Building Galaxy Labs to advance life science research

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

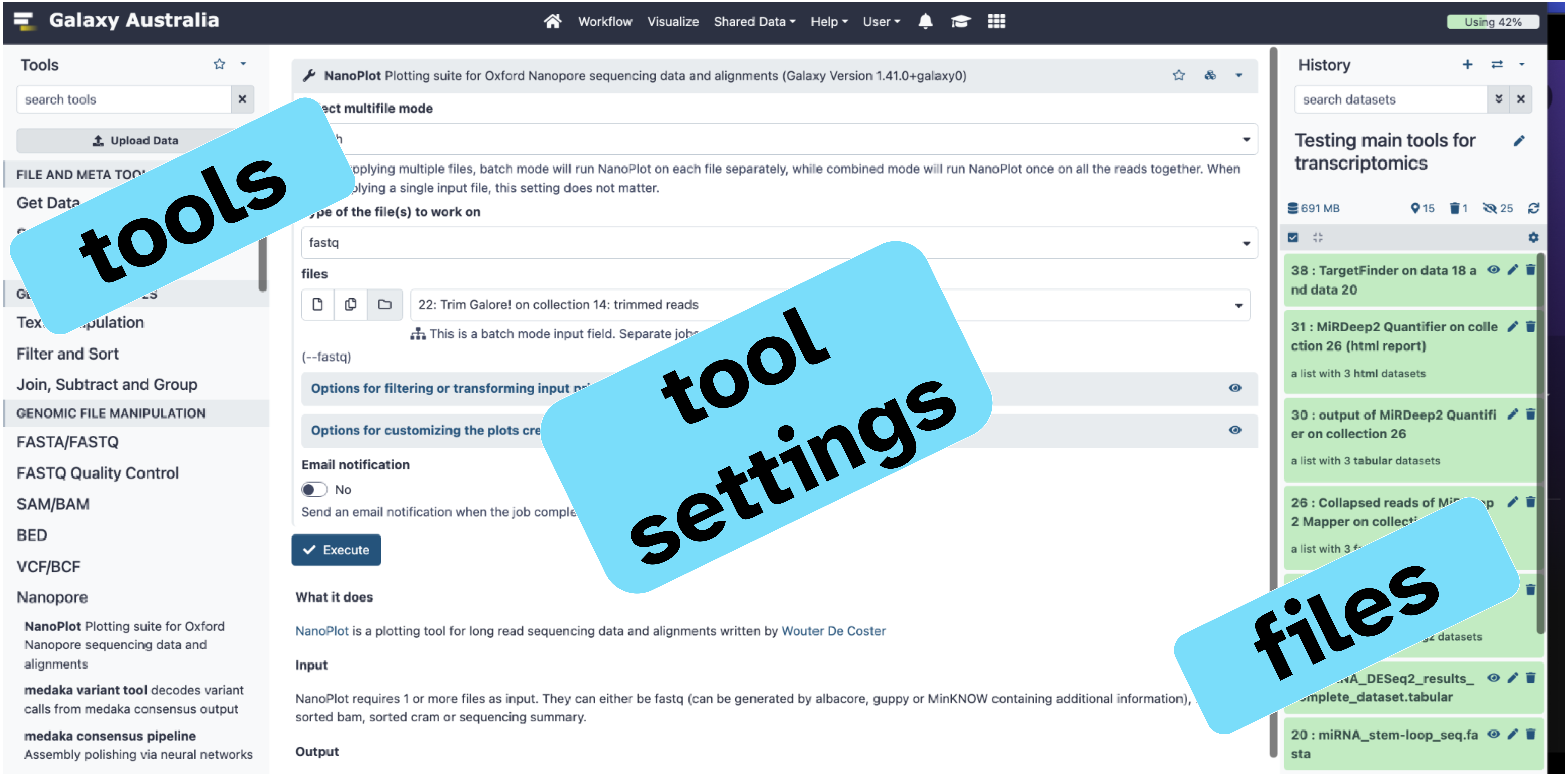
Australian BioCommons

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



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

ect multifile mode

h

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

pe 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 job

(--fastq)

Options for filtering or transforming input r

Options for customizing the plots cre

Email notification

No

Send an email notification when the job comple

Execute

What it does

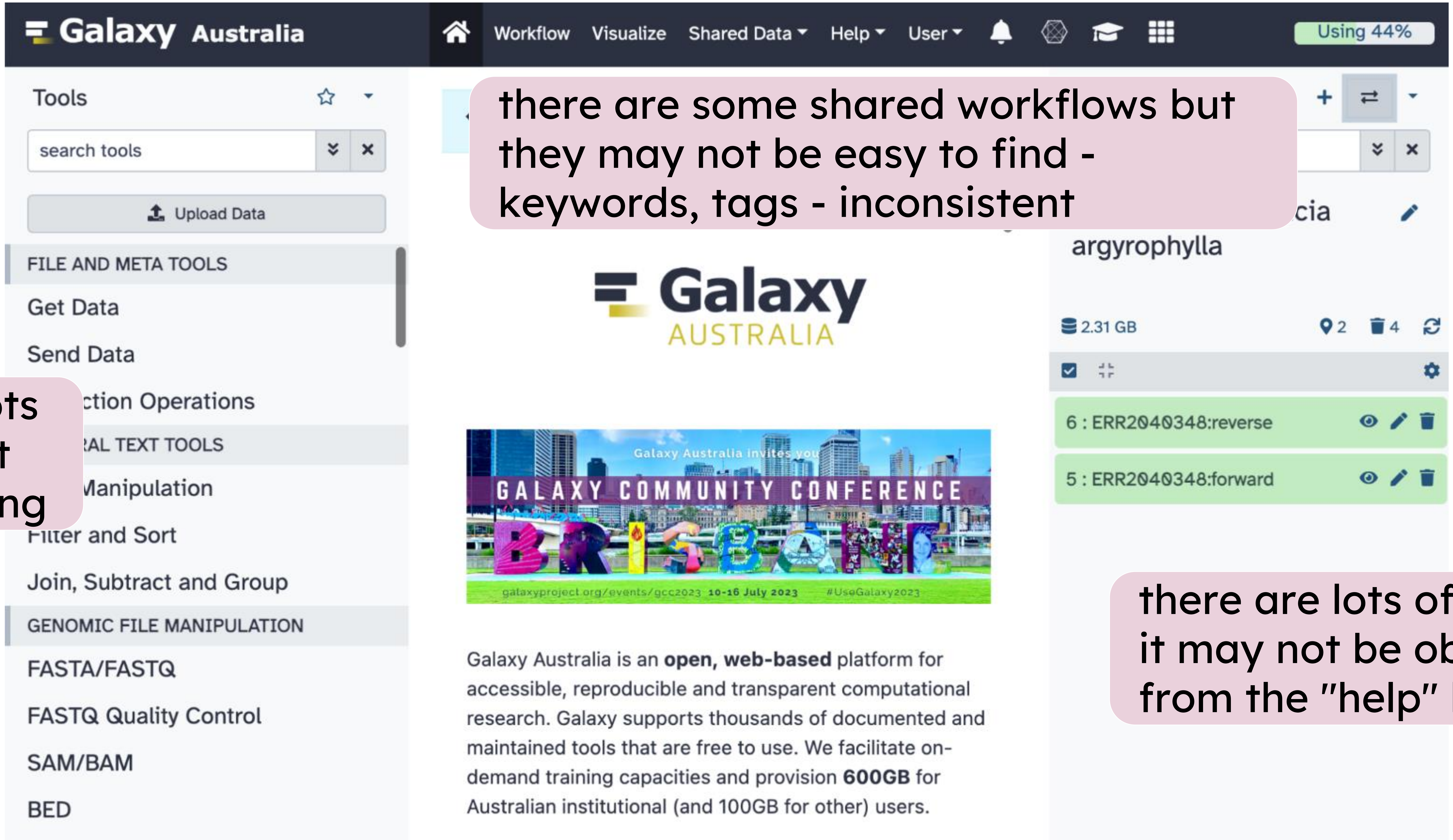
NanoPlot is a plotting tool for long read sequencing data and alignments written by Wouter De Coster

