Here's one we prepared earlier

(re)creating bioinformatics methods and workflows with Galaxy Australia

Gareth Price
Johan Gustafsson

Australian BioCommons

open, web-based platform for

accessible, reproducible and transparent

computational research

Australian BioCommons

A world of Galaxies

usegalaxy.*

Australian BioCommons

Galaxy numbers

3

Main instances

160+

Total implementations

Australian BioCommons

Galaxy numbers

3

Main instances

160+

Total implementations

1,500+

Tools available on Galaxy Australia

8,000+

Tools available in total

Australian BioCommons

Galaxy numbers

3

Main instances

160+

Total implementations

1,500+

Tools available on Galaxy Australia

61,000+

Workflow invocations

4.7M+

Jobs runs

8,000+

Tools available in total

Australian BioCommons

Galaxy numbers

3

Main instances

160+

Total implementations

1,500+

Tools available on Galaxy Australia

61,000+

Workflow invocations

4.7M+

Jobs runs

8,000+

Tools available in total

>600 GB

Storage available to each Australian users

11M+

User datasets

Australian BioCommons

Galaxy numbers

3
Main instances

160+
Total implementations

1,500+
Tools available on Galaxy Australia

61,000+
Workflow invocations

4.7M+
Jobs runs

8,000+
Tools available in total

>600 GB
Storage available to each Australian users

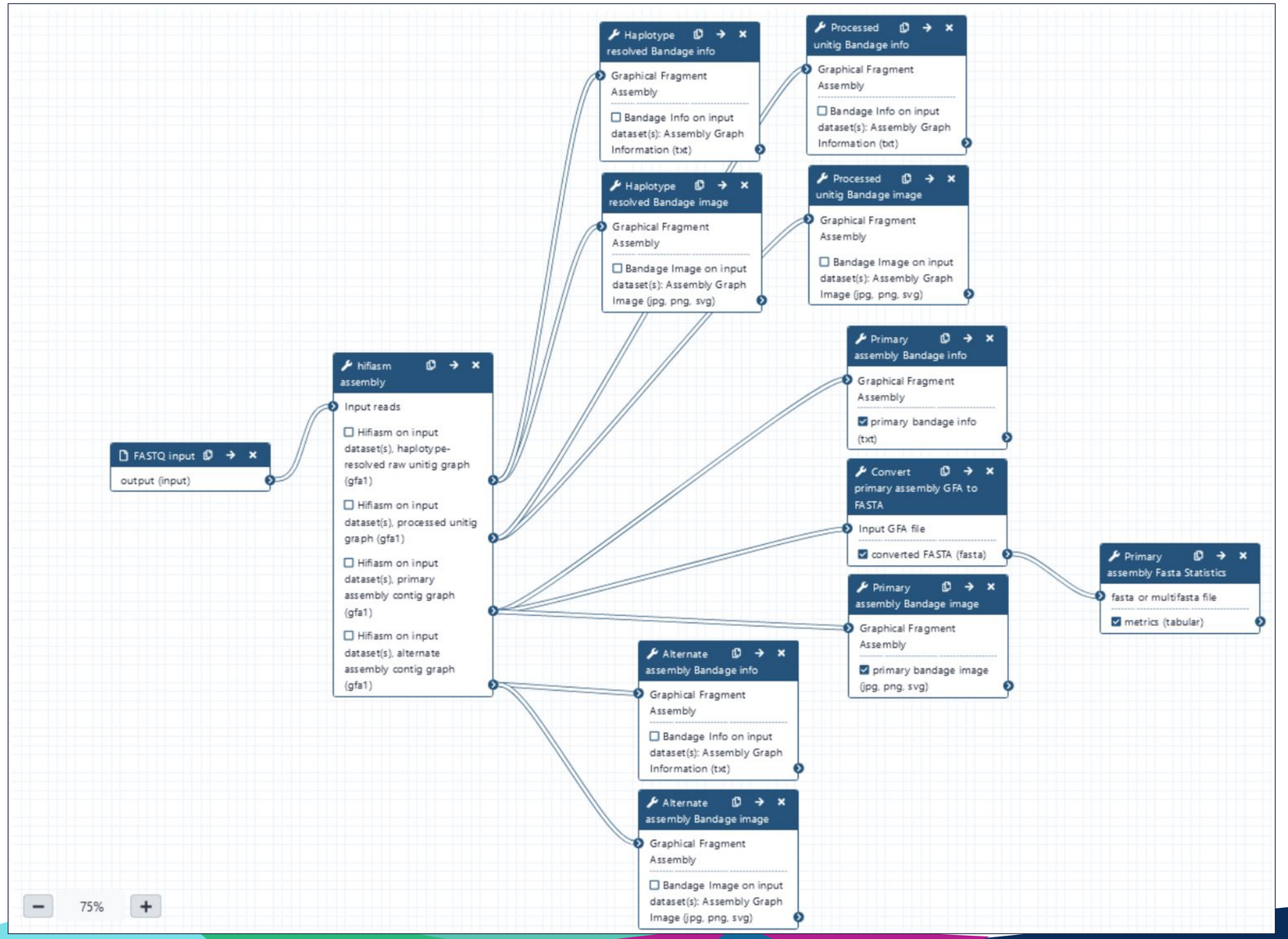
20,000+
Users on Galaxy Australia

11M+
User datasets

Australian BioCommons

Workflows

Workflows are important



Australian BioCommons

Workflows are important

321

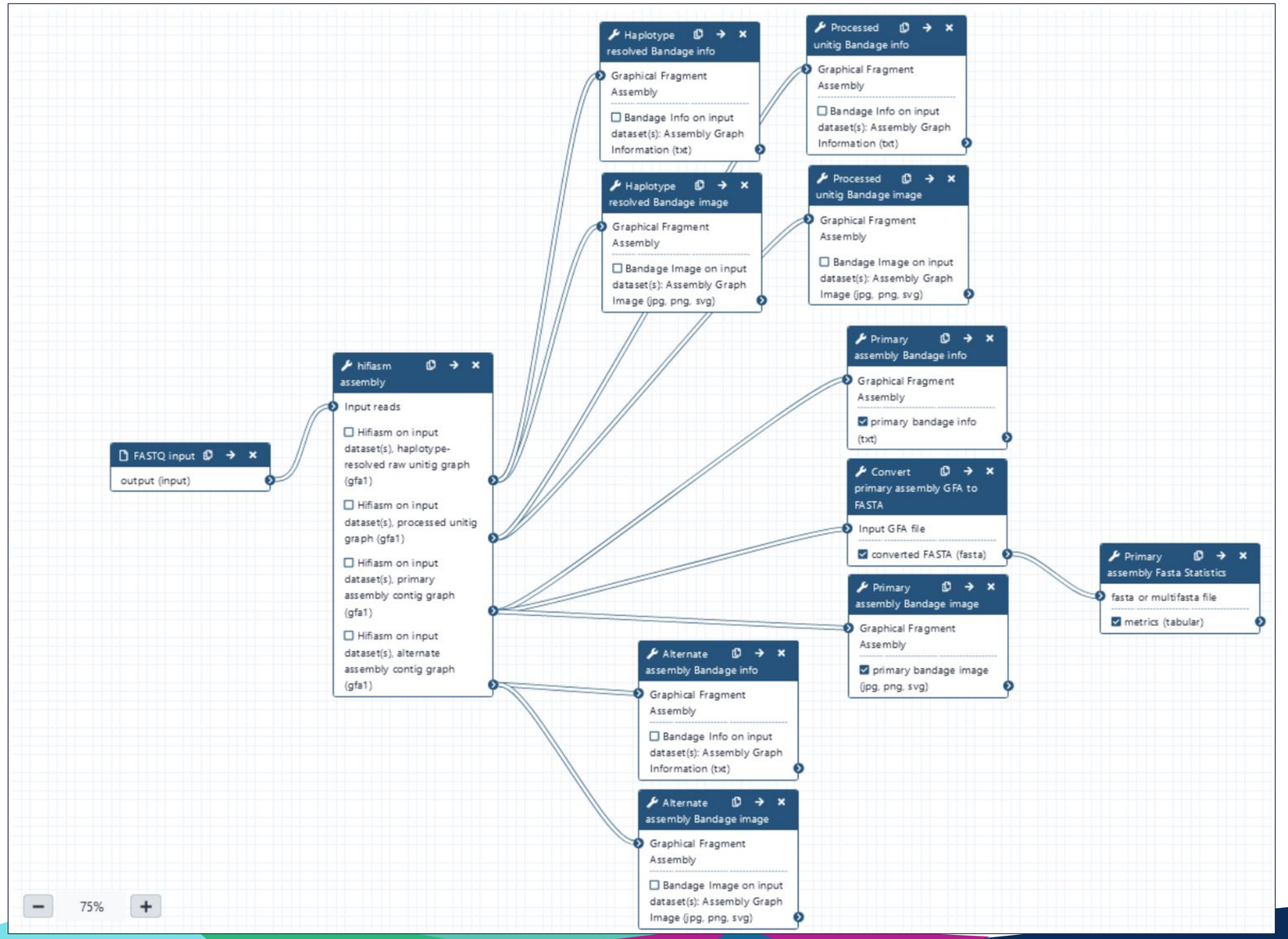
workflow languages

https://s.apache.org/existing-workflow-systems

440,068

clones for nf-core workflows

https://nf-co.re/stats



Australian BioCommons

Workflows exist in significant numbers

1,829

workflows in Dockstore
(WDL, CWL, NextFlow, Galaxy)

https://dockstore.org/search?entryType=workflows&searchMode=files

268

workflows in WorkflowHub
(all public)

https://workflowhub.eu/workflows

Australian BioCommons

Workflows exist in significant numbers

1,829

workflows in Dockstore
(WDL, CWL, NextFlow, Galaxy)

https://dockstore.org/search?entryType=workflows&searchMode=files

753

# of workflows on
Galaxy Europe instance

https://usegalaxy.eu/api/workflows

268

workflows in WorkflowHub
(all public)

https://workflowhub.eu/workflows

Australian
BioCommons

Including locally

1,829

workflows in Dockstore
(WDL, CWL, NextFlow, Galaxy)

https://dockstore.org/search?entryType=workflows&searchMode=files

753

# of workflows on
Galaxy Europe instance

https://usegalaxy.eu/api/workflows

268

workflows in WorkflowHub
(all public)

https://workflowhub.eu/workflows

303

# of workflows on
Galaxy Australia instance

https://usegalaxy.org.au/api/workflows

Australian
BioCommons

Why should I create a workflow?

Workflows can be ...

- Complex
- Time consuming
- Maintenance heavy
- Difficult to find
- Difficult to redeploy
- Duplicated by many

Workflows can be … should be …

- Complex
- Time consuming
- Maintenance heavy
- Difficult to find
- Difficult to redeploy
- Duplicated by many

- Findable
- Reusable
- Citable

Australian BioCommons

Workflows can be … should be …

… and we would say …

- Complex
- Time consuming
- Maintenance heavy
- Difficult to find
- Difficult to redeploy
- Duplicated by many

