displays a sidebar with various tool categories and sub-tools. Categories include FILE AND META TOOLS, GENERAL TEXT TOOLS, GENOMIC FILE MANIPULATION, and others like FASTA/FASTQ, SAM/BAM, BED, VCF/BCF, Nanopore, Convert Formats, and Lift-Over. A search bar at the top of the sidebar allows users to search for specific tools.

**View Panel:** The central panel displays a banner for the Galaxy platform 2024 update, highlighting the latest developments in accessible, reproducible, and collaborative data analysis. It also features a brief description of Galaxy Australia's mission and capabilities.

**Current History:** The right panel shows the user's current history of datasets. The history list includes entries such as "genome\_assembly\_treynolds" (1.35 GB), "13: Quast on data 4 and data 1", "0: HTML report", "12: Busco on data 10: full table", "11: Busco on data 10: short sum", "10: Create assemblies with Unicycler on data 7, data 6, and data 5: Final Assembly", "9: Create assemblies with Unicycler on data 7, data 6, and data 5: Final Assembly Graph", "8: Create assemblies with Unicycler on data 7, data 6, and data 5: SPAdes graphs", and a note about a list with 15 gfa1 datasets.

Tool Panel

Data Analysis using Galaxy Australia (2025)

View Panel

Current History

# Galaxy Workflows



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Using 1% of 600.0 GB

treynolds

Tools

search tools

Inputs

FILE AND META TOOLS

- Get Data
- Send Data
- Collection
- Operations

GENERAL TEXT TOOLS

- Text Manipulation
- Filter and Sort
- Join, Subtract and Group

GENOMIC FILE MANIPULATION

- FASTA/FASTQ
- FASTQ Quality Control
- SAM/BAM
- BED

Unnamed Workflow

Quality Checking

Search against known database

Create Reference

BIOM format conversion

Normalisation

Plots

Name: Unnamed Workflow

Annotation:

These notes will be visible when this workflow is viewed.

License: Specify a license for this workflow.

Creator: Add a new creator - either a person or an organization.

Tags: Add Tags

Apply tags to make it easy to search for and find items with the same tag.

100%

# Uploading Data



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Using 35% of 600.0 GB    treynolds

Upload

Tools

search tools

FILE AND META TOOLS

- Get Data
- Send Data
- Collection Operations

GENERAL TEXT TOOLS

- Text Manipulation
- Filter and Sort
- Join, Subtract and Group

GENOMIC FILE MANIPULATION

- FASTA/FASTQ
- FASTQ Quality Control
- SAM/BAM
- BED
- VCF/BCF
- Nanopore
- Convert Formats
- Lift-Over

History

search datasets

RNA-Seq reads to counts

mb-workshops

975 MB    47    2    335

41: MultiQC on data 39, data 37, and others: Stats

40: MultiQC on data 39, data 37, and others: Webpage

#fastqc-raw

15: FastQC on collection 1: RawData

a

a list with 12 txt datasets

14: FastQC on collection 1: Webpage

e

e

a list with 12 html datasets

1: fastqs

i

Galaxy Australia has been upgraded to release 24.2. Check the rel...

Home News People About Support Docs

Galaxy AUSTRALIA

Galaxy platform 2024 update published!

Read the latest developments supporting accessible, reproducible, and collaborative data analyses

With contributions from 130 authors representing 60 institutions

doi.org/10.1093/nar/gkae410

usegalaxy.\*

Galaxy Australia is an **open, web-based** platform for accessible, reproducible and transparent computational research. Galaxy supports thousands of documented and maintained tools that are free to use. We facilitate on-demand training capacities and provision **600GB** for Australian institutional (and 100GB for other) users.

# Uploading Data



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Upload Tools search tools

Tools

FILE AND META 1

Get Data Send Data Collection Operations

Workflow Invocations

Histories

History Multiview

Datasets

Libraries

Notifications

More

Upload from Disk or Web to RNA-Seq reads to counts

Regular Composite Collection Rule-based

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

New File 0 b Auto-detect unspecified (?) 0% [trash]

Download data from the web by entering URLs (one per line) or directly paste content.

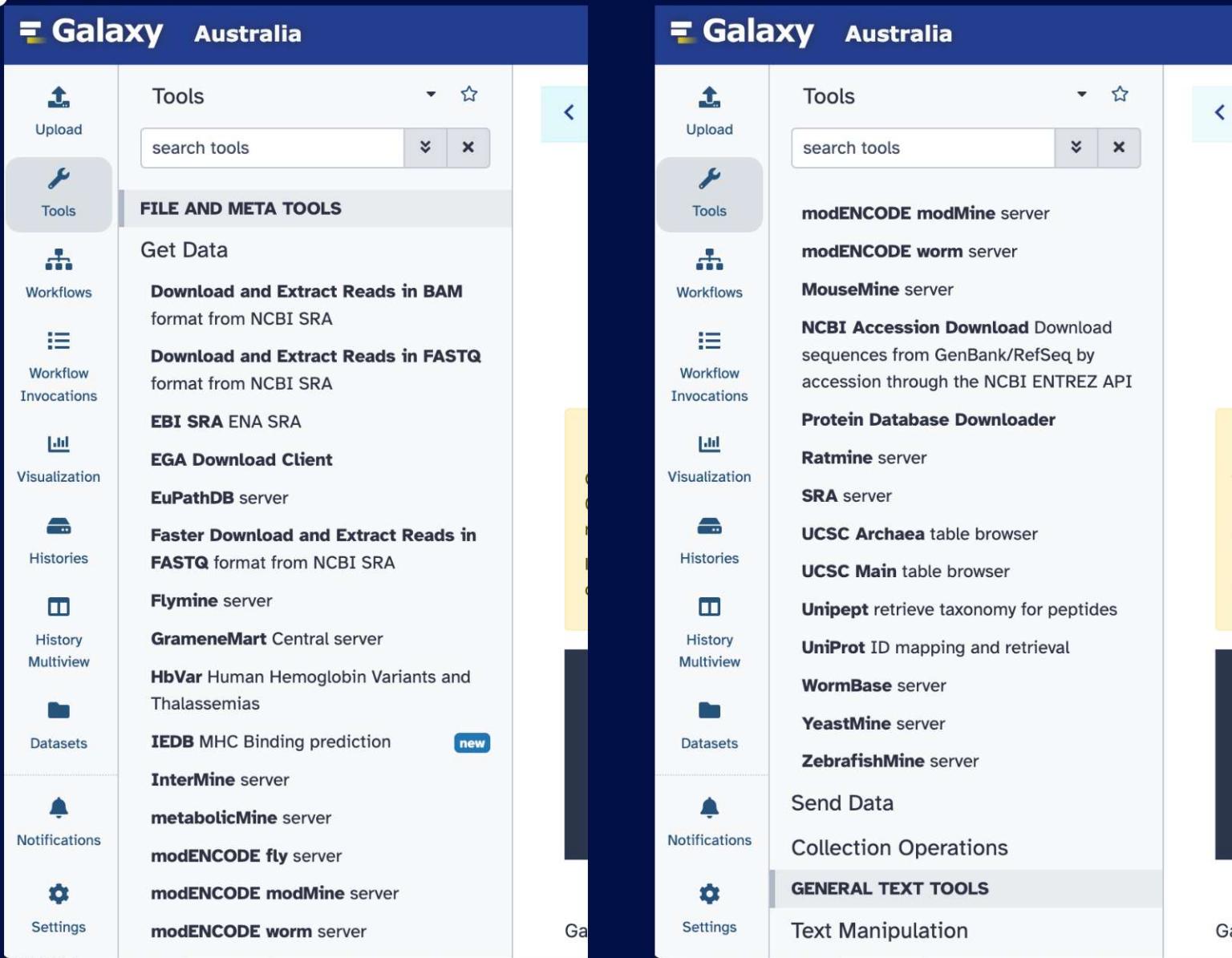
Type (set all): Auto-detect Reference (set all): unspecified (?)

Choose local file Choose remote files Paste/Fetch data Start Pause Reset Close

Galaxy Australia is an open, web-based platform for accessible, reproducible and

The screenshot shows the Galaxy Australia interface for uploading data. On the left, a sidebar lists various tools and data types. The main area is titled "Upload from Disk or Web to RNA-Seq reads to counts". It features tabs for "Regular", "Composite", "Collection", and "Rule-based". A message indicates "You added 1 file(s) to the queue. Add more files or click 'Start' to proceed." Below this are fields for "New File" (0 b), "Auto-detect", "unspecified (?)", and a progress bar at 0%. A note says "Download data from the web by entering URLs (one per line) or directly paste content." At the bottom, there are buttons for "Type (set all)" (Auto-detect), "Reference (set all)" (unspecified ?), "Choose local file" (highlighted with a red border), "Choose remote files", "Paste/Fetch data" (highlighted with a red border), "Start", "Pause", "Reset", and "Close". A footer message states "Galaxy Australia is an open, web-based platform for accessible, reproducible and".

# Uploading Data



The screenshot shows the Galaxy Australia web interface. The left panel displays the 'FILE AND META TOOLS' section under the 'Tools' category. The right panel displays the 'GENERAL TEXT TOOLS' section under the 'Tools' category.

**Left Panel (FILE AND META TOOLS):**

- Get Data
  - Download and Extract Reads in BAM** format from NCBI SRA
  - Download and Extract Reads in FASTQ** format from NCBI SRA
  - EBI SRA ENA SRA**
  - EGA Download Client**
  - EuPathDB** server
  - Faster Download and Extract Reads in FASTQ** format from NCBI SRA
  - Flymine** server
  - GrameneMart** Central server
  - HbVar** Human Hemoglobin Variants and Thalassemias
  - IEDB MHC Binding prediction** (new)
  - InterMine** server
  - metabolicMine** server
  - modENCODE fly** server
  - modENCODE modMine** server
  - modENCODE worm** server

**Right Panel (GENERAL TEXT TOOLS):**

- modENCODE modMine** server
- modENCODE worm** server
- MouseMine** server
- NCBI Accession Download** Download sequences from GenBank/RefSeq by accession through the NCBI ENTREZ API
- Protein Database Downloader**
- Ratmine** server
- SRA** server
- UCSC Archaea** table browser
- UCSC Main** table browser
- Unipept** retrieve taxonomy for peptides
- UniProt** ID mapping and retrieval
- WormBase** server
- YeastMine** server
- ZebrafishMine** server
- Send Data**
- Collection Operations**

# Running a Tool



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Using 35% of 600.0 GB    treynolds

Tools    bowtie2    Run Tool

Tool Parameters

Is this single or paired library

Single-end

FASTA/Q file \*

304: fastqs\_v2

accepted formats

The supplied input will be mapped over this tool.

Must be of datatype "fastqsanger" or "fasta"

Write unaligned reads (in fastq format) to separate file(s)

No

--un/--un-conc (possibly with -gz or -bz2); This triggers --un parameter for single reads and --un-conc for paired reads

Write aligned reads (in fastq format) to separate file(s)

No

--al/--al-conc (possibly with -gz or -bz2); This triggers --al parameter for single reads and --al-conc for paired reads

Will you select a reference genome from your history or use a built-in index?

Use a built-in genome index

Built-ins were indexed using default options. See 'Indexes' section of help below.