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

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



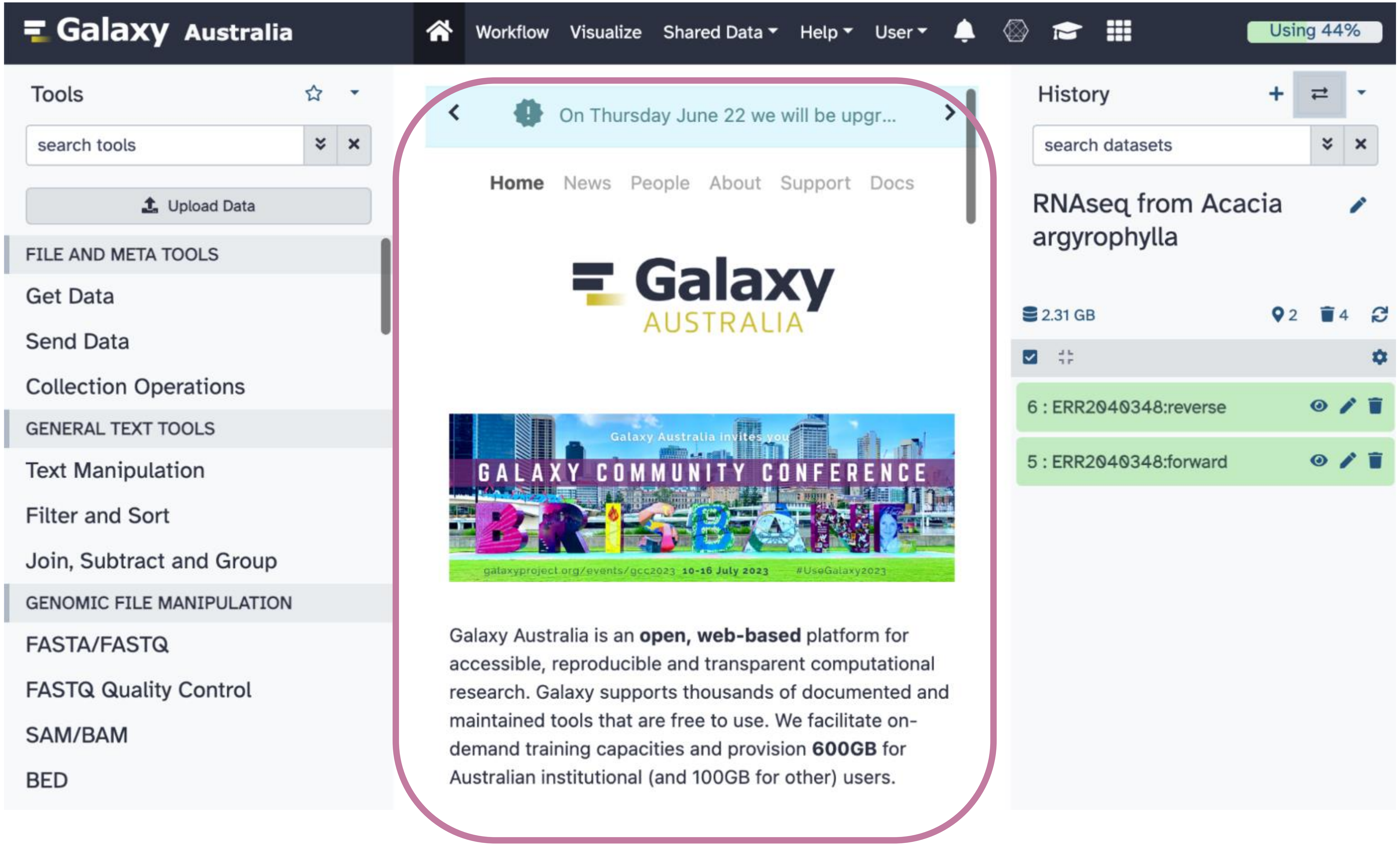
Galaxy Australia is an **open, web-based** platform for accessible, reproducible and transparent computational research. Galaxy supports thousands of documented and maintained tools that are free to use. We facilitate on-demand training capacities and provision **600GB** for Australian institutional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

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

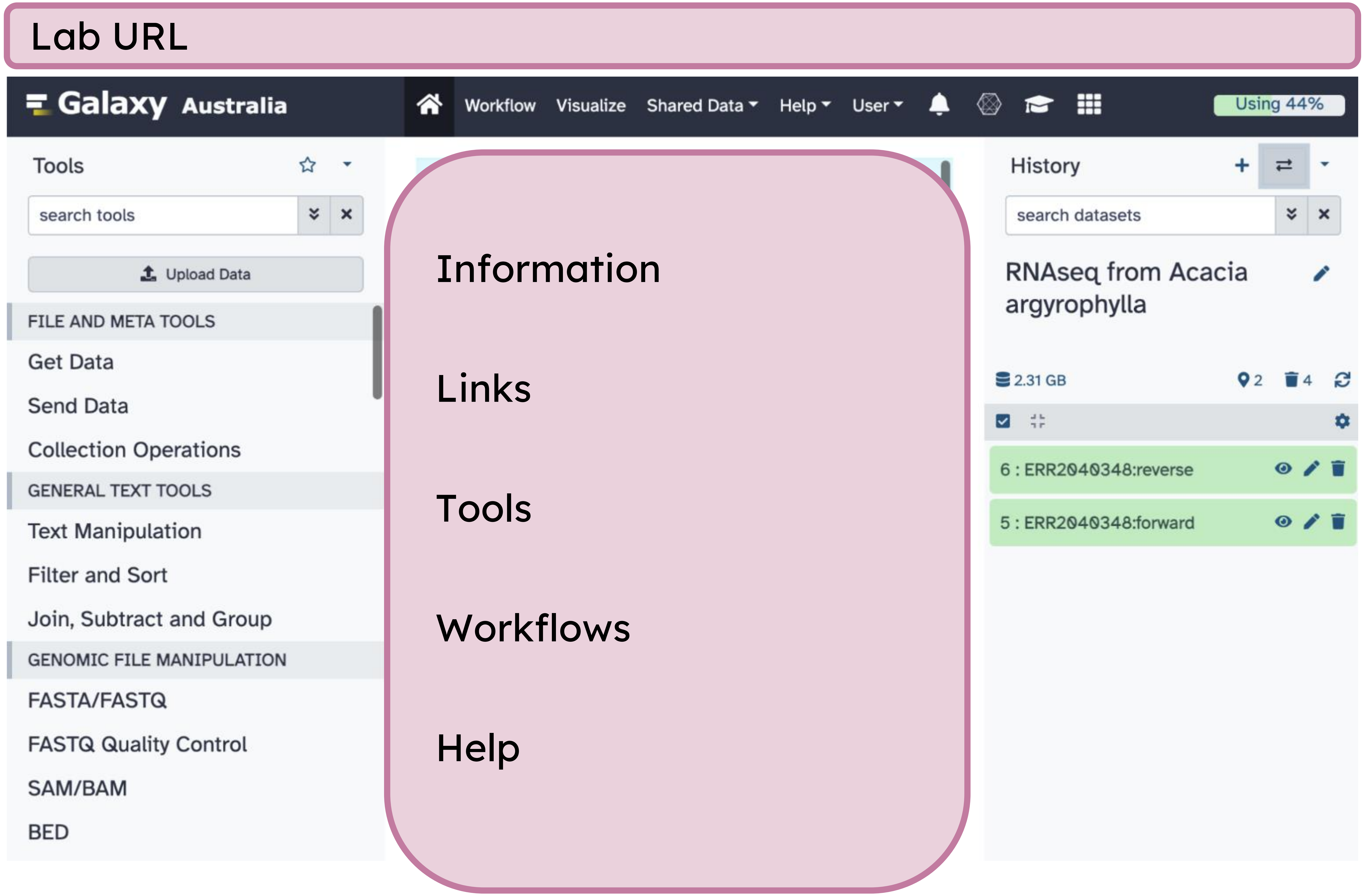
We can make better use of this space to display relevant resources



On Thursday June 22 we will be upgr...

