

Building Galaxy Labs to advance life science research

Anna Syme, Madeline Bassetti, Winnie Mok, Cameron Hyde, Nigel Ward, **Gareth Price**

Galaxy! What is it?

- <https://usegalaxy.org.au/>
- New milestone: >7 million jobs submitted
- Current average 1200 active users per month
- Used by researchers
- Used in training
- Free to use
- Funded and supported by various organisations
(including Australian BioCommons, NCRIS, Bioplatforms Australia, ARDC, UniMelb, UQ, AARNet, Pawsey and others)

usegalaxy.org.au

Galaxy Australia Workflow Visualize Shared Data Help User Using 42%

Tools search tools Upload Data

FILE AND META TOOLS

Get Data

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

NanoPlot Plotting suite for Oxford Nanopore sequencing data and alignments

medaka variant tool decodes variant calls from medaka consensus output

medaka consensus pipeline Assembly polishing via neural networks

NanoPlot Plotting suite for Oxford Nanopore sequencing data and alignments (Galaxy Version 1.41.0+galaxy0)

Select multifile mode

Applying multiple files, batch mode will run NanoPlot on each file separately, while combined mode will run NanoPlot once on all the reads together. When applying a single input file, this setting does not matter.

Type of the file(s) to work on

fastq

files

22: Trim Galore! on collection 14: trimmed reads

This is a batch mode input field. Separate jobs

(--fastq)

Options for filtering or transforming input reads

Options for customizing the plots created

Email notification

No

Send an email notification when the job completes

Execute

What it does

NanoPlot is a plotting tool for long read sequencing data and alignments written by Wouter De Coster

Input

NanoPlot requires 1 or more files as input. They can either be fastq (can be generated by albacore, guppy or MinKNOW containing additional information), sorted bam, sorted cram or sequencing summary.

Output

History search datasets

Testing main tools for transcriptomics

691 MB 15 1 25

38 : TargetFinder on data 18 and data 20

31 : MiRDeep2 Quantifier on collection 26 (html report)

a list with 3 html datasets

30 : output of MiRDeep2 Quantifier on collection 26

a list with 3 tabular datasets

26 : Collapsed reads of MiRDeep2 Mapper on collection 26

a list with 3 files

miRNA_DESeq2_results_complete_dataset.tabular

20 : miRNA_stem-loop_seq.fasta

tools

tool settings

files

Galaxy has lots of features but they can be hard to find

The screenshot displays the Galaxy Australia web interface. On the left, a 'Tools' sidebar is open, showing a search bar and a list of tool categories under 'FILE AND META TOOLS' and 'GENOMIC FILE MANIPULATION'. The main workspace shows a workflow for 'argyrophylla' with two steps: '6 : ERR2040348:reverse' and '5 : ERR2040348:forward'. A banner for the 'Galaxy Community Conference' is visible in the center. The top navigation bar includes 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'User' menus, along with a 'Using 44%' indicator.

there are some shared workflows but they may not be easy to find - keywords, tags - inconsistent

there are lots of tools! but the list is long

there are lots of tutorials, but it may not be obvious just from the "help" button