History

search datasets

RNA-Seq reads to counts

mb-workshops

975 MB    47    2    335

41: MultiQC on data 39, data 37, and others: Stats

40: MultiQC on data 39, data 37, and others: Webpage

#fastqc-raw

15: FastQC on collection 1: RawData

a

a list with 12 txt datasets

14: FastQC on collection 1: Webpage

e

e a list with 12 html datasets

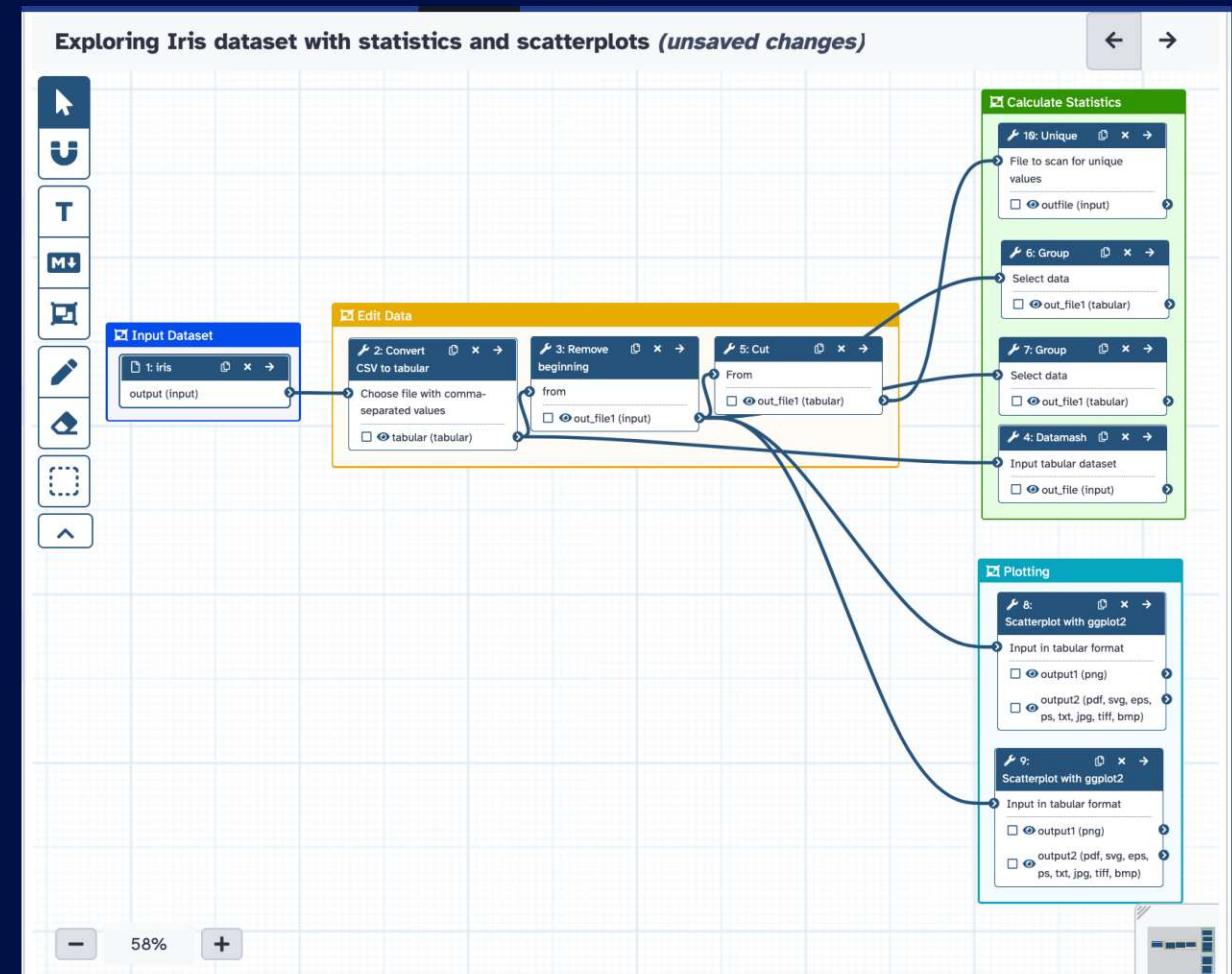
1: fastqs

i

a list with 12 fastqsanger.gz datasets

# Galaxy Workflows

- Workflows logically connect a collection of steps including:
  - Inputs, Tools, Workflows
- Links the outputs from one tool to the inputs of another tool
- Enable performing the same analysis steps on different datasets for consistency/reproducibility
- Saves tool parameters (however these can be updated for a specific Workflow run)
- Automates running of steps in the Workflow (steps will be scheduled once all inputs are available)



# Import Published Workflows



THE UNIVERSITY OF  
MELBOURNE

The screenshot displays two web interfaces side-by-side: WorkflowHub and Dockstore.

**WorkflowHub (Left):**

- Header:** WorkflowHub, Search, About, Help, Register, Log in.
- Section:** Workflows.
- Search Bar:** Search here... (with a magnifying glass icon).
- Filter:** Workflow type: Galaxy.
- List:** Shows several Galaxy workflows:
  - gromacs-dctmd/main: Perform dcTMD free energy simulations and.
  - tissue-microarray-analysis/main: Complete multiplex tissue image (MTI) analysis.
- Filters:** Created At (Any time), Updated At (Any time), Tool (MultiQC, fastp, SAMtools, BUSCO, BWA, Minimap2, More).

**Dockstore (Right):**

- Header:** Dockstore, Search, Organizations, About, Docs, Forum, Login, Register.
- Section:** Organizations / The Galaxy Intergalactic Workflow Commission.
- Organization Details:** The Galaxy Intergalactic Workflow Comission (37 members).
- Variant Calling:** Collections 17, Members 9, Updates 10.
- Vertebrates Genome Project:** Galaxy Workflows for the Vertebrates Genome Project.
- About the Organization:** URL: <https://github.com/galaxyproject/iwc>.
- IWC - Intergalactic Workflow Commission:** Reproducibility is important. Our goals are to:
  - foster workflow use
  - incorporate versioning
  - capture more metadata: (names, versions, authors, use cases, etc.)
  - help scientists find workflows!

# Create Workflow/Pipeline from History



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Using 1% of 600.0 GB treynolds

History + ⌛ ⌂

You have 5 histories.

- Show Histories Side-by-Side
- Resume Paused Jobs
- Copy this History
- Delete this History
- Export Tool Citations
- Export History to File
- Archive History
- Extract Workflow** (highlighted)
- Show Invocations
- Share or Publish
- Set Permissions
- Make Private

69: featureCounts on collection 44: Summary

a list with 2 datasets

Geneid	GSM461177_subsampled
Geneid	GSM461177_subsampled
FBgn0286778	15347
FBgn0000559	13752
FBgn0026372	9622
FBgn0263782	9570
FBgn0283427	9022
FBgn0261602	7688
FBgn0001942	7658
FBgn0023169	7053
FBgn0004167	6953
FBgn0263986	6531
FBgn0011828	6521
FBgn0261800	6115
FBgn0001114	5995
FBgn0000299	5859
FBti0019356	5778
FBgn0001507	5402

Upload

Tools sort ⌄ ⌁

Show Sections

Sort data in ascending or descending order

Sort data in ascending or descending order

**Sort collection**

**Sort assembly**

Sort.seqs put sequences in different files in the same order

SortSam sort SAM/BAM dataset

Sort a row according to their columns

Filter with SortMeRNA of ribosomal RNAs in metatranscriptomic data

bedtools SortBED order the intervals

Workflows

Workflow Invocations

Visualization

Histories

History Multiview

Datasets

Settings

# Create Workflow



THE UNIVERSITY OF  
MELBOURNE

Workflows

My workflows Workflows shared with me Public workflows

+ Create Import

Search my workflows by query or use the advanced filtering options

Sort by: Name Update time Filter: Show deleted Show bookmarked

Galaxy Australia

Using 35% of 600.0 GB treyolds

edited about 20 hours ago  
**Unnamed Workflow**

never run Add Tags

Attributes Inputs Tools Workflows Report Best Practices Changes Notifications Options Save + Exit

Genome assembly Add Tags

FASTQ Quality Control  
FastQC Read Quality reports  
FASTA/FASTQ

Falco An alternative, more performant implementation of FastQC for high throughput sequence quality control

Convert Formats  
Tabular to FASTQ converter

Mothur

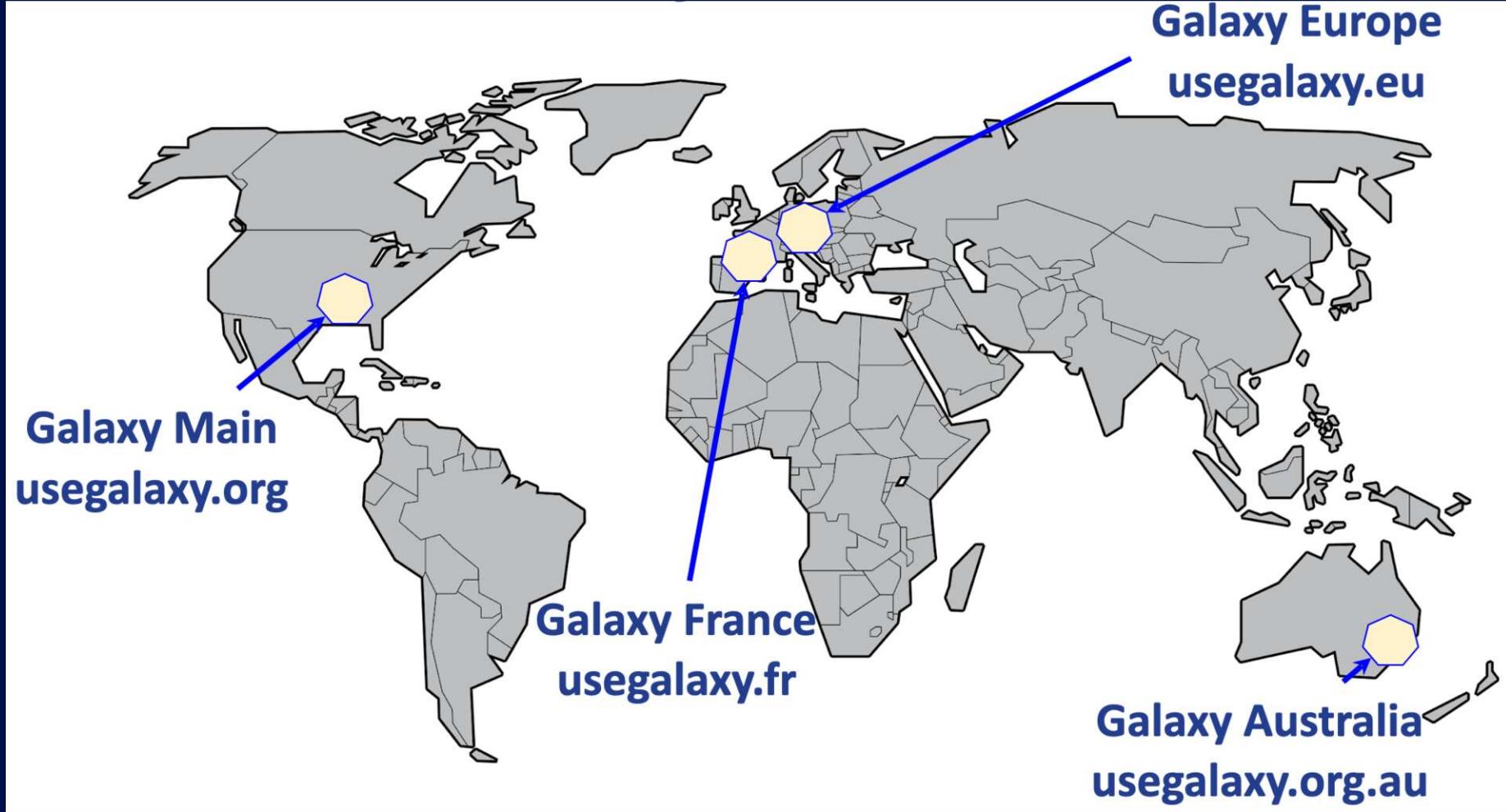
Make.fasta Convert fasta and quality to fastq