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.

This space has been designed to be a Galaxy Lab



Lab URL

Information

Links

Tools

Workflows

Help

What is a Lab, in words

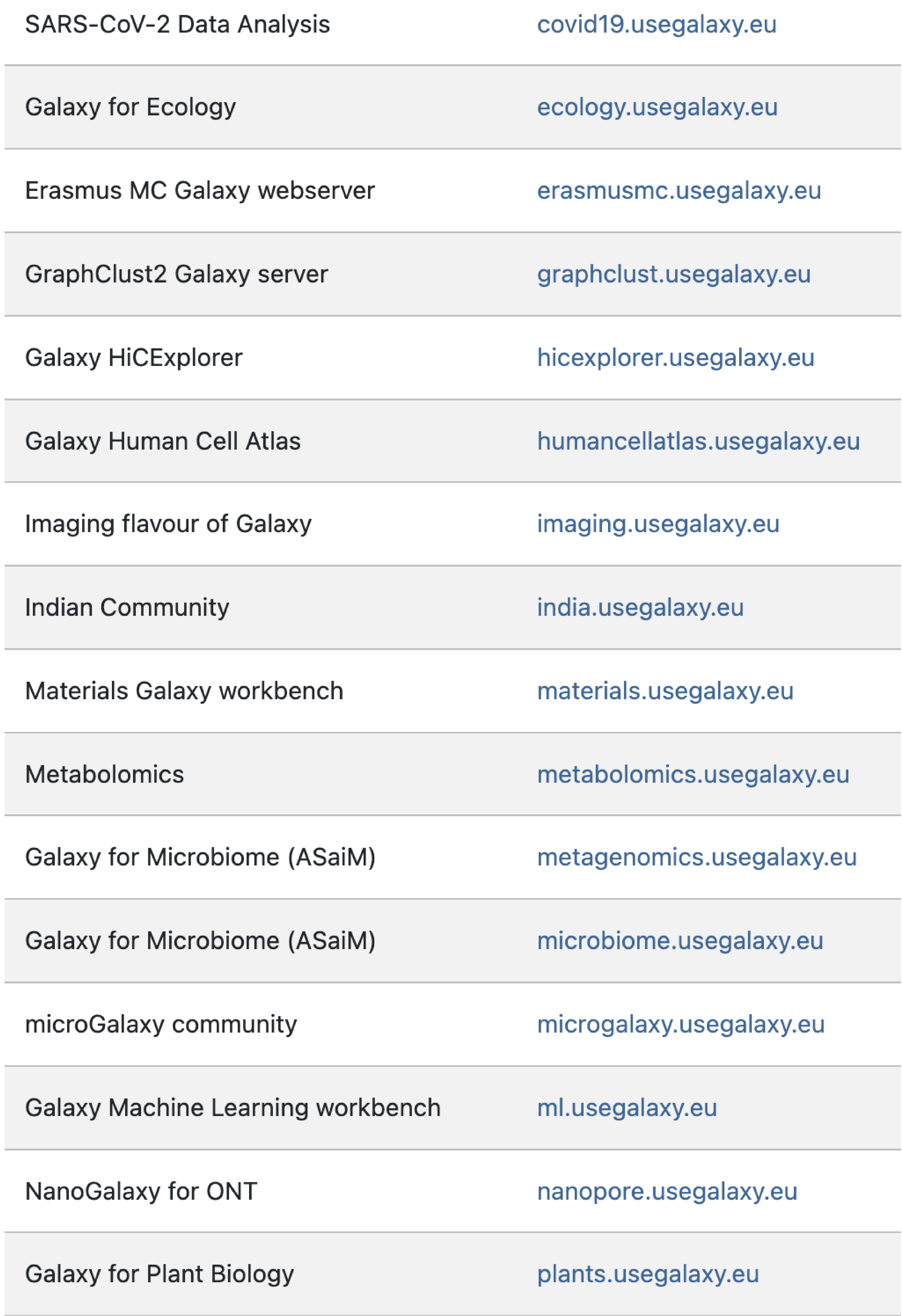
- A focused landing page in Galaxy Australia that provides quick access to tools, workflows and tutorials for a particular research area
- URL is the *labname* + *usegalaxy.org.au*
- It is connected to the same server as *usegalaxy.org.au* - the user's history, jobs and data quota are shared with the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

Other names

- Subdomain
- Site
- Page
- Centre
- Community

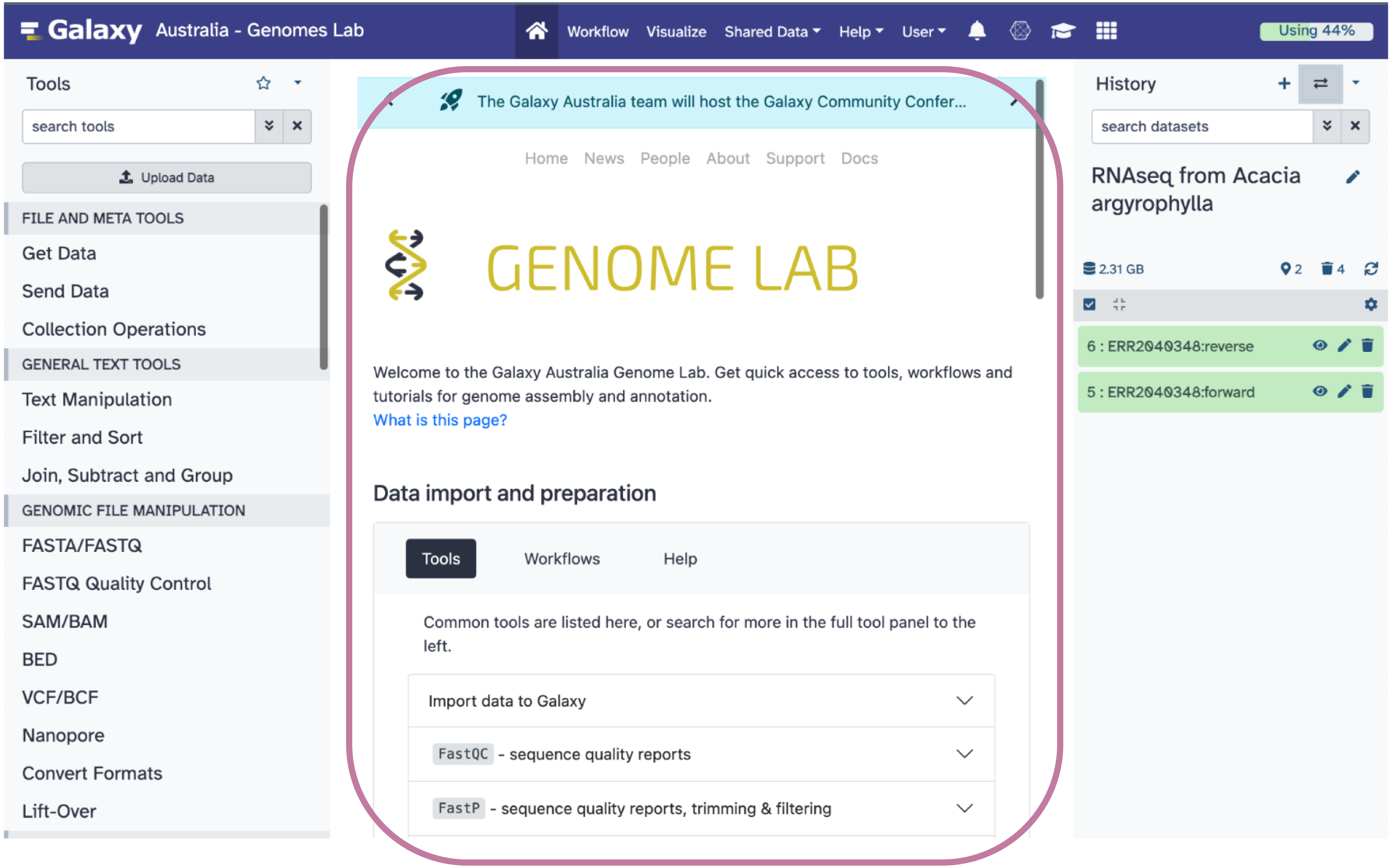
Similar pages for particular Galaxy communities