- Findable
- Reusable
- Citable

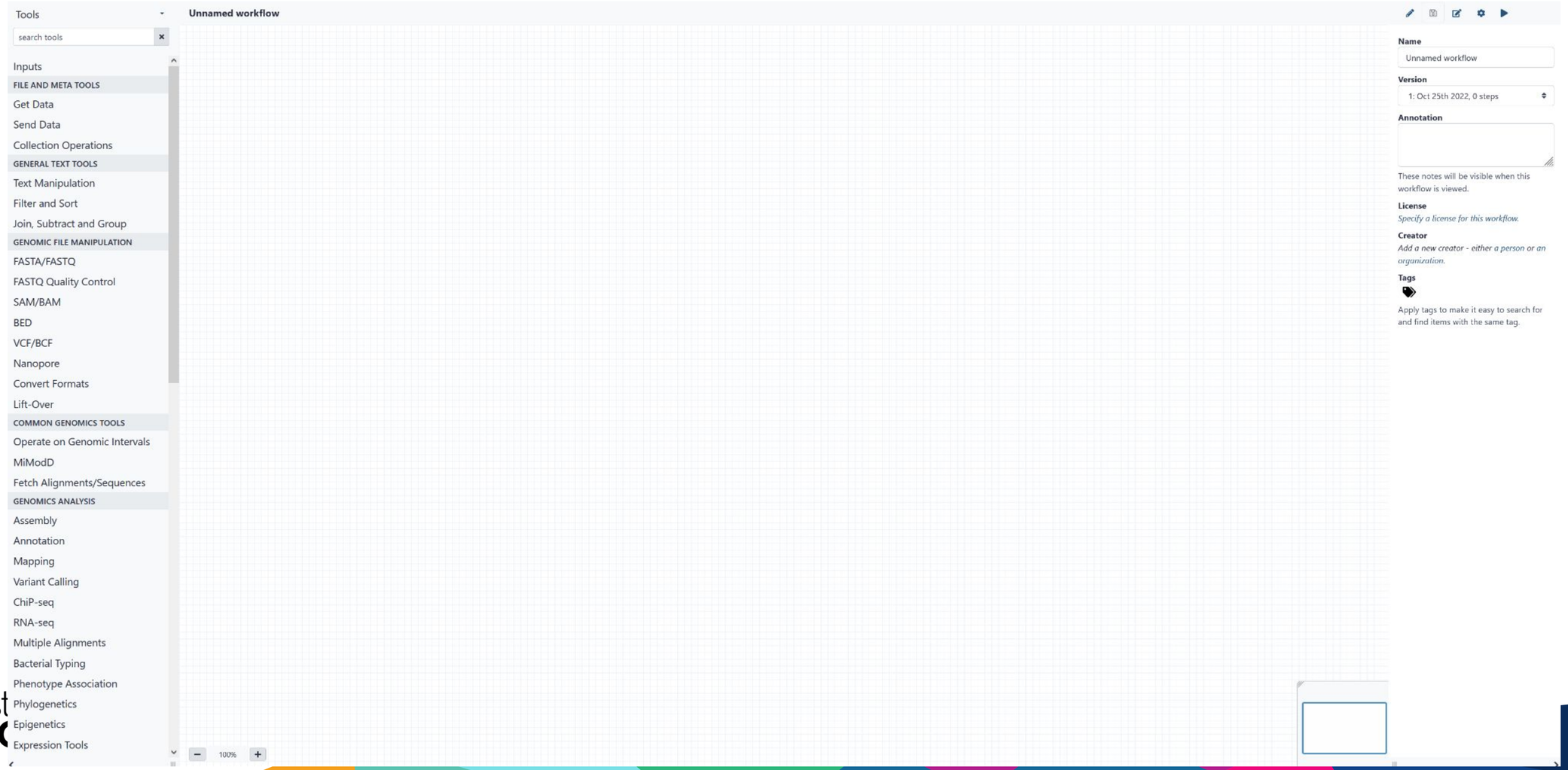
- Straightforward to construct
- Compute agnostic
- User friendly