1: Input Dataset output (input)

2: FastQC Raw read data from your current history  
Contaminant list  
Adapter list  
Submodule and Limit specifying file  
html\_file (html)  
text\_file (txt)

100%

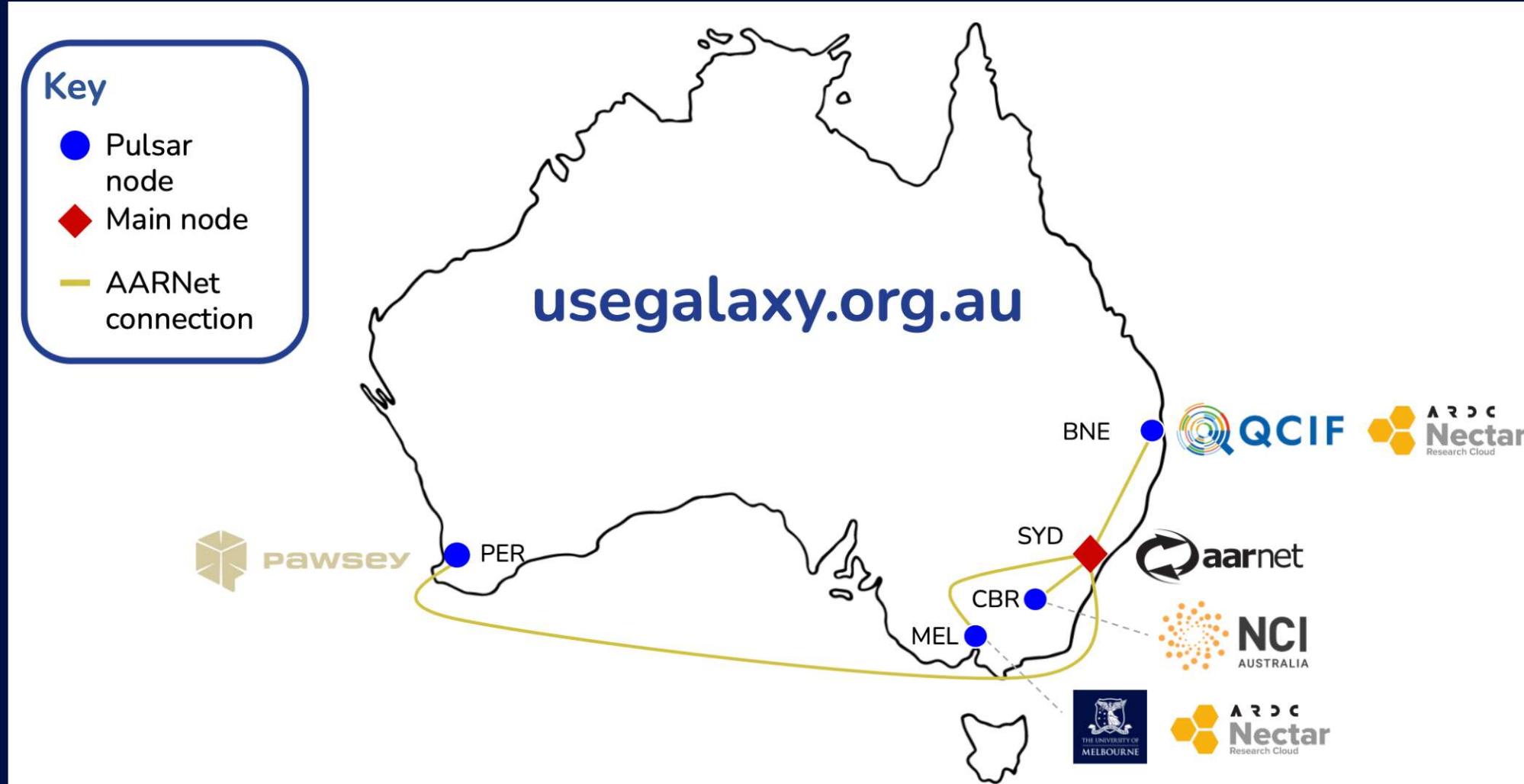
# UseGalaxy.\* Members 2025



# Galaxy Australia Deployment (2024)



THE UNIVERSITY OF  
MELBOURNE



# Benefits of Galaxy Australia



- 600 GB of storage if registered with an institutional/university email (can update if moving to new institution)
- 100 GB of storage for international users
- Support (help desk - [help@genome.edu.au](mailto:help@genome.edu.au))
- Access to high performance computing, cluster resources without requiring system account and security credentials
- Accessible on desktop, laptop, tablet (and even smartphone)
- 10,000+ available analysis tools and 100's of precomputed reference genomes available

# Accessing Specialised Tools



THE UNIVERSITY OF  
MELBOURNE

The screenshot shows the Galaxy Australia web interface. On the left, a sidebar lists various tools and data types. The main content area is titled "Request access to specialist tools". It contains a brief description of the access process, a link to contact support, and four highlighted tool applications: AlphaFold 2, FGENESH++, Cell Ranger, and DiaNN. A red box highlights the first two applications. On the right, a history panel shows several datasets, including FASTQ files and QC reports.

Can request access

Request access to specialist tools

Go to support menu

Galaxy Australia users from Australian institutions can apply for access to specialised tools. Access to these tools is restricted due to licensing and resource constraints. Access will usually be granted automatically if your account email address is on our list of institutions.

Please [contact us](#) if you have any further queries about accessing these tools.

Predict protein structures with AlphaFold 2

Annotate genomes with FGENESH++

Perform single cell RNA-seq analyses with Cell Ranger

Analyse data-independent acquisition (DIA) proteomics data with DiaNN

Apply now

Apply now

Apply now

Apply now

History

Kombucha ITS - Test for QC

526 MB 12 163 152

247: SRR28912966\_1.fastq.gz

246: SRR28912963\_2.fastq.gz

237: MultiQC-web.html

236: MultiQC-stats

a list with 3 tabular datasets

219: FastQC-files-raw.txt

a list with 8 txt datasets

218: FastQC-files-web.html

a list with 8 html datasets

191: MultiQC-web.html

Predict protein structures with AlphaFold 2

Annotate genomes with FGENESH++

Perform single cell RNA-seq analyses with Cell Ranger

Analyse data-independent acquisition (DIA) proteomics data with DiaNN

Apply now

Apply now

Apply now

Apply now

# Sharing

- Galaxy Histories and Workflows can be exported and downloaded as files
  - Can be shared and uploaded to any Galaxy server (public or local)
- Galaxy Histories and Workflows can also be shared within Galaxy Australia
  - Share with specific Galaxy Australia users
  - Share as a link
  - Publish to make available to all users
- Galaxy Workflows can also be published to [workflowhub.eu](https://workflowhub.eu) and [dockstore.org](https://dockstore.org)
  - Can import published Workflows directly into Galaxy Australia



# The Galaxy Training Network



THE UNIVERSITY OF  
MELBOURNE

Galaxy Training! Contributors Learning Pathways Help Settings Search Tutorials

## Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

### Galaxy for Scientists

We have separated the tutorials into several categories based on field and technology. We are exploring other ways to organise the tutorials going forward!

### Start Here

Topic	Tutorials
Introduction to Galaxy Analyses	13
Using Galaxy and Managing your Data	22

**Not sure where to start?**

Try the NGS Basics Learning Path! [Start Learning](#)

### Scientific Fields

Topic	Tutorials
Galaxy for Developers	1
Galaxy for Teachers	1

### Quickstart

Learning Pathways	Galaxy for SysAdmins

Galaxy for Developers	Galaxy for Teachers

### Upcoming Events

Check out upcoming events around the Galaxy!

January 28, 2025  
Galaxy at SURF Research Cloud workshop

[training.galaxyproject.org/](http://training.galaxyproject.org/)

Scientific Fields	
Topic	Tutorials
Climate	12
Computational chemistry	9
SARS-CoV-2	9
Foundations of Data Science	49
Ecology	23
Evolution	9
FAIR Data, Workflows, and Research	23
Genome Annotation	20
Imaging	9
Materials Science	1
Microbiome	23
One Health	9
Plants	9
Statistics and machine learning	20
Visualisation	5

Methodologies	
Topic	Tutorials
Assembly	19
Epigenetics	10
GMOD	14
Metabolomics	9
Proteomics	32
Sequence analysis	8
Single Cell	35
Synthetic Biology	3
Transcriptomics	24
Variant Analysis	17

# GTN Tutorial Resources



THE UNIVERSITY OF  
MELBOURNE

## A short introduction to Galaxy

Author(s) Anna Syme Nicola Soranzo  
Editor(s) Bérénice Batut Ahmed Hamid Awan Phil Reed  
Reviewers

### Overview

#### Questions:

- How to get started in Galaxy

#### Objectives:

- Learn how to upload a file
- Learn how to use a tool
- Learn how to view results
- Learn how to view histories
- Learn how to extract and run a workflow
- Learn how to share a history

#### Time estimation:

40 minutes

#### Level:

Introductory

#### Supporting Materials:

Slides Datasets Workflows

FAQs

Recordings

Available on these Galaxies

#### Published:

Aug 27, 2018

#### Last modification:

Mar 31, 2025

License: Tutorial Content is licensed under [Creative Commons Attribution 4.0 International License](#). The GTN Framework is licensed under [MIT](#)

PURL: <https://gxy.io/GTN:T00190>

Rating: 4.7 (104 recent ratings, 667 all time)

Revision: 43

## Tutorial Recording - 18 March 2025



GTN Tutorial: A short introduction to galaxy 20...