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



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 Galaxy Australia team will host the Galaxy Community Confer...

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

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

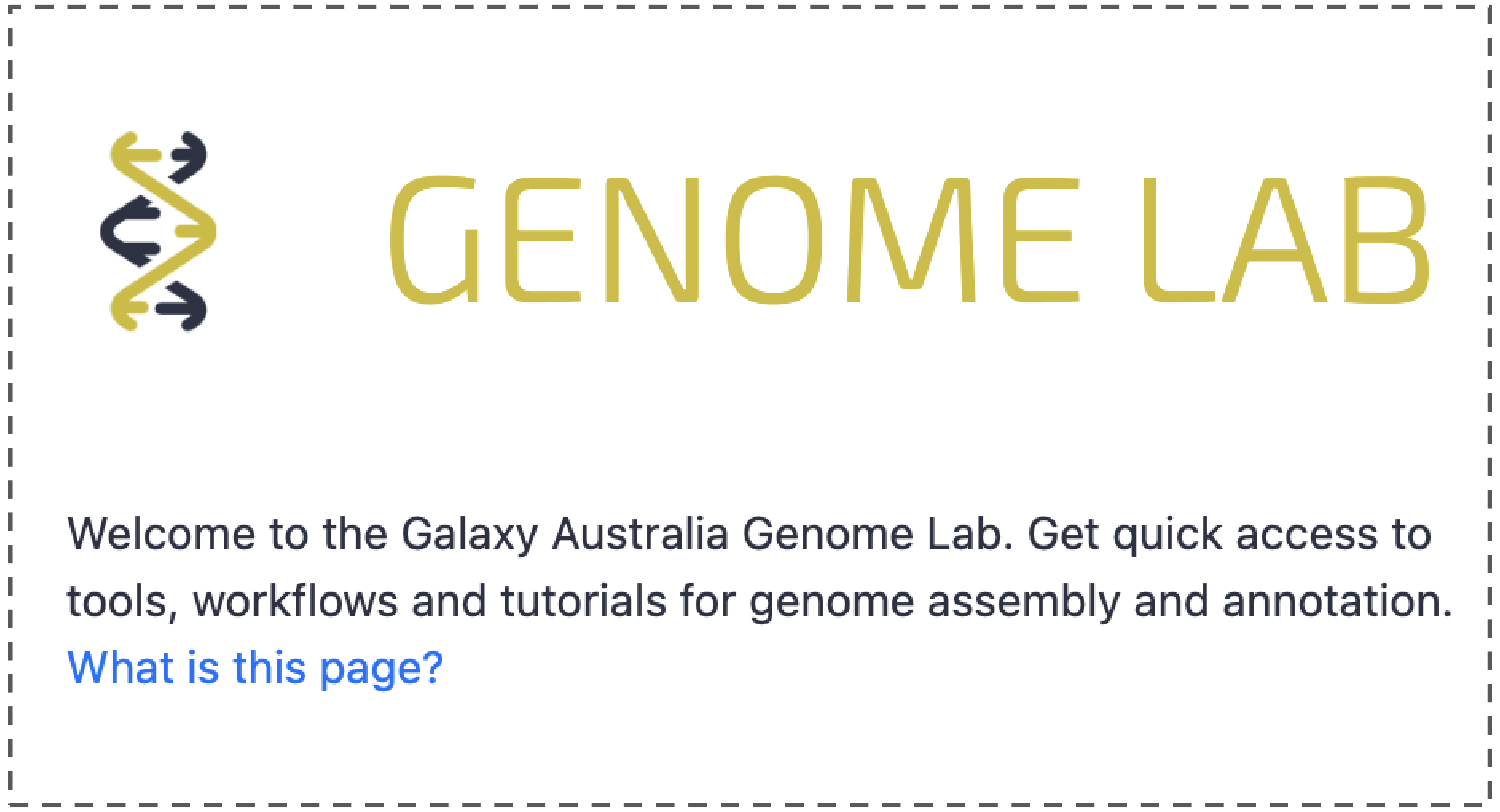
Import data to Galaxy

`FastQC` - sequence quality reports

`FastP` - sequence quality reports, trimming & filtering

Content

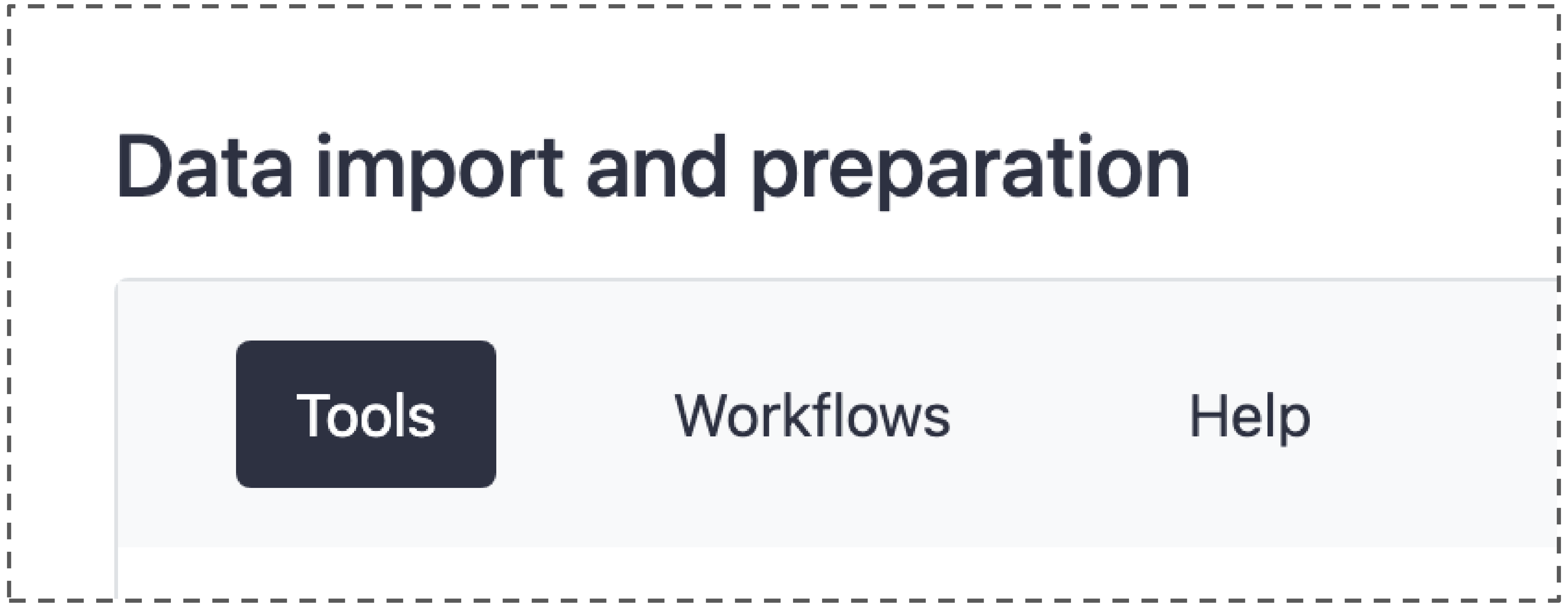
- Genome assembly and annotation (not the more comprehensive "genomics")
- Trying to keep it simple and uncluttered
- Keep start section very short, with link to more information.
- Three topics
    - **Data import and preparation**
    - **Genome assembly**
    - **Genome annotation**