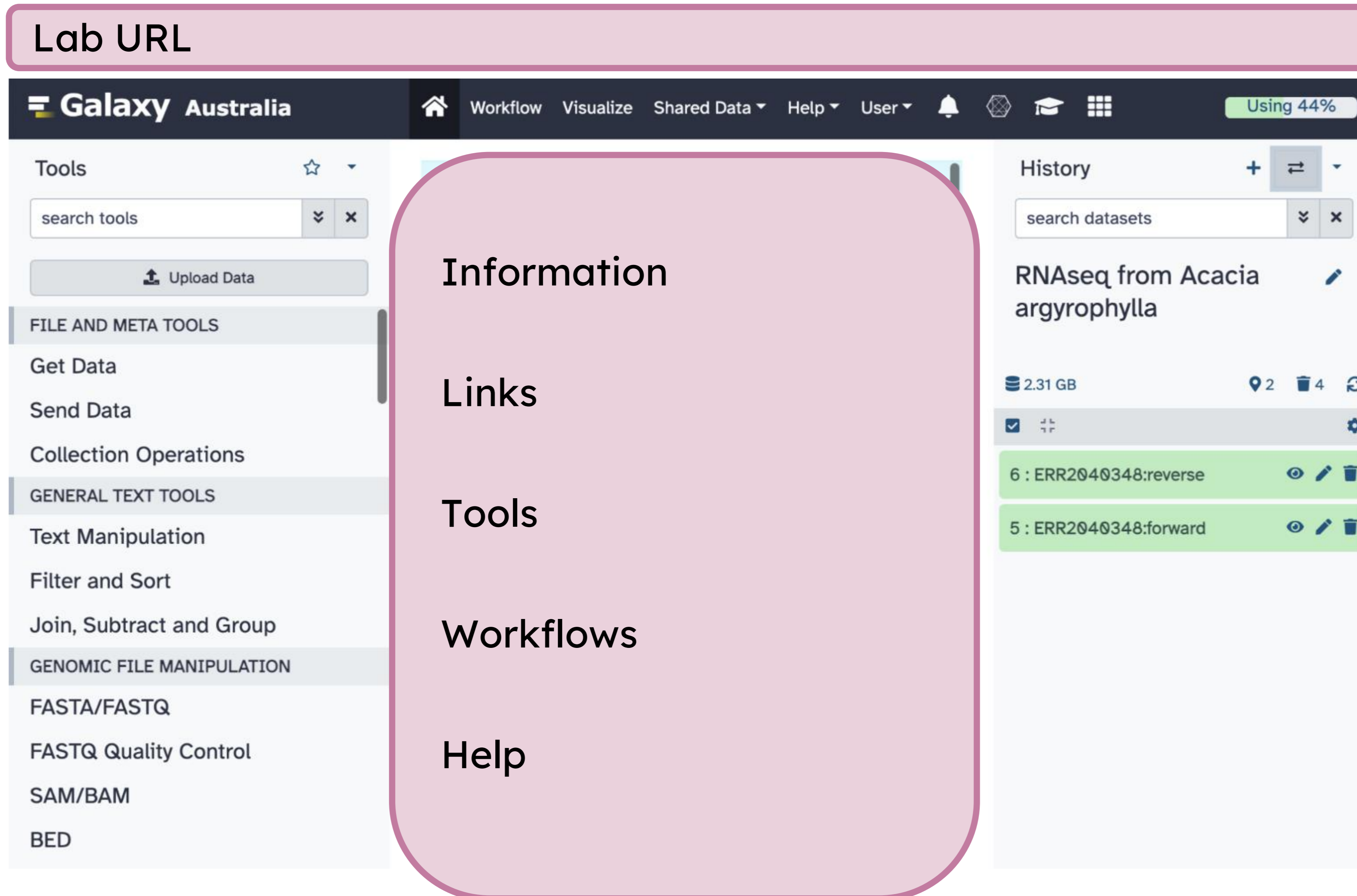
Galaxy Australia is an **open, web-based** platform for accessible, reproducible and transparent computational research. Galaxy supports thousands of documented and maintained tools that are free to use. We facilitate on-demand training capacities and provision **600GB** for Australian institutional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

We can make better use of this space to display relevant resources

The screenshot displays the Galaxy Australia web interface. The top navigation bar includes the Galaxy Australia logo, a home icon, and menu items for Workflow, Visualize, Shared Data, Help, and User. A status indicator shows 'Using 44%'. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by categorized tool lists: FILE AND META TOOLS (Get Data, Send Data, Collection Operations), GENERAL TEXT TOOLS (Text Manipulation, Filter and Sort, Join, Subtract and Group), and GENOMIC FILE MANIPULATION (FASTA/FASTQ, FASTQ Quality Control, SAM/BAM, BED). The central content area features a notification about an upgrade on Thursday June 22, a navigation menu (Home, News, People, About, Support, Docs), the Galaxy AUSTRALIA logo, a banner for the 'Galaxy Community Conference' (10-16 July 2023), and a paragraph describing the platform as an open, web-based platform for accessible, reproducible, and transparent computational research. The right sidebar shows a 'History' section with a search bar for datasets, displaying two entries: '6 : ERR2040348:reverse' and '5 : ERR2040348:forward', each with view, edit, and delete icons. The total size of the datasets is 2.31 GB, with 2 locations and 4 files.