Copy link

## GTN Video Tutorial



Watch on [training.galaxyproject.org](#)

### About this Recording

Speaker Tristan Reynolds

Captioner Tristan Reynolds

Length 27 minutes

Created 18 March 2025

Archive [View tutorial at time of recording](#)

License CC-BY

### Supporting Materials:

Slides

Datasets

Workflows

FAQs

Recordings

Available on these Galaxies

# Commonly Encountered Issues

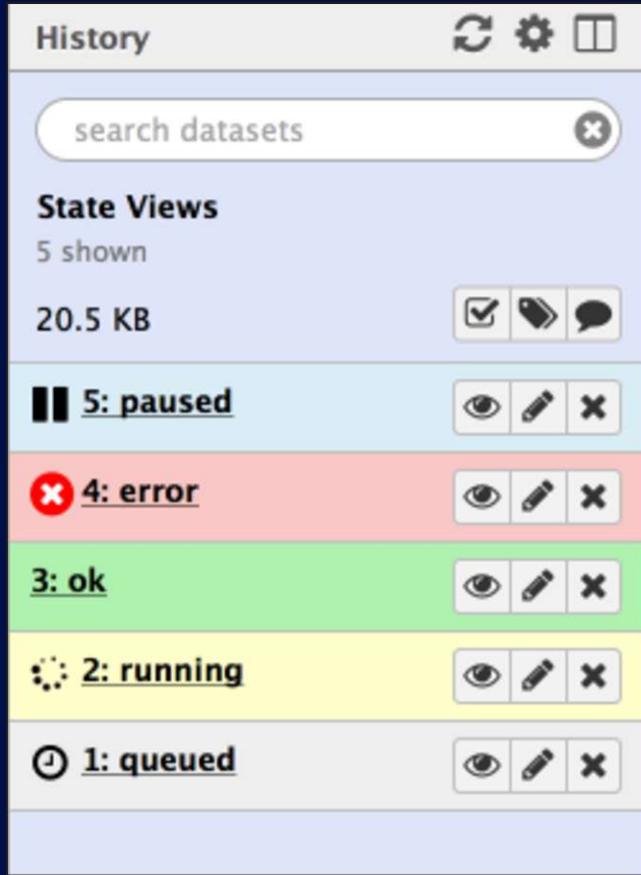
- Registration confirmation/account activation email gets captured by institutional spam filters
  - Contact Galaxy support if account activation email not received - [help@genome.edu.au](mailto:help@genome.edu.au)
- Make sure to register for Galaxy Australia ([usegalaxy.org.au](http://usegalaxy.org.au))
  - Limited support if registered with North American or European Galaxy servers
  - Use Galaxy Australia to insure that all data remains within Australia (some users may have restrictions on data sharing/transfer outside of Australia)
- Galaxy lowers, but does not remove the learning curve for computational data analysis
  - Contact Galaxy support with any questions or to request assistance - [help@genome.edu.au](mailto:help@genome.edu.au)
  - Take advantage of the library of training material - [Galaxy Training Network](#)
  - Jobs will fail



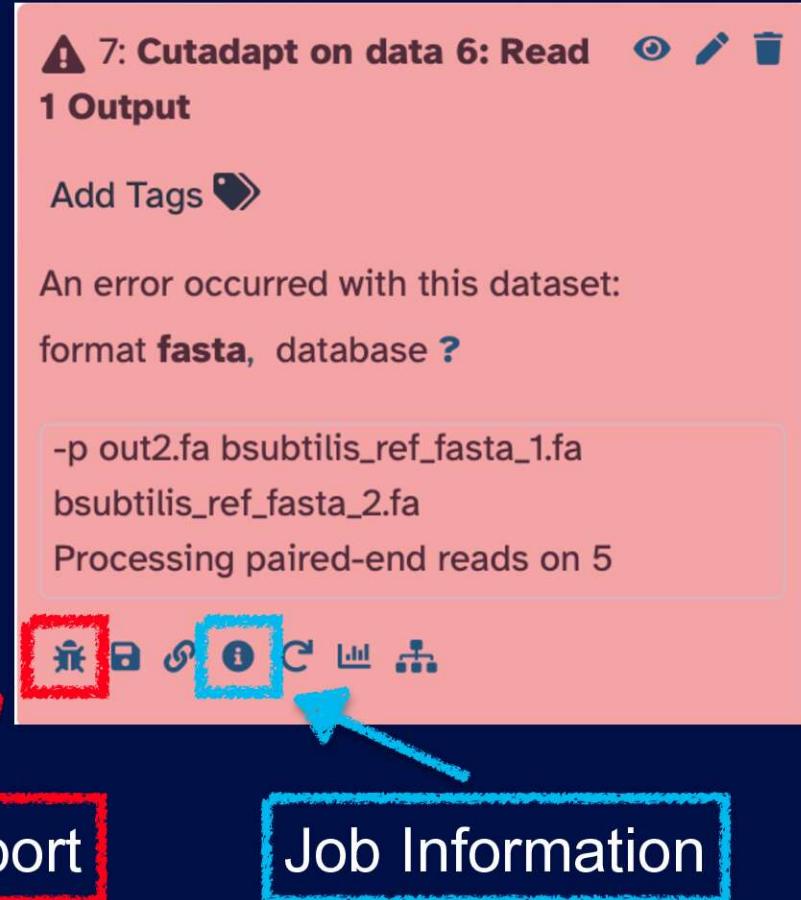
Galaxy Training!

# Troubleshooting

## Job Status



The screenshot shows the Galaxy History interface with a search bar and a sidebar titled "State Views". It lists five categories: "20.5 KB", "5: paused", "4: error", "3: ok", "2: running", and "1: queued". Each category has a set of three icons: a checkmark, a tag, and a speech bubble.



The screenshot shows the Galaxy Job Details interface for a dataset with ID 7. The top section displays an error message: "7: Cutadapt on data 6: Read" with a warning icon, followed by "1 Output" and "Add Tags". Below this, it says "An error occurred with this dataset: format **fasta**, database ?". The command line shown is "-p out2.fa bsubutilis\_ref\_fasta\_1.fa bsubutilis\_ref\_fasta\_2.fa". It also mentions "Processing paired-end reads on 5". At the bottom, there is a toolbar with several icons, one of which is highlighted with a blue border. Two arrows point from labels below the interface to these icons: a red arrow points to the icon labeled "Error Report" (highlighted in red), and a blue arrow points to the icon labeled "Job Information" (highlighted in blue).

# Investigating Job Failures



THE UNIVERSITY OF  
MELBOURNE

## Dataset Error Report

An error occurred while running the tool  
[toolshed.g2.bx.psu.edu/repos/lparsons/cutadapt/cutadapt/4.9+galaxy1](http://toolshed.g2.bx.psu.edu/repos/lparsons/cutadapt/cutadapt/4.9+galaxy1).

### Details

Execution resulted in the following messages:

Fatal error: Exit code 1 ()

Tool generated the following standard error:

```
Traceback (most recent call last):
  File "/usr/local/lib/python3.12/site-packages/cutadapt/runners.py", line 105, in run
    for index, chunks in enumerate(self._read_chunks(*files)):
      ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "/usr/local/lib/python3.12/site-packages/cutadapt/runners.py", line 122, in _read_chunks
    for chunks in dnaio.read_paired_chunks(
      ^^^^^^^^^^^^^^
  File "/usr/local/lib/python3.12/site-packages/dnaio/chunks.py", line 222, in read_paired_chunks
    raise ValueError()
ValueError: FASTA/FASTQ records do not fit into buffer of size 4000000

Traceback (most recent call last):
  File "/usr/local/lib/python3.12/site-packages/cutadapt/runners.py", line 105, in run
    for index, chunks in enumerate(self._read_chunks(*files)):
      ^^^^^^^^^^^^^^
```

⚠ 7: Cutadapt on data 6: Read 1 Output

Add Tags ↗

An error occurred with this dataset:  
format **fasta**, database ?

```
-p out2.fa bsubutilis_ref.fasta_1.fa
bsubutilis_ref.fasta_2.fa
Processing paired-end reads on 5 cores
```

✖️ 📁 ⚙️ ⓘ ⚡

## Job Outputs

Tool Outputs	Dataset
Cutadapt on : Read 1 Output	⚠ 7: Cutadapt on data 6: Read 1 Output
Cutadapt on : Read 2 Output	⚠ 8: Cutadapt on data 6: Read 2 Output

### Job Information

Galaxy Tool ID: [toolshed.g2.bx.psu.edu/repos/lparsons/cutadapt/cutadapt/4.9+galaxy1](#)

Job State: error

Command Line: `ln -f -s '/mnt/user-data-volA/data22/8/d/2/dataset_8d26eaea-48f1-47be-b...'`

Tool Standard Output:

```
This is cutadapt 4.9 with Python 3.12.7
Command line parameters: -j=5 -a AACG --error-rate=0.1 --times=1 --
overlap=3 --action=trim --minimum-length=1 -o out1.fa -p out2.fa
bsubutilis_ref.fasta_1.fa bsubutilis_ref.fasta_2.fa
Processing paired-end reads on 5 cores ...
```

Tool Standard Error:

```
Traceback (most recent call last): File "/usr/local/lib/python3.12/site...
```

Tool Exit Code: 1

- desc: Fatal error: Exit code 1 ()
- error\_level: 3
- exit\_code: 1
- type: exit\_code

Job Messages:

Job API ID: a6e389a98c2d16788d15b347e415df1a

# Galaxy Australia Team



AUSTRALIA



Gareth  
Price



Catherine  
Bromhead



Justin  
Lee



Nuwan  
Goonasekera



Michael  
D'Silva



Igor  
Makunin



Michael  
Thang



Cameron  
Hyde



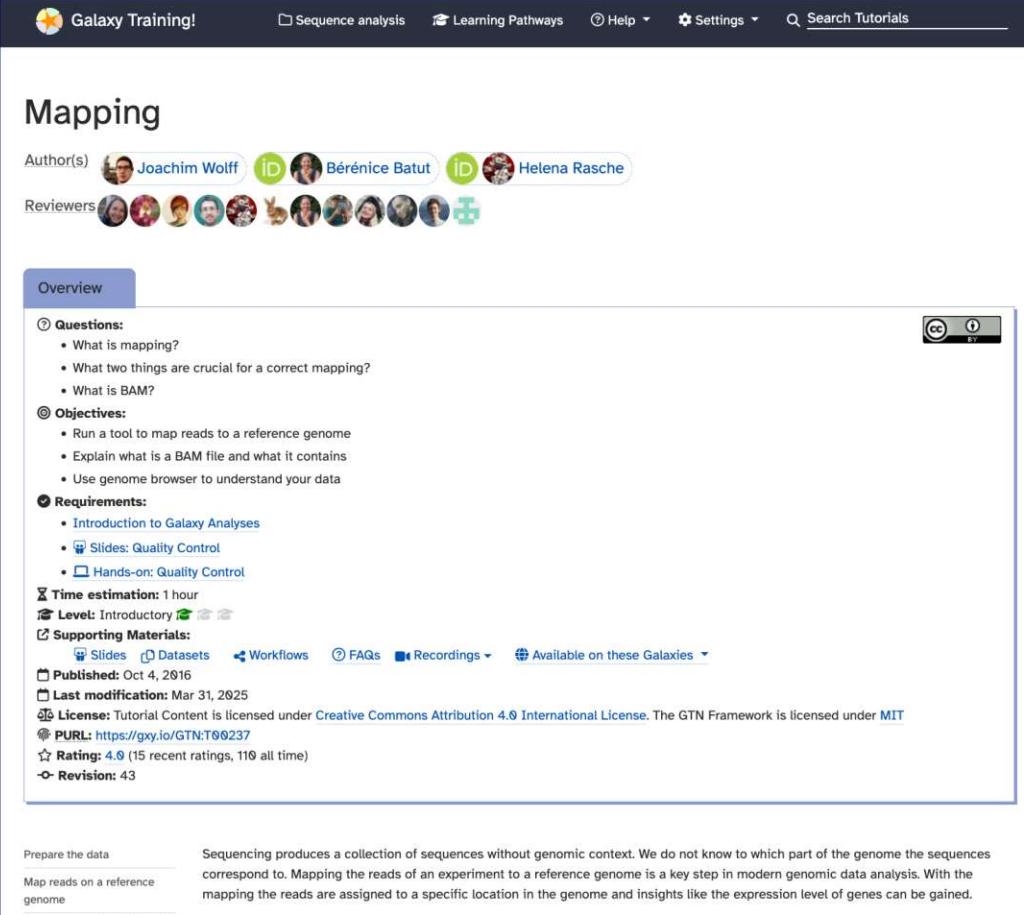
Tristan  
Reynolds



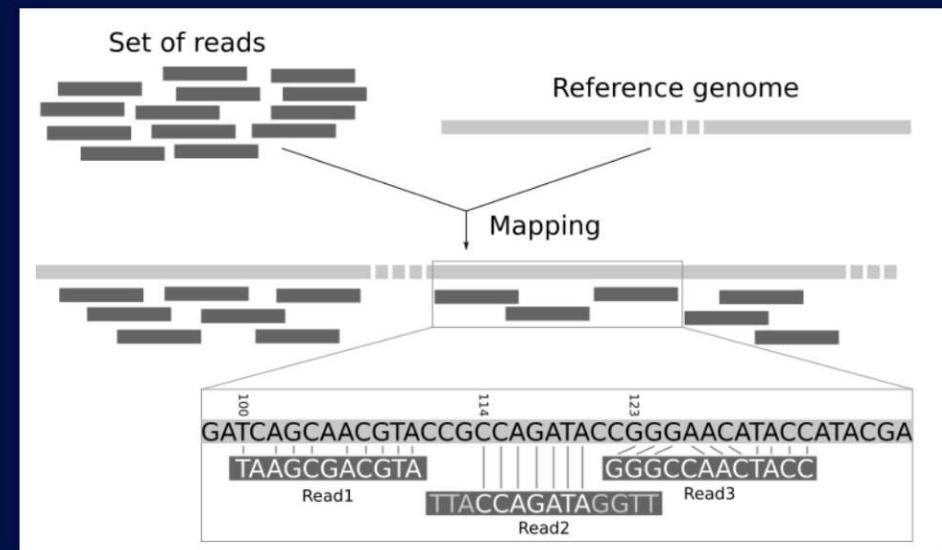
Anna  
Syme



# Hands On - Mapping



The screenshot shows the 'Mapping' tutorial page on the Galaxy Training! website. The top navigation bar includes links for Galaxy Training!, Sequence analysis, Learning Pathways, Help, Settings, and Search Tutorials. Below the title 'Mapping', it lists 'Author(s)' Joachim Wolff, Bérénice Batut, and Helena Rasche, and 'Reviewers'. The main content area is titled 'Overview' and contains sections for 'Questions', 'Objectives', 'Requirements', 'Time estimation', 'Level', 'Supporting Materials', 'Published', 'Last modification', 'License', 'PURL', 'Rating', and 'Revision'. A note at the bottom explains that sequencing produces a collection of sequences without genomic context, and mapping is the process of assigning these sequences to specific locations in a reference genome to gain insights into gene expression.

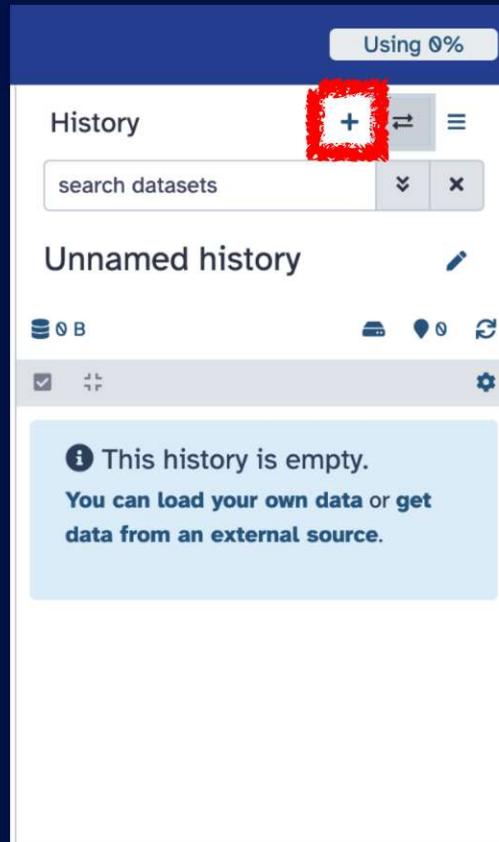


<https://training.galaxyproject.org/training-material/topics/sequence-analysis/tutorials/mapping/tutorial.html>

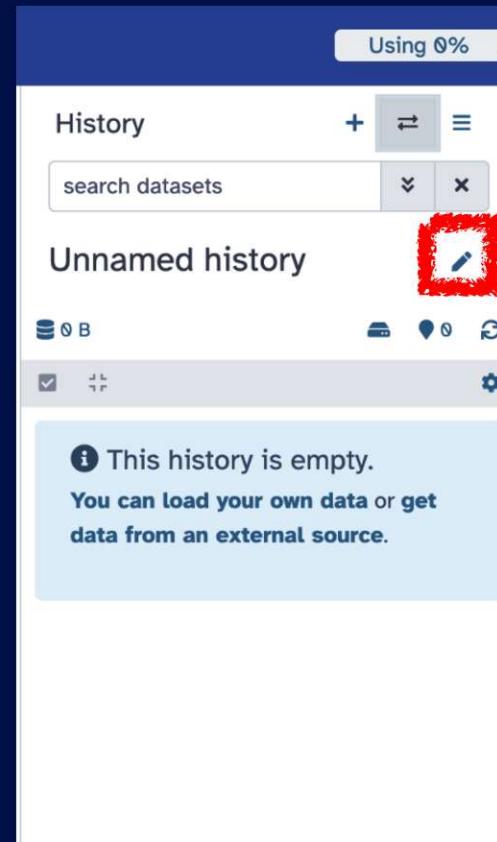
# Reference Slides

# History Functionality

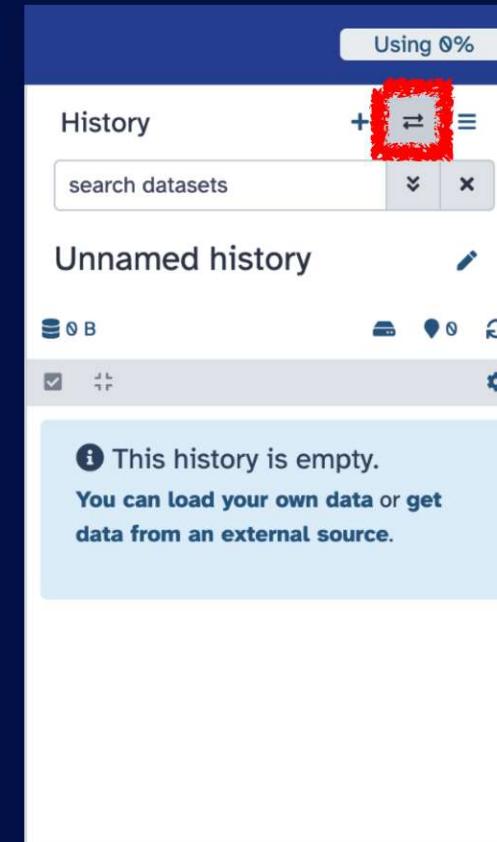
## Create New



## Rename



## Switch



# History Functionality



THE UNIVERSITY OF  
MELBOURNE

The screenshot shows the Galaxy Australia web interface. On the left, a sidebar lists various tools and data management options. A modal window titled "Switch to history" is open, displaying a list of histories. The histories listed are:

- Galaxy Basics for everyone - Iris Species (Current) (17 items, 7 days ago)
- Collection Testing (328 items, about 21 hours ago)
- Unnamed history (37 items, 5 days ago)
- Testing Cancelled Job Message (21 items, 14 days ago)
- RNA-seq counts to genes (17 items, 21 days ago)

At the bottom of the modal, it says "Loaded 10 out of 14 histories" and "Click a history to switch to it".

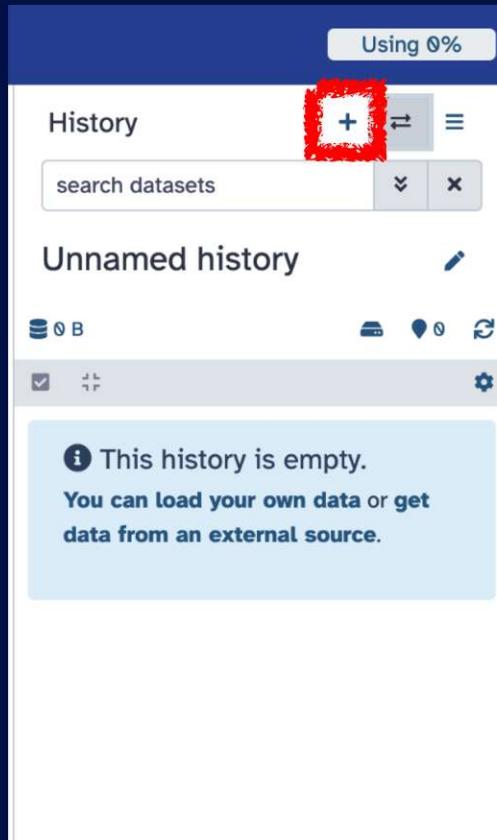
In the main workspace, a history titled "Galaxy Basics for everyone - Iris Species" is visible, containing 381 kB of data. The history list includes:

- 16: iris petal scatterplot: pdf #iris
- 15: iris petal scatterplot: png #iris
- 10: iris sepal scatterplot: pdf #iris
- 9: iris sepal scatterplot: png #iris
- 8: iris summary and statistics #iris
- 7: iris samples per species group #iris

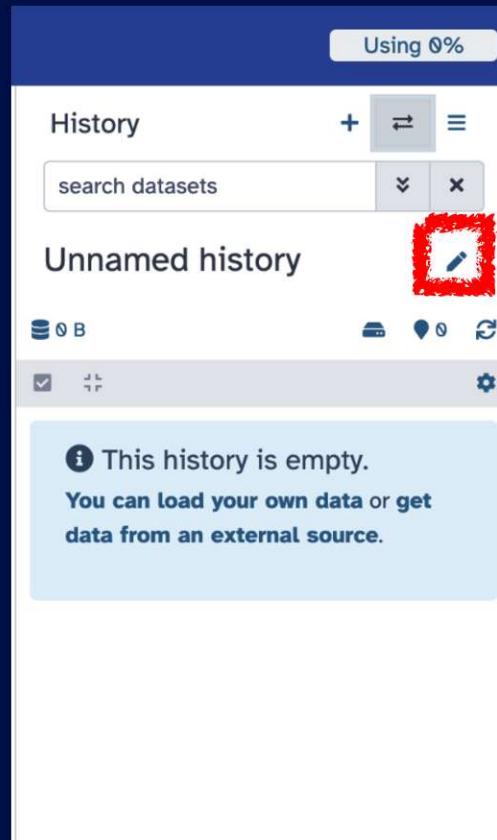
At the bottom of the workspace, a footer message reads: "Galaxy Australia is an open, web-based platform for accessible, reproducible and transparent computational research. Galaxy supports thousands of".