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

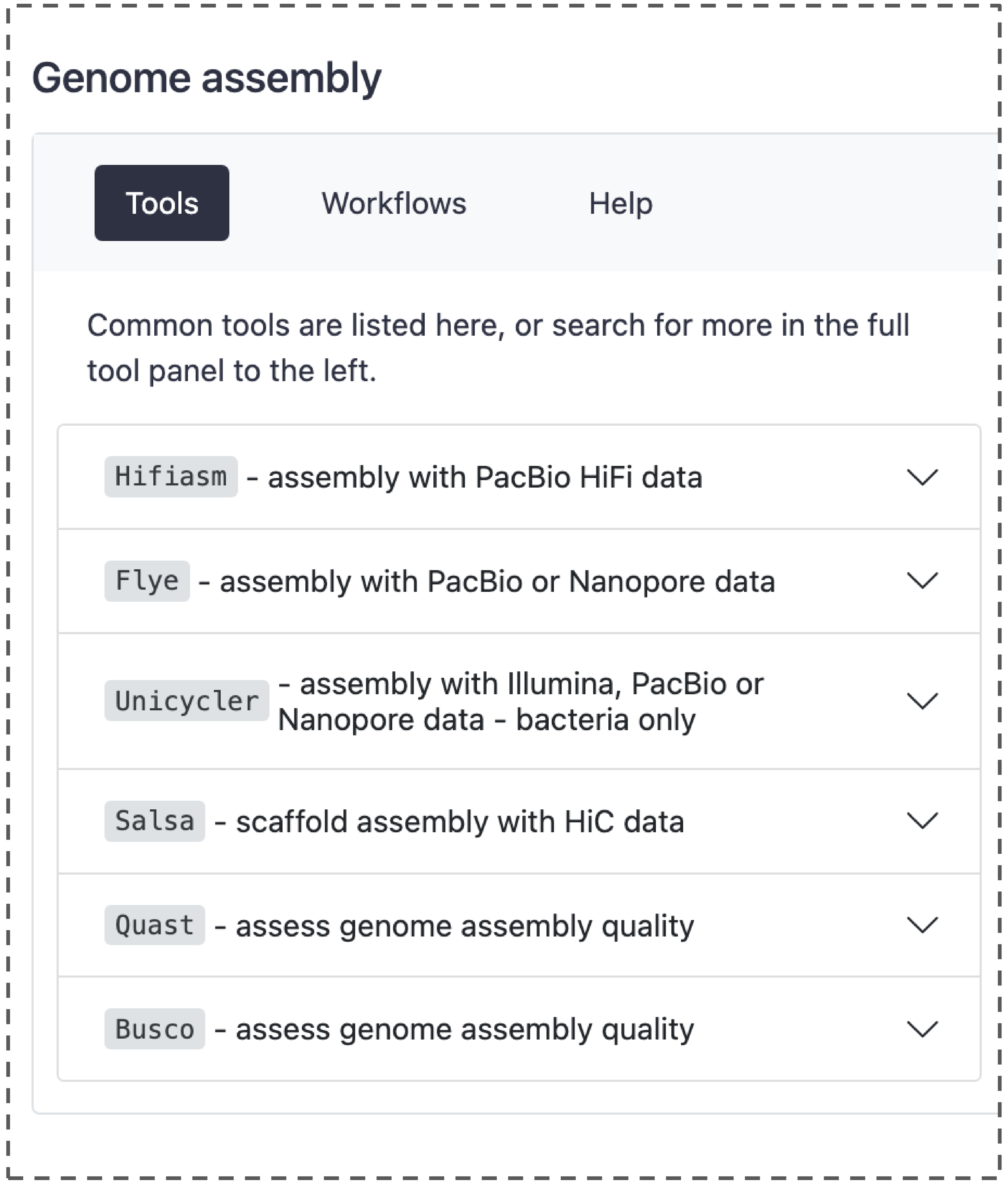
- Developed ideas about content in concert with thinking about the resulting layout
- Wanted to show all the things available but not in a big wall of text: nested information
- Decided to show a tabbed navigation bar for each of the topics: each has **tools**, **workflows**, **help**.



Data import and preparation

Tools 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



Genome assembly

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

`Hifiasm` - assembly with PacBio HiFi data

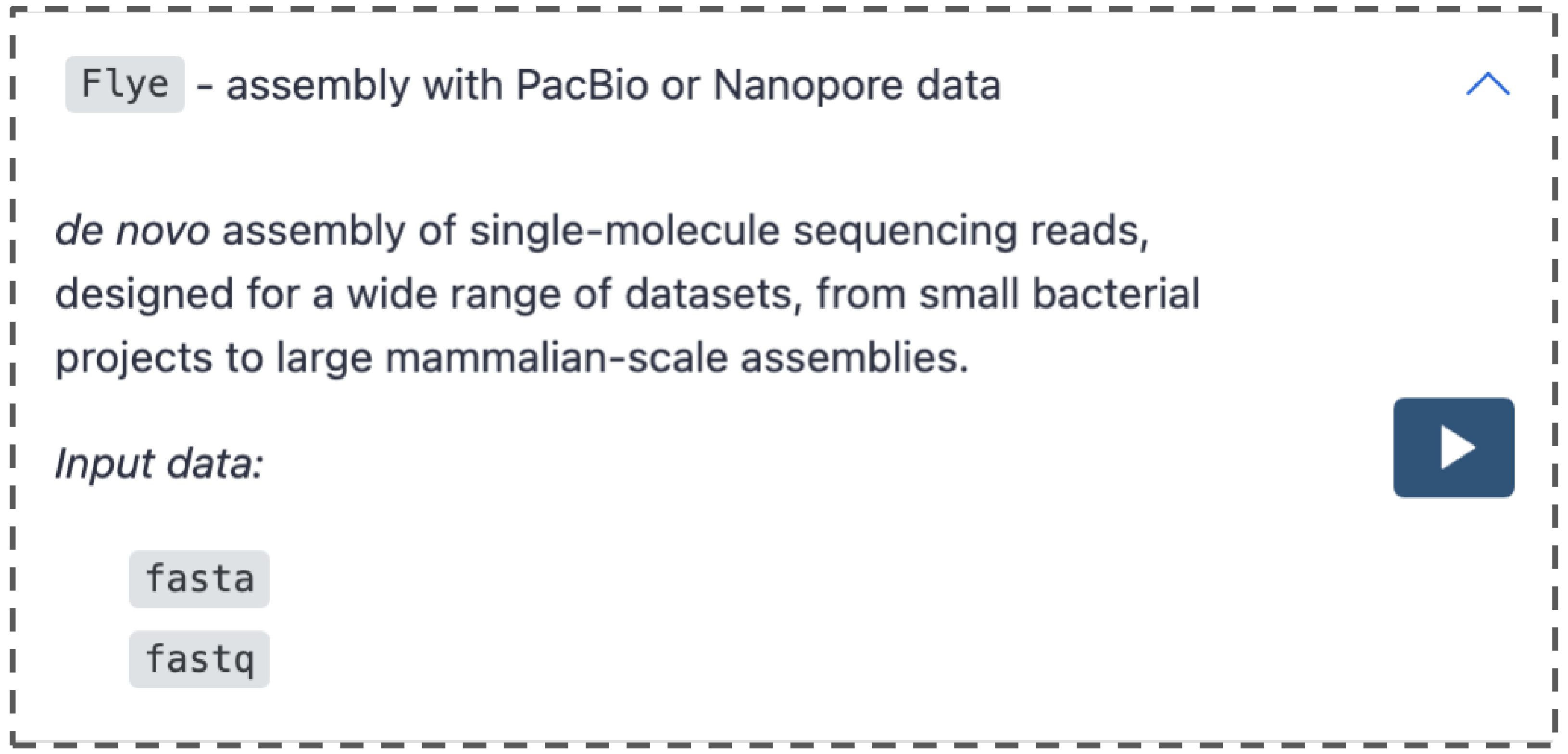
`Flye` - assembly with PacBio or Nanopore data