This space has been designed to be a Galaxy Lab

Lab URL



The screenshot shows the Galaxy Australia web interface. At the top, there is a dark navigation bar with the 'Galaxy Australia' logo, a home icon, and menu items for 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'User'. On the right of this bar, it indicates 'Using 44%' of resources. Below the navigation bar, the interface is divided into three main sections: a left sidebar for 'Tools', a central workspace, and a right sidebar for 'History'. The 'Tools' sidebar lists categories like 'FILE AND META TOOLS', 'GENERAL TEXT TOOLS', and 'GENOMIC FILE MANIPULATION'. The 'History' sidebar shows a search for 'RNaseq from Acacia argyrophylla' and lists two datasets: '6 : ERR2040348:reverse' and '5 : ERR2040348:forward'. A large, rounded rectangular overlay in the center of the workspace contains a list of navigation options: 'Information', 'Links', 'Tools', 'Workflows', and 'Help'. A pink bar at the top of the screenshot contains the text 'Lab URL'.

Galaxy Australia

Workflow Visualize Shared Data Help User Using 44%

Tools

search tools

Upload Data

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

Information

Links

Tools

Workflows

Help

History

search datasets

RNaseq from Acacia argyrophylla

2.31 GB 2 4

6 : ERR2040348:reverse

5 : ERR2040348:forward

What is a Lab, in words

- A focused landing page in Galaxy Australia that provides quick access to tools, workflows and tutorials for a particular research area
- URL is the [labname](#) + [usegalaxy.org.au](#)
- It is connected to the same server as [usegalaxy.org.au](#) - the user's history, jobs and data quota are shared with the main server

Other names

- Subdomain
- Site
- Page
- Centre
- Community

Similar pages for particular Galaxy communities

- Usually called subdomains
 - Tend to subset the tools in the tool panel
 - And present links to workflows and GTN tutorials
- Examples =>

SARS-CoV-2 Data Analysis	covid19.usegalaxy.eu
Galaxy for Ecology	ecology.usegalaxy.eu
Erasmus MC Galaxy webserver	erasmusmc.usegalaxy.eu
GraphClust2 Galaxy server	graphclust.usegalaxy.eu
Galaxy HiCExplorer	hicexplorer.usegalaxy.eu
Galaxy Human Cell Atlas	humancellatlas.usegalaxy.eu
Imaging flavour of Galaxy	imaging.usegalaxy.eu
Indian Community	india.usegalaxy.eu
Materials Galaxy workbench	materials.usegalaxy.eu
Metabolomics	metabolomics.usegalaxy.eu
Galaxy for Microbiome (ASaiM)	metagenomics.usegalaxy.eu
Galaxy for Microbiome (ASaiM)	microbiome.usegalaxy.eu
microGalaxy community	microgalaxy.usegalaxy.eu
Galaxy Machine Learning workbench	ml.usegalaxy.eu
NanoGalaxy for ONT	nanopore.usegalaxy.eu
Galaxy for Plant Biology	plants.usegalaxy.eu

First Lab - the Genome Lab

The screenshot shows the Galaxy Australia Genomes Lab interface. The top navigation bar includes the Galaxy logo, the text "Australia - Genomes Lab", and various navigation options like "Workflow", "Visualize", "Shared Data", "Help", "User", and a "Using 44%" indicator. The left-hand menu lists various tool categories such as "FILE AND META TOOLS", "GENERAL TEXT TOOLS", and "GENOMIC FILE MANIPULATION". The central workspace, which is circled in purple, features a header with a navigation menu ("Home", "News", "People", "About", "Support", "Docs") and a large "GENOME LAB" logo. Below the logo, there is a welcome message and a "Data import and preparation" section with tabs for "Tools", "Workflows", and "Help". The "Tools" tab is active, showing a list of tools including "Import data to Galaxy", "FastQC", and "FastP". The right-hand panel displays the "History" section, showing a search for "RNaseq from Acacia argyrophylla" and a list of datasets, including "ERR2040348:reverse" and "ERR2040348:forward".

Content

- Genome assembly and annotation (not the more comprehensive "genomics")
- Trying to keep it simple and uncluttered
- Keep start section very short, with link to more information.
- Three topics
 - **Data import and preparation**
 - **Genome assembly**
 - **Genome annotation**



GENOME LAB

Welcome to the Galaxy Australia Genome Lab. Get quick access to tools, workflows and tutorials for genome assembly and annotation.
[What is this page?](#)

The influence of layout

- Developed ideas about content in concert with thinking about the resulting layout
- Wanted to show all the things available but not in a big wall of text: nested information
- Decided to show a tabbed navigation bar for each of the topics: each has **tools, workflows, help.**

Data import and preparation

Tools

Workflows

Help

Tools in Galaxy

- the problem that Labs are trying to solve - ***complicated tool views*** - is a problem for all of Galaxy.
- Tool box re-engineering is happening globally.
- Subsetting of the tool panel on the left is difficult.
- Presenting researchers with best-practice tools was the chosen outcome

Tools ☆ ▾

search tools ▾ ✕

⬆ Upload Data

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

COMMON GENOMICS TOOLS

- Operate on Genomic Intervals
- MiModD
- Fetch Alignments/Sequences

GENOMICS ANALYSIS

- Assembly

Genome assembly

Tools Workflows Help

Common tools are listed here, or search for more in the full tool panel to the left.