# History Functionality

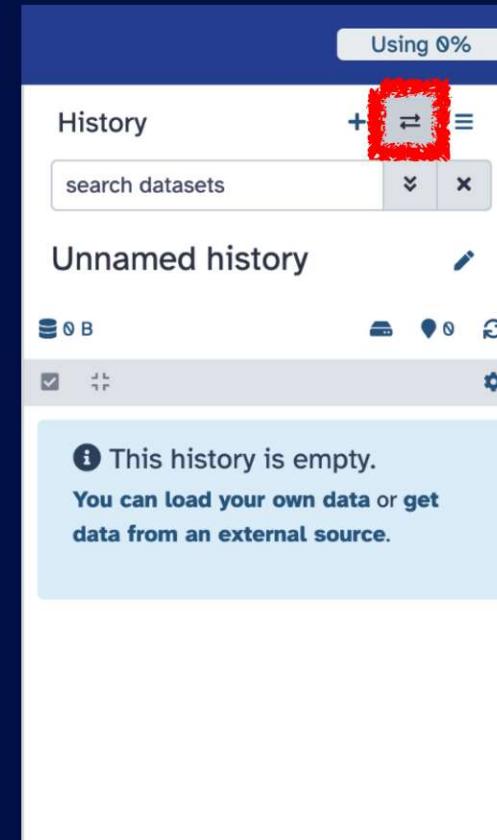
## Create New



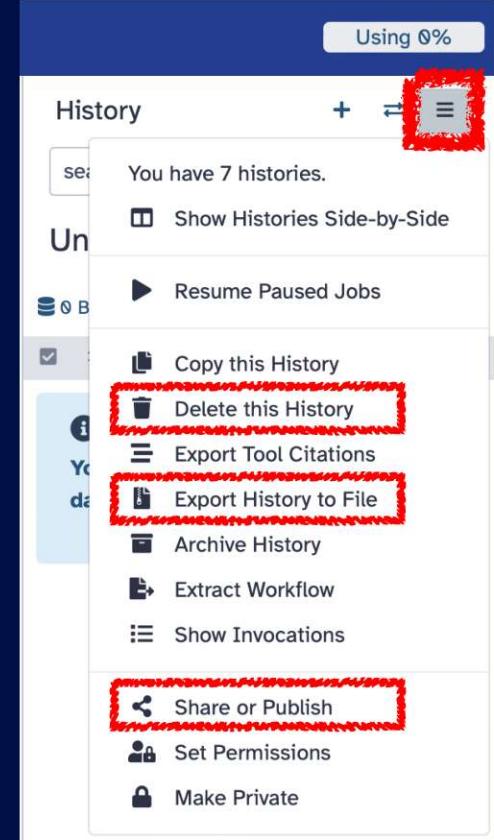
## Rename



## Switch



## Options



# Uploading Data



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Using 1% of 600.0 GB treynolds

Upload

Tools

search tools

FILE AND META TOOLS

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

History

search datasets

Unnamed history

0 B

This history is empty.  
You can load your own data or get data from an external source.

Galaxy Australia has been upgraded to release 24.1. Che...

Home News People About Support Docs

Galaxy AUSTRALIA

Galaxy platform 2024 update published!

Read the latest developments supporting accessible, reproducible, and collaborative data analyses

With contributions from 130 authors representing 60 institutions

doi.org/10.1093/nar/gkae410

usegalaxy.\*

Galaxy Australia is an **open, web-based** platform for accessible, reproducible and transparent computational research. Galaxy supports thousands of documented and maintained tools that are free to use. We facilitate on-demand training capacities and provision **600GB** for Australian institutional (and 100GB for other) users.

# Uploading Data



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Upload

Tools

search tools

FILE AND MEDIA

Get Data

Send Data

Collection

GENERAL TEXT

Text Manipulation

Filter and Sort

Join, Subtract

GENOMIC FILES

FASTA/FASTQ

FASTQ Quality

SAM/BAM

BED

VCF/BCF

Nanopore

Convert Formats

Lift-Over

COMMON GENOMICS TOOLS

Upload from Disk or Web to Galaxy Basics for everyone - Iris Species

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

New File 299 b Auto-detect unspecified (?) 0%

Download data from the web by entering URLs (one per line) or directly paste content.

`https://zenodo.org/record/6457007/files/GSM461177_1_subsampled.fastqsanger  
https://zenodo.org/record/6457007/files/GSM461177_2_subsampled.fastqsanger  
https://zenodo.org/record/6457007/files/GSM461180_1_subsampled.fastqsanger  
https://zenodo.org/record/6457007/files/GSM461180_2_subsampled.fastqsanger`

Type (set all): Auto-detect Reference (set all): unspecified (?)    Paste/Fetch data

Galaxy Australia is an open, web-based platform for accessible, reproducible and transparent computational research. Galaxy supports thousands of

Basics for everyone - Iris Species

catterplot: pdf

catterplot: png

catterplot: pdf

catterplot: png

#iris

7: iris samples per species group

#iris

# Uploaded Datasets in the Current History



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Using 1% of 600.0 GB treynolds

Upload Tools search tools FILE AND META TOOLS Get Data Send Data Collection Operations GENERAL TEXT TOOLS Text Manipulation Filter and Sort Join, Subtract and Group GENOMIC FILE MANIPULATION FASTA/FASTQ FASTQ Quality Control SAM/BAM BED VCF/BCF Nanopore

Galaxy Australia has been upgraded to release 24.1. Che...

Home News People About Support Docs

Galaxy AUSTRALIA

Galaxy platform 2024 update published!

Read the latest developments supporting accessible, reproducible, and collaborative data analyses

With contributions from 130 authors representing 60 institutions

doi.org/10.1093/nar/gkae410

usegalaxy.\*

History

search datasets

training\_test

618 MB

8: GSM461180\_2\_subsampled.fa  
stqsanger

7: GSM461180\_1\_subsampled.fa  
stqsanger

6: GSM461177\_2\_subsampled.fa  
stqsanger

5: GSM461177\_1\_subsampled.fa  
stqsanger

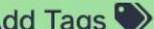
The dataset entries from 8 to 5 are highlighted with a red border.

# Dataset Options

## View data

5: GSM461177\_1\_subsample 

d.fastqsanger

Add Tags 

136.2 MB

format fastqsanger, database ?

uploaded fastqsanger file



```
@SRR031714.4 HWI-EAS299_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACGTCCGTGATCGTGTGCCAG  
+  
IIIIIIIIIIIIIIIIII$II$II(I8I;II63  
@SRR031714.6 HWI-EAS299_130MNEAAXX:2:1:1747:
```

This dataset is large and only the first megabyte is shown below.  
[Show all](#) | [Save](#)

```
@SRR031714.4 HWI-EAS299_130MNEAAXX:2:1:1729:593/1  
GCCGGGAAGAGGATACGTCCGTGATCGTGTGCCAG  
+  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII  
@SRR031714.6 HWI-EAS299_130MNEAAXX:2:1:1747:596/1  
GTGTAAGGCACTAAAGTATATCTCCTTTAACATCTC  
+  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII  
@SRR031714.8 HWI-EAS299_130MNEAAXX:2:1:1758:603/1  
GTGATATTAAACGTGATATGAGCGAAAACAAGTCGAA  
+  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII  
@SRR031714.29 HWI-EAS299_130MNEAAXX:2:1:1665:1684/1  
ATCGCGAAGTTCTTCATAACGCTCACGTTCTATTG  
+  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII  
@SRR031714.33 HWI-EAS299_130MNEAAXX:2:1:1559:1932/1  
AAGACGTGCATGCTGGAAAAATCATGTGCAACCAT  
+  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII  
@SRR031714.36 HWI-EAS299_130MNEAAXX:2:1:1721:567/1  
GCAAGAAGAAGGCATTACCAAGGCCAGCAAGAACGT  
+  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII  
@SRR031714.56 HWI-EAS299_130MNEAAXX:2:1:1255:1045/1  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
+  
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII  
@SRR031714.70 HWI-EAS299_130MNEAAXX:2:1:1618:605/1  
GCTTCCCTCTGTATGGTGTCTCCCTGAAGTTCTCAGA
```

History 

deleted:any visible:any 

training\_test 

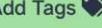
1.83 GB  30  6  58 



d.fastqsanger

5: GSM461177\_1\_subsample 

d.fastqsanger

Add Tags 

136.2 MB

format fastqsanger, database ?

uploaded fastqsanger file



```
@SRR031714.4 HWI-EAS299_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACGTCCGTGATCGTGTGCCAG  
+  
IIIIIIIIIIIIIIIIII$II$II(I8I;II63  
@SRR031714.6 HWI-EAS299_130MNEAAXX:2:1:1747:
```

 4: GSM461180\_2.fastqsanger 

 3: GSM461180\_1.fastqsanger 

# Dataset Options



THE UNIVERSITY OF  
MELBOURNE

## Add Tags

5: GSM461177\_1\_subsample

d.fastqsanger

**Add Tags**

136.2 MB

format **fastqsanger**, database ?

uploaded fastqsanger file

```
@SRR031714.4 HWI-EAS299_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACTCCGTCGATCGTGTGCCAG  
+  
IIIIIIIIIIIIIIIIIIII$II(I8I;II63  
@SRR031714.6 HWI-EAS299_130MNEAAXX:2:1:1747:
```

## Select Existing Tags or Create New

5: GSM461177\_1\_subsample

d.fastqsanger

**Add Tags**

#IWC  
#PathoGFAIR  
#microGalaxy  
#Collection  
GUCFG2Galaxy  
shotgun  
metagenomics  
transcriptomics

## Applied Tags

5: GSM461177\_1\_subsample

d.fastqsanger

**#sample1** **reads**

Add Tags

136.2 MB

format **fastqsanger**, database ?