`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

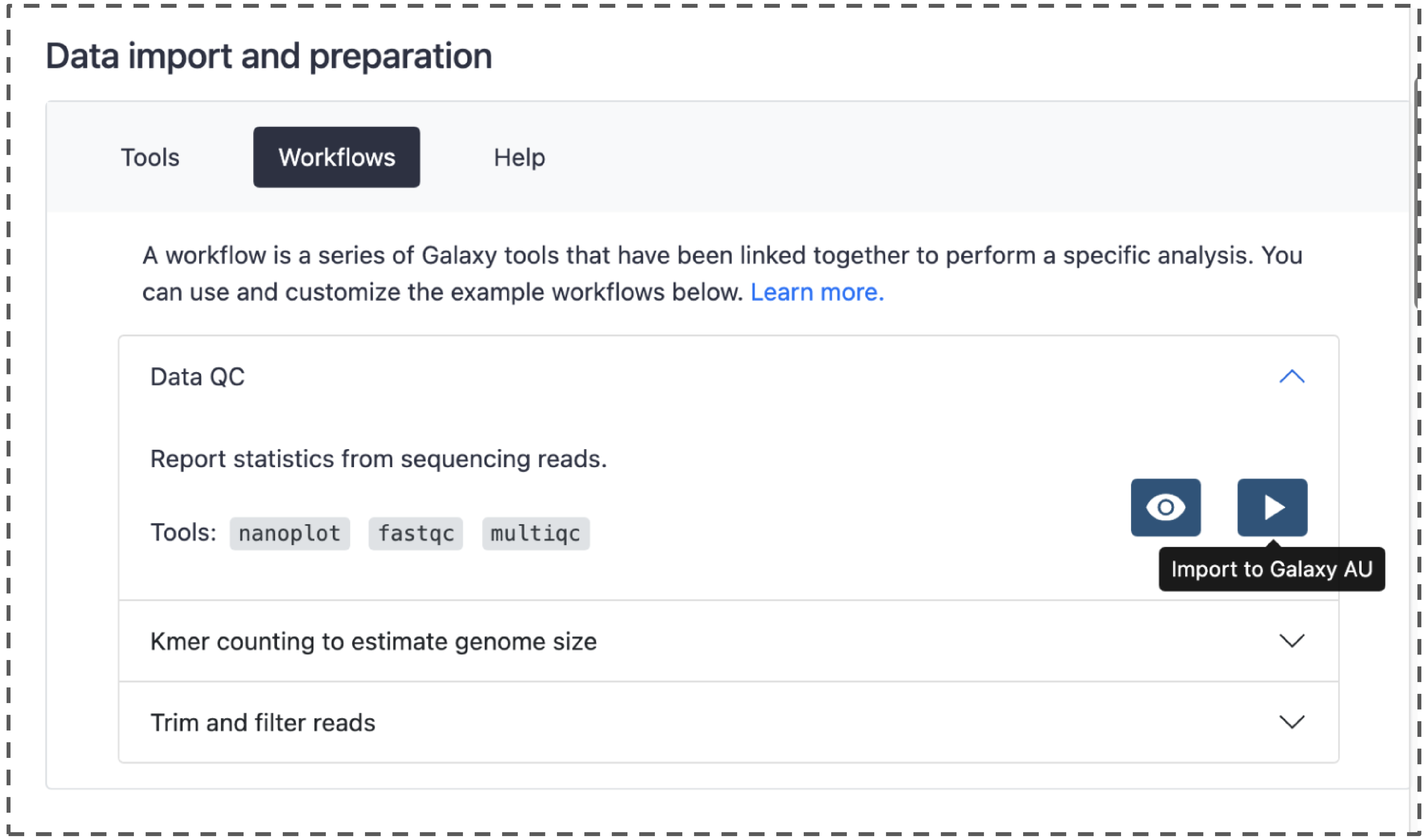
de novo assembly of single-molecule sequencing reads, designed for a wide range of datasets, from small bacterial projects to large mammalian-scale assemblies.

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



Data import and preparation

A workflow is a series of Galaxy tools that have been linked together to perform a specific analysis. You can use and customize the example workflows below. Learn more.

Data QC

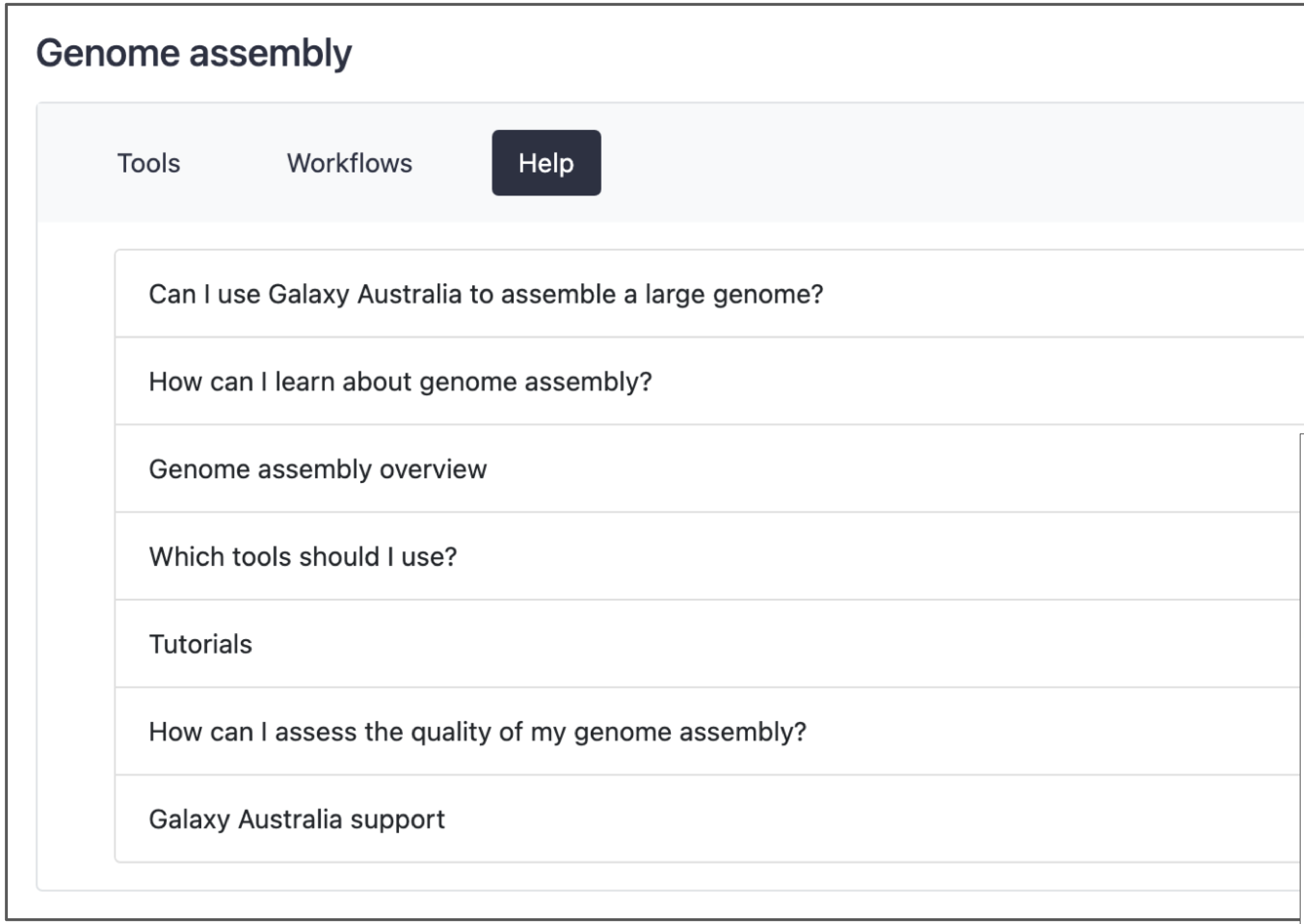
Report statistics from sequencing reads.