- Hifiasm - assembly with PacBio HiFi data ▾
- Flye - assembly with PacBio or Nanopore data ▾
- Unicycler - assembly with Illumina, PacBio or Nanopore data - bacteria only ▾
- Salsa - scaffold assembly with HiC data ▾
- Quast - assess genome assembly quality ▾
- Busco - assess genome assembly quality ▾

Flye - assembly with PacBio or Nanopore data ▴

de novo assembly of single-molecule sequencing reads, designed for a wide range of datasets, from small bacterial projects to large mammalian-scale assemblies.

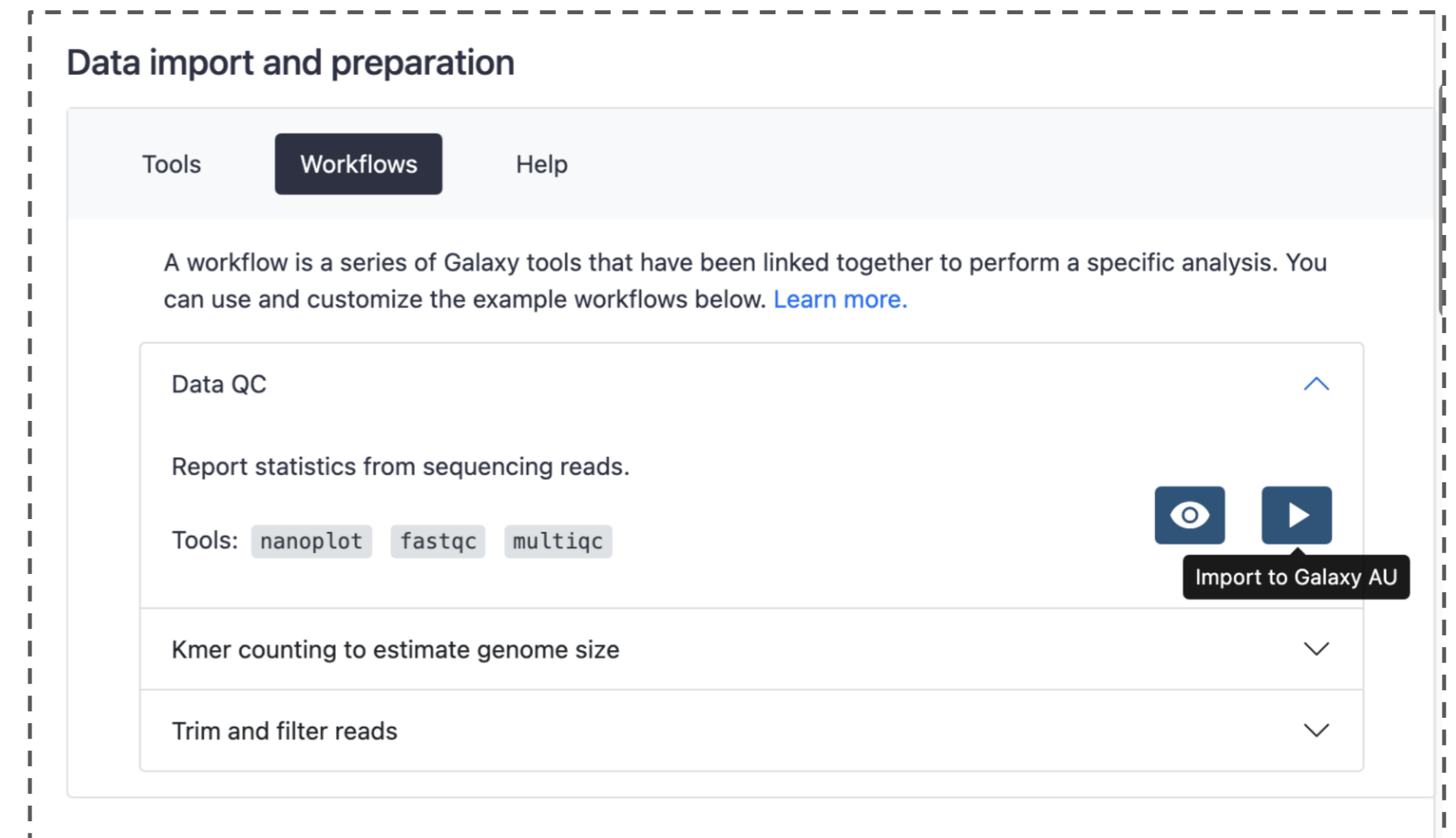
Input data:

- fasta
- fastq

▶

More about the workflows section

- A short list of workflows, from mainly the Galaxy Training Network tutorials, or from Australian Biocommons / Galaxy Australia
- Ideally hosted on workflowhub.eu (or Dockstore) for version control
- Import button



The screenshot displays the 'Data import and preparation' section of a Galaxy interface. At the top, there are navigation tabs for 'Tools', 'Workflows', and 'Help'. Below this, a text block explains that a workflow is a series of Galaxy tools linked together for a specific analysis, with a 'Learn more' link. A list of workflow cards is shown, with the first card expanded. The 'Data QC' card includes a description 'Report statistics from sequencing reads.', a list of tools 'nanoplot', 'fastqc', and 'multiqc', and an 'Import to Galaxy AU' button. Other workflow cards visible are 'Kmer counting to estimate genome size' and 'Trim and filter reads'.

More about the help section

Genome assembly

Tools Workflows **Help**

Can I use Galaxy Australia to assemble a large genome? ▼

How can I learn about genome assembly? ▼

Genome assembly overview

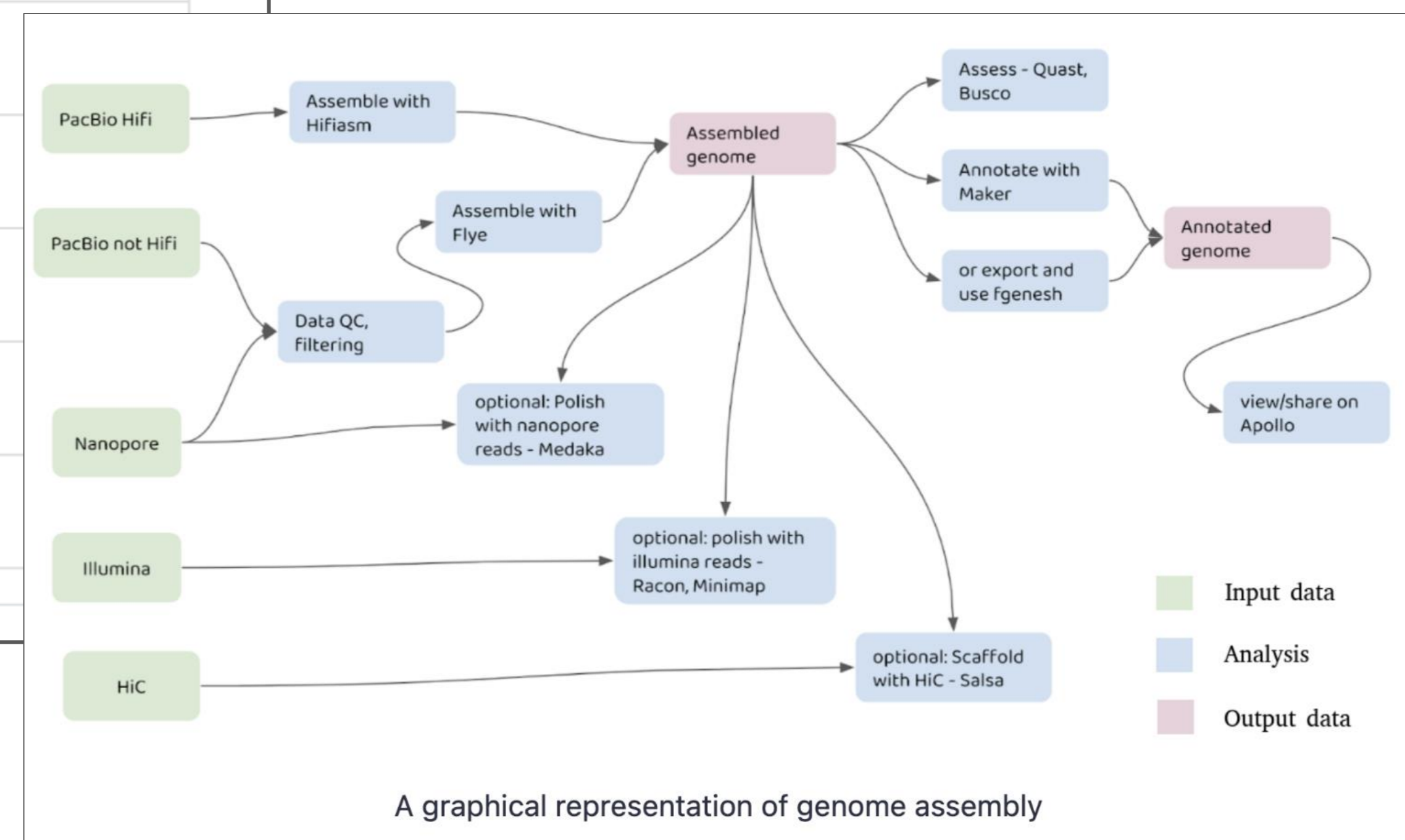
Which tools should I use?