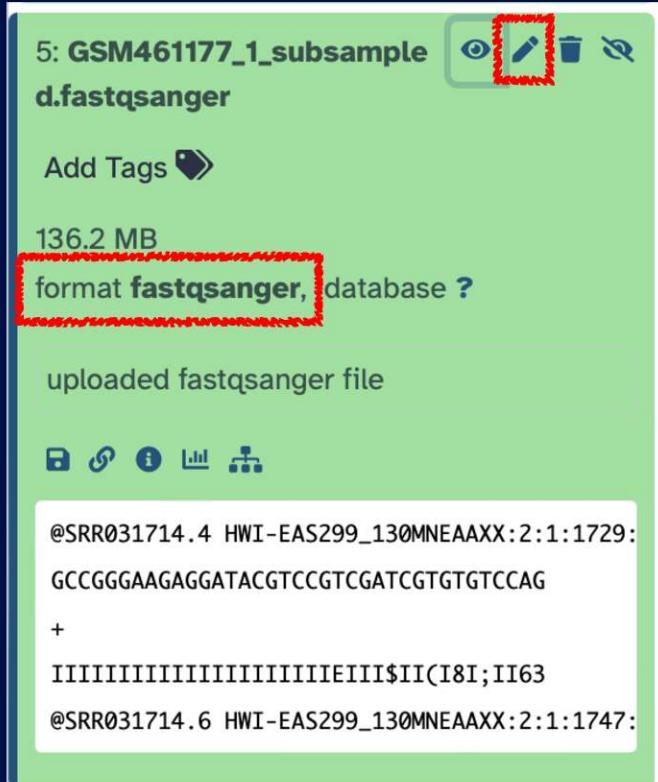
uploaded fastqsanger file

```
@SRR031714.4 HWI-EAS299_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACTCCGTCGATCGTGTGCCAG  
+  
IIIIIIIIIIIIIIIIIIII$II(I8I;II63  
@SRR031714.6 HWI-EAS299_130MNEAAXX:2:1:1747:
```

NOTE: #tags will be applied to output when tagged input data is used with a tool

# Dataset Options

## Edit Attributes



5: GSM461177\_1\_subsample d.fastqsanger

Add Tags

136.2 MB

format fastqsanger, database ?

uploaded fastqsanger file

Sequence data preview:

```
@SRR031714.4 HWI-EAS299_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACGTCCGTCGATCGTGTGTCCAG  
+  
IIIIIIIIIIIIIIIIIIII$II(I8I;II63  
@SRR031714.6 HWI-EAS299_130MNEAAXX:2:1:1747:
```

### Edit Dataset Attributes

Attributes Datatypes Permissions

Name: GSM461177\_1\_subsampled.fastqsanger

Info: uploaded fastqsanger file

Annotation: optional

Database/Build: optional

Save Auto-detect

### Edit Dataset Attributes

Attributes Datatypes Permissions

#### Assign Datatype

New Type: fastqsanger

This will change the datatype of the existing dataset but not modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

Save Auto-detect

#### Convert to Datatype

Target datatype: fqtoc (using 'Convert FASTQ files to seek locations')

This will create a new dataset with the contents of this dataset converted to a new format.

Create Dataset

# Dataset Options



THE UNIVERSITY OF  
MELBOURNE

## Information

5: GSM461177\_1\_subsample

d.fastqsanger

Add Tags

136.2 MB

format fastqsanger, database ?

uploaded fastqsanger file

@SRR031714.4 HWI-EAS299\_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACTCGTCGATCGTGTGCCAG  
+  
IIIIIIIIIIIIIIIIIIII\$II(I8I;II63  
@SRR031714.6 HWI-EAS299\_130MNEAAXX:2:1:1747:

Dataset Information

Number	5
Name	GSM461177_1_subsampled.fastqsanger
Created	Saturday Sep 7th 14:13:05 2024 GMT+10
Filesize	<b>136.2 MB</b>
Dbkey	?
Format	fastqsanger
File contents	contents
History Content API ID	a6e389a98c2d16788bfd17e9d980285a
History API ID	1700b9ee8cfca6a6
UUID	266c6bdc-44ea-480c-ad7d-c907fdf9dc69
Full Path	/mnt/user-data-volA/data22/2/6/6/dataset_266c6bdc-44ea-480c-ad7d-c907fdf9dc69.dat
Sources	<ul style="list-style-type: none"><li><a href="https://zenodo.org/record/6457007/files/GSM461177_1_subsampled.fastqsanger">https://zenodo.org/record/6457007/files/GSM461177_1_subsampled.fastqsanger</a> </li></ul>

Tool Parameters

Input Parameter	Value
request_version	1

```
{"targets": [{"destination": {"type": "hdas"}, "elements": [{"name": null, "dbkey": "?", "ext": "auto", "space_to_tab": false, "to_posix_lines": true, "deferred": false, "src": "url", "url": "https://zenodo.org/record/6457007/files/GSM461177_1_subsampled.fastqsanger", "hashes": []}, {"name": null, "dbkey": "?", "ext": "auto", "space_to_tab": false, "to_posix_lines": true, "deferred": false, "src": "url", "url": "https://zenodo.org/record/6457007/files/GSM461177_1_subsampled.fastqsanger", "hashes": []}], "in_place": false, "purge_source": false, "object_id": 22597492}], "name": null, "dbkey": "?", "ext": "auto", "space_to_tab": false, "to_posix_lines": true, "deferred": false, "src": "url", "url": "https://zenodo.org/record/6457007/files/GSM461177_1_subsampled.fastqsanger", "hashes": []}
```

History

deleted:false visible:any	
---------------------------	--

training\_test

1.82 GB	
---------	--

d.fastqsanger

6: GSM461177\_2\_subsample

d.fastqsanger

5: GSM461177\_1\_subsample

d.fastqsanger

Add Tags

136.2 MB

format fastqsanger, database ?

uploaded fastqsanger file

@SRR031714.4 HWI-EAS299\_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACTCGTCGATCGTGTGCCAG  
+  
IIIIIIIIIIIIIIIIIIII\$II(I8I;II63  
@SRR031714.6 HWI-EAS299\_130MNEAAXX:2:1:1747:

# Dataset Options

## Download

5: GSM461177\_1\_subsample    

d.fastqsanger

Add Tags 

136.2 MB

format **fastqsanger**, database ?

uploaded fastqsanger file

```
@SRR031714.4 HWI-EAS299_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACTCCGTGATCGTGTGTCCAG  
+  
IIIIIIIIIIIIIIIIIIII$II(I8I;II63  
@SRR031714.6 HWI-EAS299_130MNEAAXX:2:1:1747:
```

## Delete

5: GSM461177\_1\_subsample    

d.fastqsanger

Add Tags 

136.2 MB

format **fastqsanger**, database ?

uploaded fastqsanger file

```
@SRR031714.4 HWI-EAS299_130MNEAAXX:2:1:1729:  
GCCGGGAAGAGGATACTCCGTGATCGTGTGTCCAG  
+  
IIIIIIIIIIIIIIIIIIII$II(I8I;II63  
@SRR031714.6 HWI-EAS299_130MNEAAXX:2:1:1747:
```

## Undelete

94: MultiQC on data 28, data 2   

6, and others: Stats

Add Tags 

5 lines 7 columns

format **tabular**, database ?

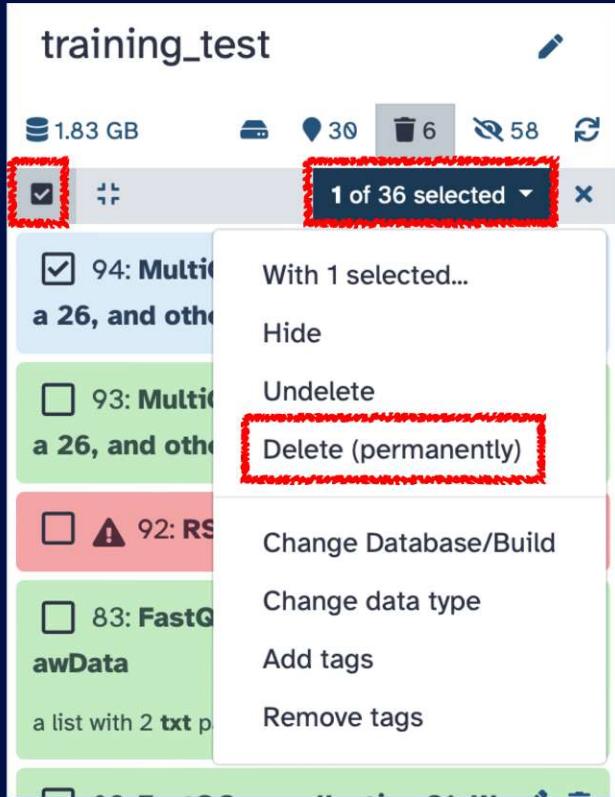
total 8  
-rw-r--r-- 1 galaxy galaxy 500 Jan 5 23:30  
fastqc-status-check-heatmap.txt

1	2
Sample	FastQC_mqc_gene
GSM461177_subsampled_forward	23.626646327645
GSM461177_subsampled_reverse	25.197947201769
GSM461180_subsampled_forward	23.763525491807
GSM461180_subsampled_reverse	8.1318023850958

# Dataset Options

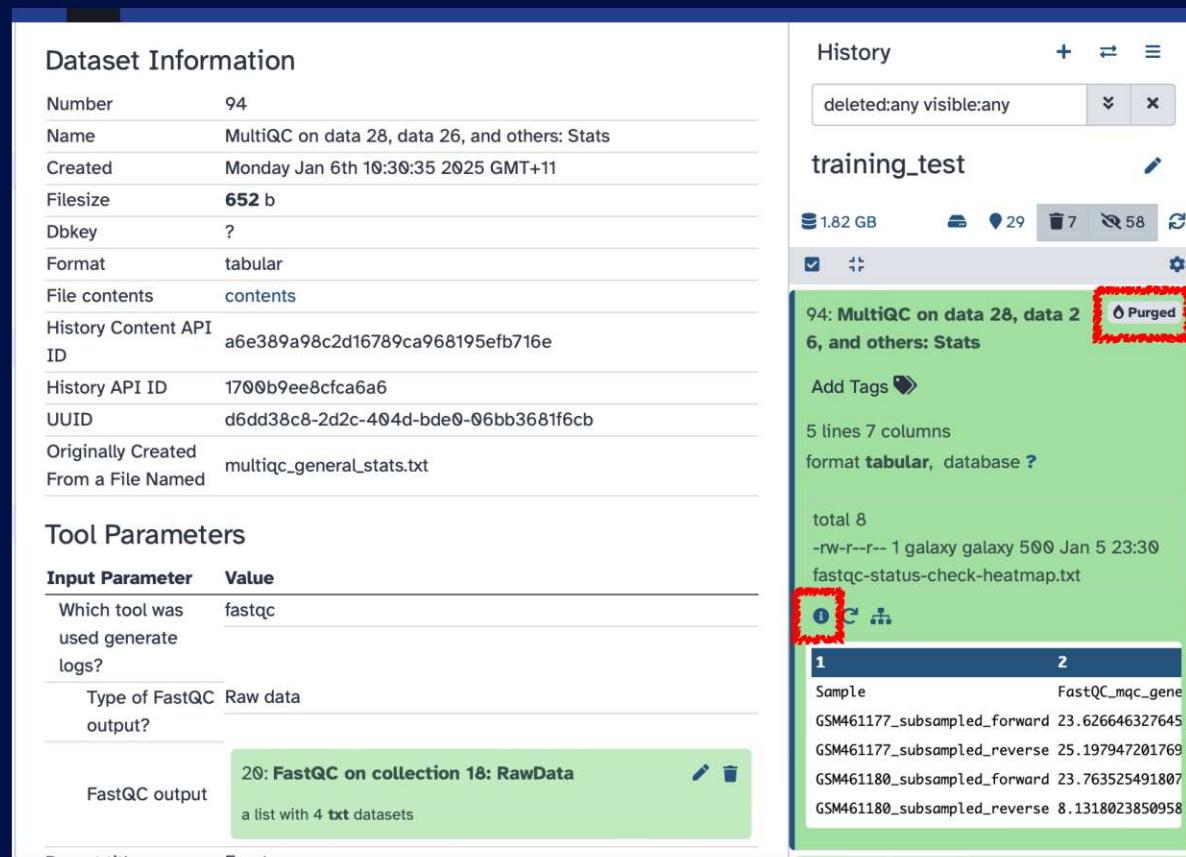
## Purge



A screenshot of the Galaxy interface showing a list of datasets under the collection 'training\_test'. A context menu is open over the first dataset, which has a checked checkbox. The menu items are: 'With 1 selected...', 'Hide', 'Undelete', 'Delete (permanently)', 'Change Database/Build', 'Change data type', 'Add tags', and 'Remove tags'. The 'Delete (permanently)' option is highlighted with a red box.