Tools: `nanoplot` `fastqc` `multiqc`

Kmer counting to estimate genome size

Trim and filter reads

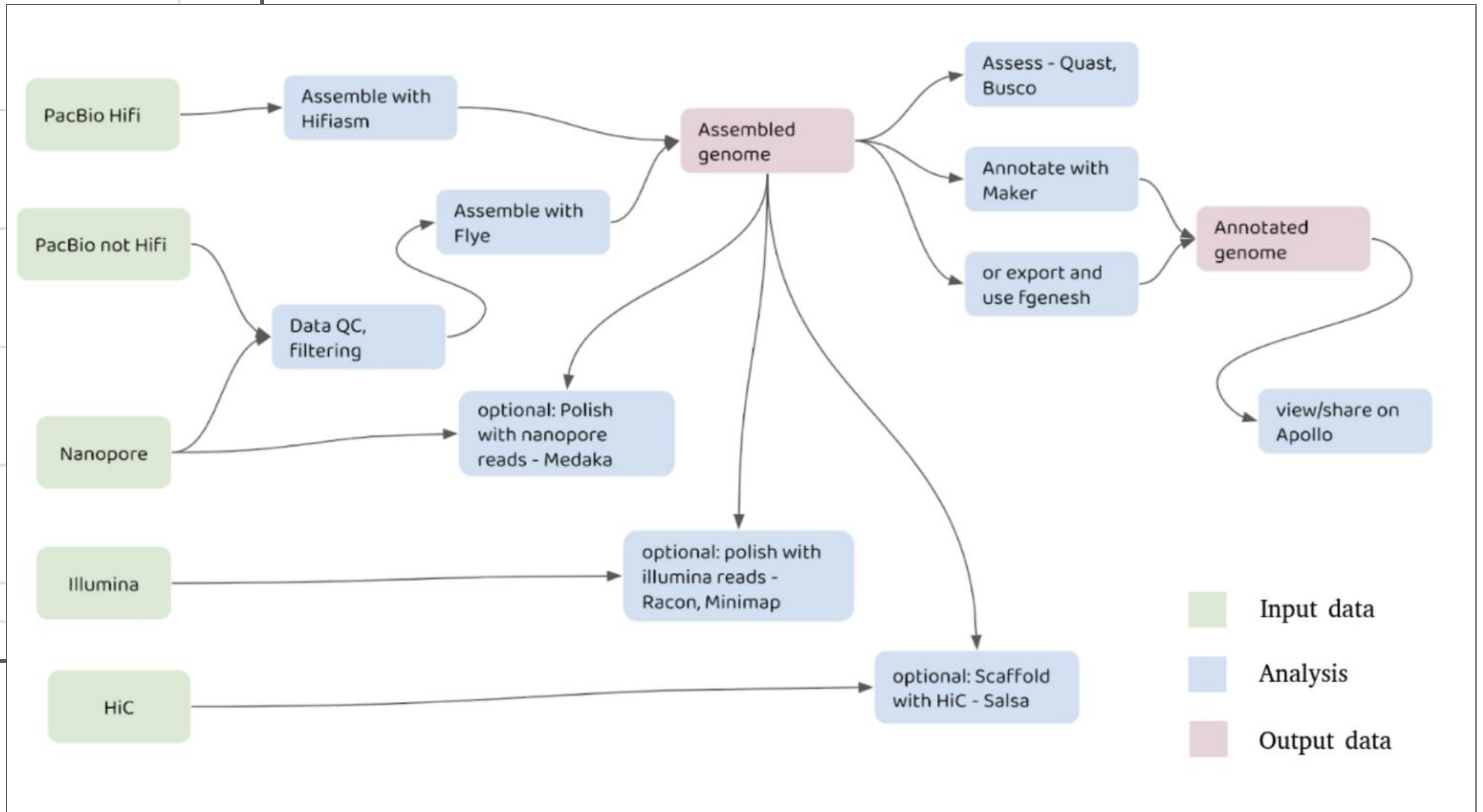
More about the help section



Genome assembly

- 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



A graphical representation of genome assembly

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

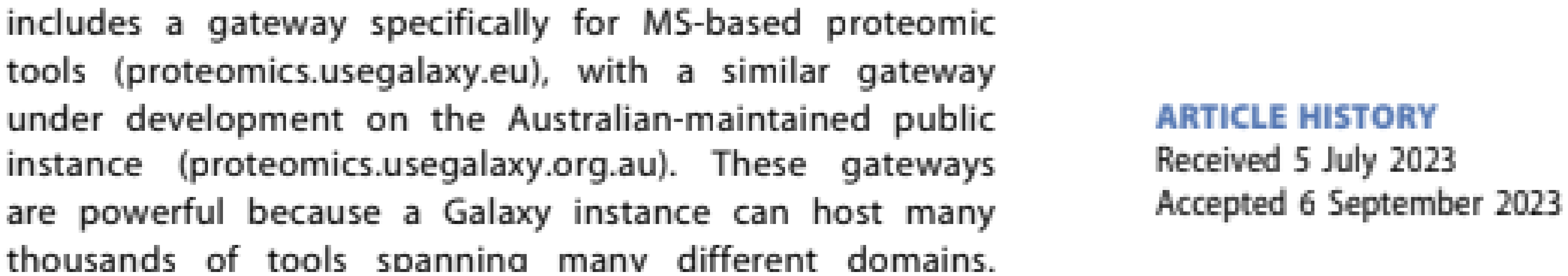
Taylor & Francis
Taylor & Francis Group

REVIEW

Check for updates

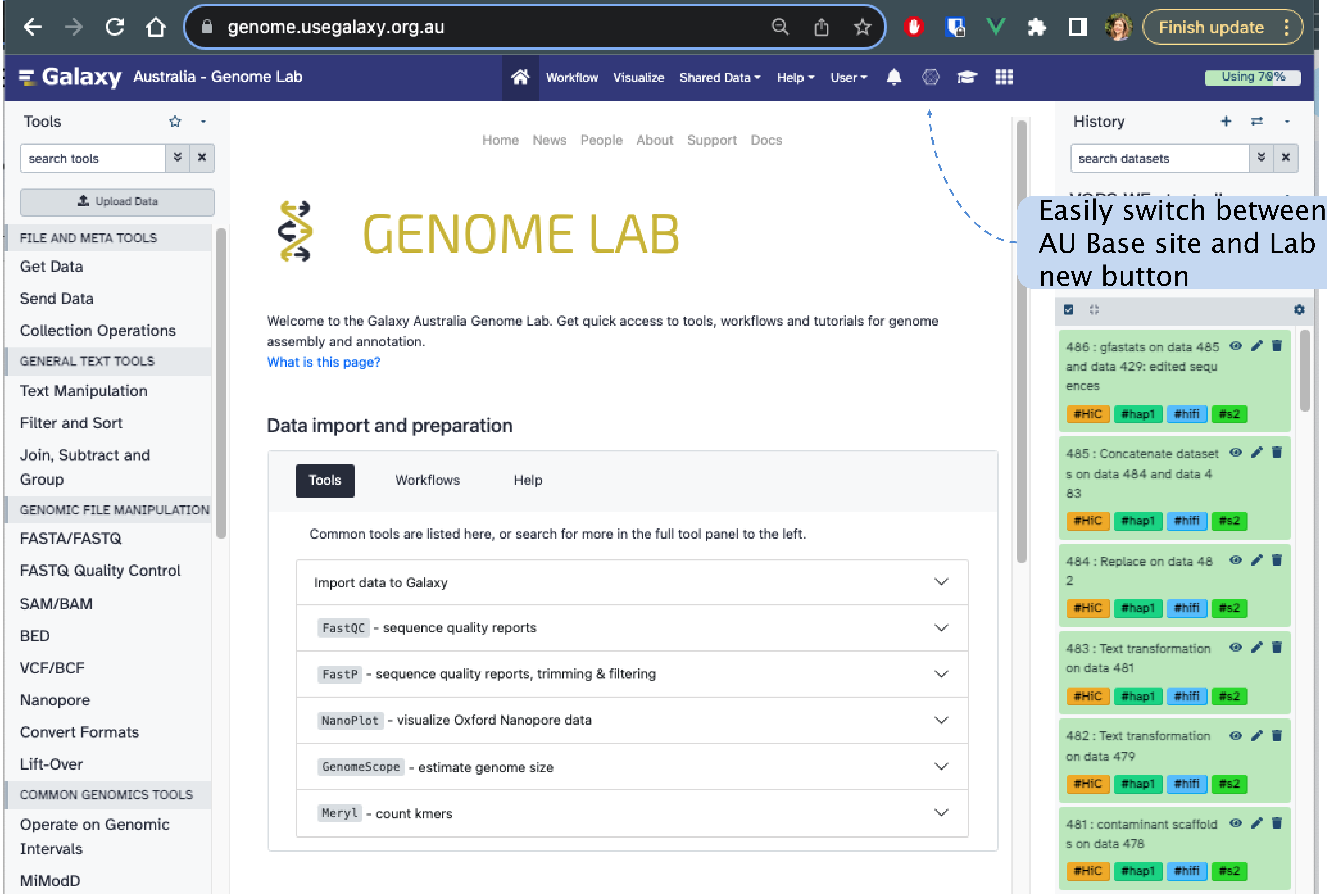
A Galaxy of informatics resources for MS-based proteomics

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[j], Valentin Loux[k,l], Andrew T. Rajczewski[a], Oliver Schilling[d,e], Yves Vandenbrouck[m], Ove Johan Ragnar Gustafsson[n], 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



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

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?

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