Tutorials

How can I assess the quality of my genome assembly?

Galaxy Australia support

- Tutorials
- Short help text and links
- Support



Technical - How we made the page

- In the existing Galaxy Media Site (a Django website)
- This currently displays various information for Galaxy Australia, such as the centre panel on the landing page
- Content in HTML templates
- Uses the Bootstrap UI library
- Made with reuse in mind

EXPERT REVIEW OF PROTEOMICS
<https://doi.org/10.1080/14789450.2023.2265062>



REVIEW



A Galaxy of informatics resources for MS-based proteomics

Subina Mehta^a, Matthias Bernt^b, Matthew Chambers^c, Matthias Fahrner^{d,e}, Melanie Christine Föll^{d,e,f}, Bjoern Gruening^g, Carlos Horro^{h,i}, James E. Johnson^l, Valentin Loux^{k,l}, Andrew T. Rajczewski^a, Oliver Schilling^{d,e}, Yves Vandebrouck^m, Ove Johan Ragnar Gustafssonⁿ, W. C. Mike Thang^{o,p}, Cameron Hyde^{o,q}, Gareth Price^{o,p}, Pratik D. Jagtap^a and Timothy J. Griffin^a

includes a gateway specifically for MS-based proteomic tools (proteomics.usegalaxy.eu), with a similar gateway under development on the Australian-maintained public instance (proteomics.usegalaxy.org.au). These gateways are powerful because a Galaxy instance can host many thousands of tools spanning many different domains.

ARTICLE HISTORY

Received 5 July 2023

Accepted 6 September 2023

Tools

search tools

Upload Data

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

COMMON GENOMICS TOOLS

- Operate on Genomic Intervals
- MiModD



Welcome to the Galaxy Australia Genome Lab. Get quick access to tools, workflows and tutorials for genome assembly and annotation.

What is this page?

Data import and preparation

- Tools
- Workflows
- Help

Common tools are listed here, or search for more in the full tool panel to the left.

- Import data to Galaxy
- FastQC - sequence quality reports
- FastP - sequence quality reports, trimming & filtering
- NanoPlot - visualize Oxford Nanopore data
- GenomeScope - estimate genome size
- Meryl - count kmers

History

search datasets

Easily switch between Galaxy AU Base site and Lab sites via new button

- 486 : gfastats on data 485 and data 429: edited sequences
- 485 : Concatenate datasets on data 484 and data 483
- 484 : Replace on data 482
- 483 : Text transformation on data 481
- 482 : Text transformation on data 479
- 481 : contaminant scaffolds on data 478

What's next?

- Documentation re: Lab deployment to help with future labs
- More Labs - e.g. Proteomics Lab
- Enhancements e.g.
 - Tag tools with analysis; fit with user input data and aims; display relevant workflows and help
 - Better community integration/management (additional tabs for discussion?)
 - Show example histories for workflows

Please try it out and let us know what you think!

What's happening in Australian genomics research?

Explore

What do you think of the Genome Lab?

Give feedback

Thanks to Galaxy AU and the Genome Lab team



Cameron Hyde, Winnie Mok, Gareth Price, Madeline Bassetti, Anna Syme



Gareth Price



Catherine Bromhead



Justin Lee



Nuwan Goonasekera



Igor Makunin



Michael Thang



Cameron Hyde



Tom Harrop



Anna Syme



Thanks!

Any questions?

You can email Galaxy Australia at:
help@genome.edu.au