**IMPORTANT:** Purged datasets **CANNOT** be recovered!

Dataset information (e.g. tool/tool parameters run to produce output) are retained allowing for reproducing data as long original input data exists.



A screenshot of the Galaxy interface showing the 'Dataset Information' and 'Tool Parameters' sections for dataset number 94. The dataset details include: Name: MultiQC on data 28, data 26, and others: Stats; Created: Monday Jan 6th 10:30:35 2025 GMT+11; Filesize: 652 b; Dbkey: ?; Format: tabular; File contents: contents; History Content API ID: a6e389a98c2d16789ca968195efb716e; History API ID: 1700b9ee8fcfa6a6; UUID: d6dd38c8-2d2c-404d-bde0-06bb3681f6cb; Originally Created From a File Named: multiqc\_general\_stats.txt.

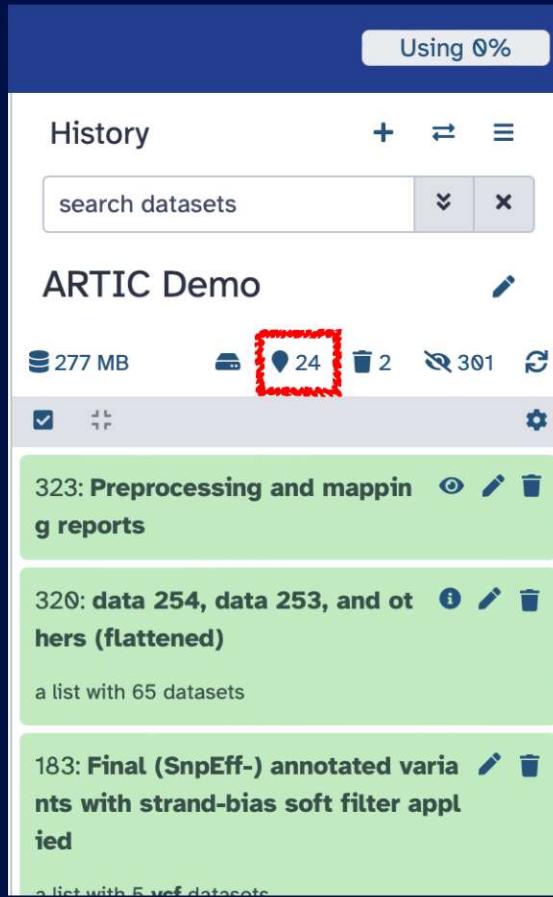
The 'Tool Parameters' section shows: Input Parameter: Which tool was used generate logs? Value: fastqc; Type of FastQC output? Raw data output?

The 'History' section shows the dataset has been purged, indicated by a red box around the 'Purged' status in the history entry.

The 'Dataset Information' section shows the dataset has been purged, indicated by a red box around the 'Purged' status in the dataset details.

# Dataset Visibility

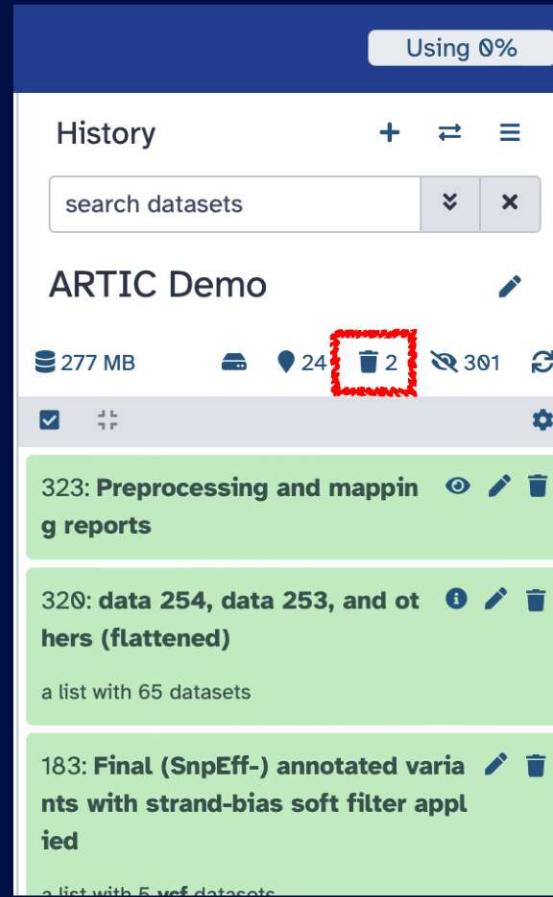
## Active



The screenshot shows the Galaxy interface with three datasets listed under the 'ARTIC Demo' project. The first dataset, '323: Preprocessing and mapping reports', has its visibility icon highlighted with a red box. The second dataset, '320: data 254, data 253, and others (flattened)', and the third dataset, '183: Final (SnpEff-) annotated variants with strand-bias soft filter applied', are shown below.

Dataset ID	Name	Size	Visibility	Annotations
323	Preprocessing and mapping reports	277 MB	Visible	24 datasets
320	data 254, data 253, and others (flattened)	277 MB	Visible	24 datasets
183	Final (SnpEff-) annotated variants with strand-bias soft filter applied	277 MB	Visible	301 datasets

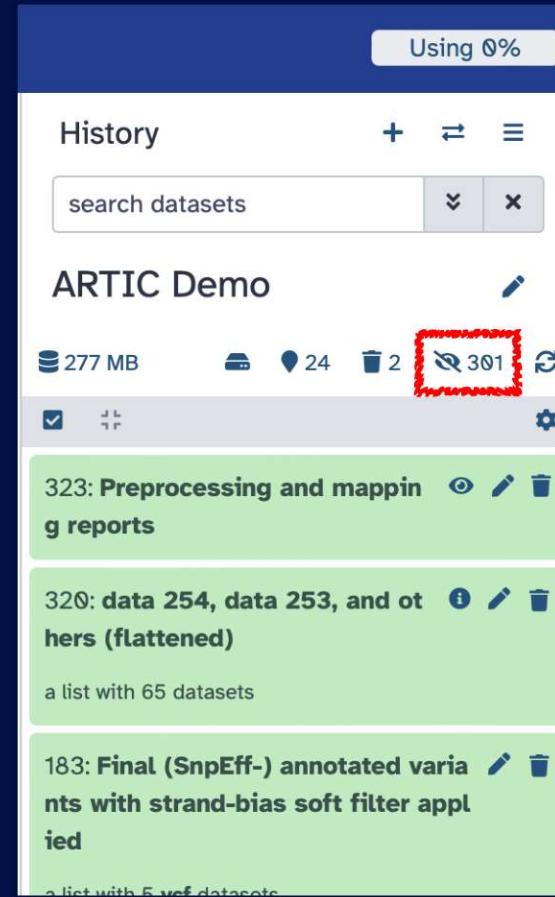
## Deleted



The screenshot shows the Galaxy interface with the same three datasets listed under the 'ARTIC Demo' project. The visibility icon for the second dataset ('320') is highlighted with a red box, indicating it is deleted.

Dataset ID	Name	Size	Visibility	Annotations
323	Preprocessing and mapping reports	277 MB	Visible	24 datasets
320	data 254, data 253, and others (flattened)	277 MB	Deleted	24 datasets
183	Final (SnpEff-) annotated variants with strand-bias soft filter applied	277 MB	Visible	301 datasets

## Hidden



The screenshot shows the Galaxy interface with the same three datasets listed under the 'ARTIC Demo' project. The visibility icon for the third dataset ('183') is highlighted with a red box, indicating it is hidden.

Dataset ID	Name	Size	Visibility	Annotations
323	Preprocessing and mapping reports	277 MB	Visible	24 datasets
320	data 254, data 253, and others (flattened)	277 MB	Visible	24 datasets
183	Final (SnpEff-) annotated variants with strand-bias soft filter applied	277 MB	Hidden	301 datasets

# Create Dataset Collection

Create a collection from a list of datasets

Collections of datasets are permanent, ordered lists of datasets that 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... [View details](#)

sample\_1\_reads.fq.gz [Discard](#)

sample\_2\_reads.fq.gz [Discard](#)

sample\_3\_reads.fq.gz [Discard](#)

sample\_4\_reads.fq.gz [Discard](#)

sample\_5\_reads.fq.gz [Discard](#)

[With 5 selected...](#)

[Unhide](#)

[Delete](#)

[Delete \(permanently\)](#)

[Build Dataset List](#) **Build Dataset Pair**

[Build List of Dataset Pairs](#)

[Build Collection from Rules](#)

[Change Database/Build](#)

[Change data type](#)

[Add tags](#)

[Remove tags](#)

Name:

[Create collection](#)

[Cancel](#)

History [+](#) [=](#) [≡](#)

deleted:false visible:any [▼](#) [X](#)

ARTIC Demo

277 MB [Download](#) 24 [Delete](#) 301 [Revert](#)

**5 of 325 selected** [▼](#) [X](#)

With 5 selected...

Unhide

Delete

Delete (permanently)

**Build Dataset List** **Build Dataset Pair**

Build List of Dataset Pairs

Build Collection from Rules

Change Database/Build

Change data type

Add tags

Remove tags

sample\_5\_reads.fq.gz [Edit](#) [Delete](#) [Revert](#)

sample\_5\_reads.fq.gz [Edit](#) [Delete](#) [Revert](#)

sample\_4\_reads.fq.gz [Edit](#) [Delete](#) [Revert](#)

sample\_3\_reads.fq.gz [Edit](#) [Delete](#) [Revert](#)

sample\_2\_reads.fq.gz [Edit](#) [Delete](#) [Revert](#)

sample\_1\_reads.fq.gz [Edit](#) [Delete](#) [Revert](#)

- **Collection:** a group datasets with the same datatype
- Using collections with Galaxy tools:
  - Using a collection as a tool's input submits separate Galaxy jobs to run the tool on each dataset and produces a collection of the individual job outputs

# Finding a Tool



THE UNIVERSITY OF  
MELBOURNE

Galaxy Australia

Using 1% of 600.0 GB treynolds

Tools

fastqc

FastQC Read Quality reports (Galaxy Version 0.74+galaxy1)

Run Tool

Tool Parameters

Raw read data from your current history \*

18: data 12, data 11, and others (flattened)

accepted formats

This is a batch mode input field. Individual jobs will be triggered for each dataset.

Contaminant list - optional

Nothing selected

accepted formats

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA

Adapter list - optional

Nothing selected

accepted formats

List of adapters adapter sequences which will be explicitly searched against the library. It should be a tab-delimited file with 2 columns: name and sequence. (--adapters)

Submodule and Limit specifying file - optional

Nothing selected

History

search datasets

training\_test

618 MB

18: data 12, data 11, and others (flattened)

a list with 4 datasets

13: 2 PE fastqs

a list with 2 pairs