486 : gfastats on data 485 and data 429: edited sequences
#HiC #hap1 #hifi #s2

485 : Concatenate datasets on data 484 and data 483
#HiC #hap1 #hifi #s2

484 : Replace on data 482
#HiC #hap1 #hifi #s2

483 : Text transformation on data 481
#HiC #hap1 #hifi #s2

482 : Text transformation on data 479
#HiC #hap1 #hifi #s2

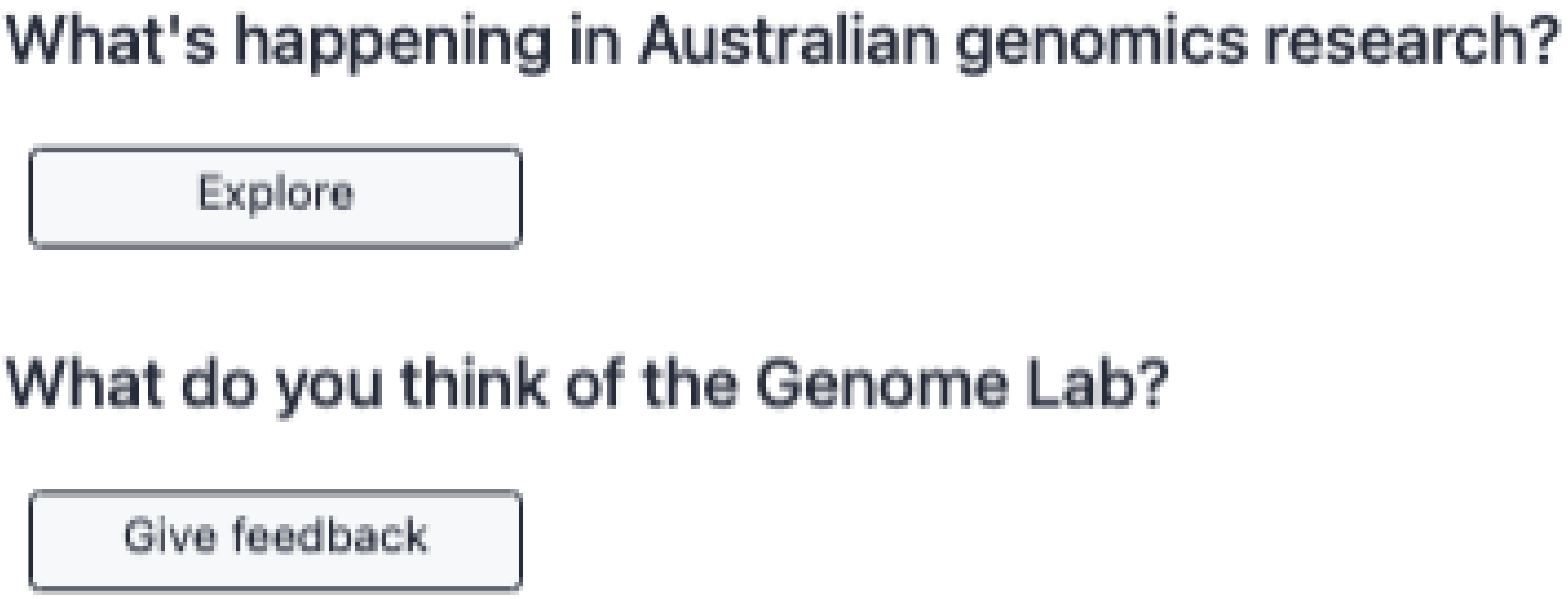
481 : contaminant scaffolds on data 478
#HiC #hap1 #hifi #s2

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

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?

What do you think of the Genome Lab?

Thanks to Galaxy AU and the Genome Lab team

Galaxy
AUSTRALIA

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

Australian BioCommons

QCIF

Melbourne Bioinformatics

aarnet

pawsey

NCI AUSTRALIA

BIOPLATFORMS AUSTRALIA

ARDC
Australian Research Data Commons

Galaxy
PROJECT

Thanks!

Any questions?

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

Australian BioCommons