Australian BioCommons

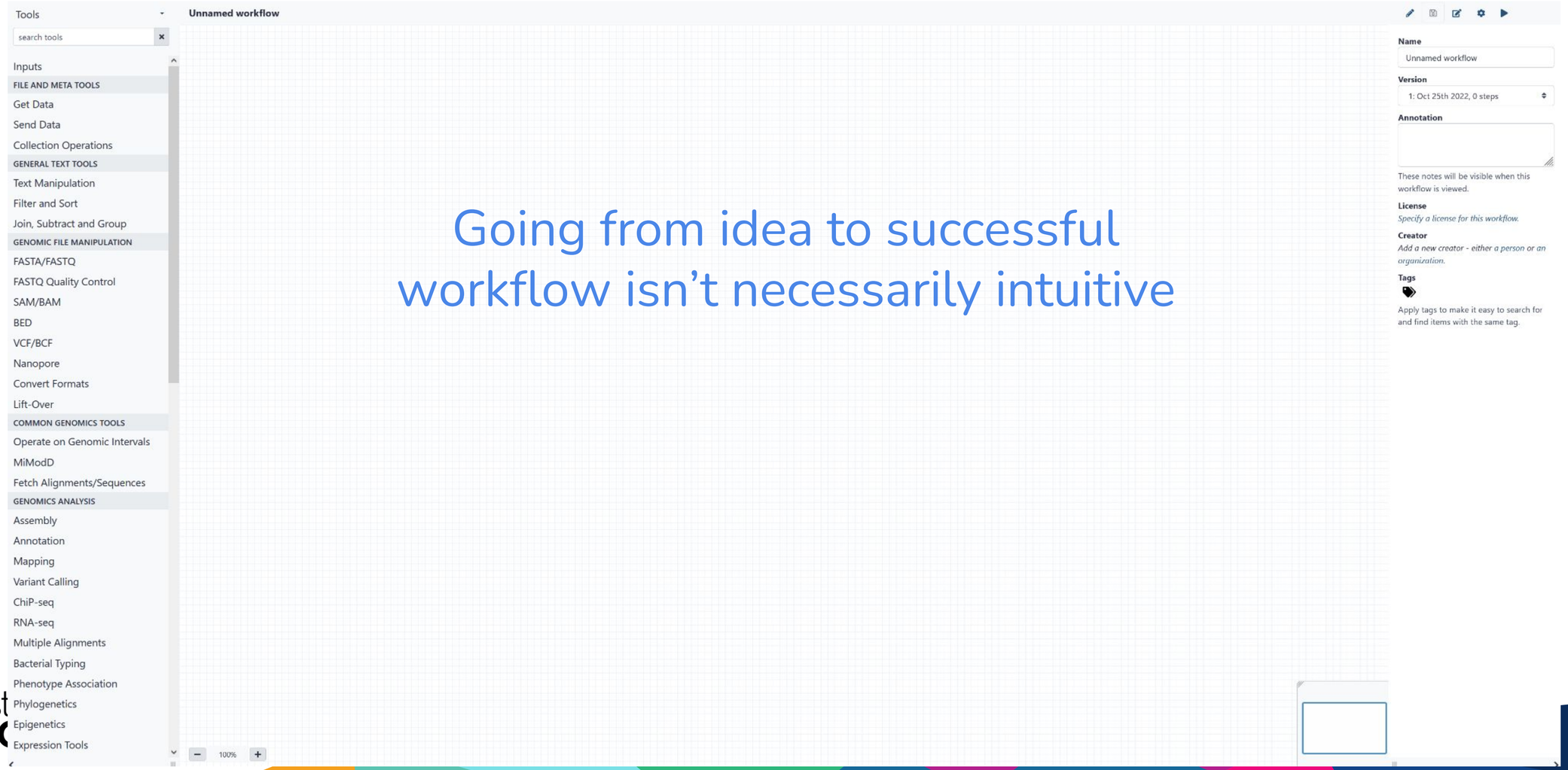
Why should I create a workflow?

Short term investment in intellect
Long term gain in reproducibility
Long term gain in visibility & citation

Galaxy can be part of the solution to these challenges



Galaxy can be part of the solution to these challenges



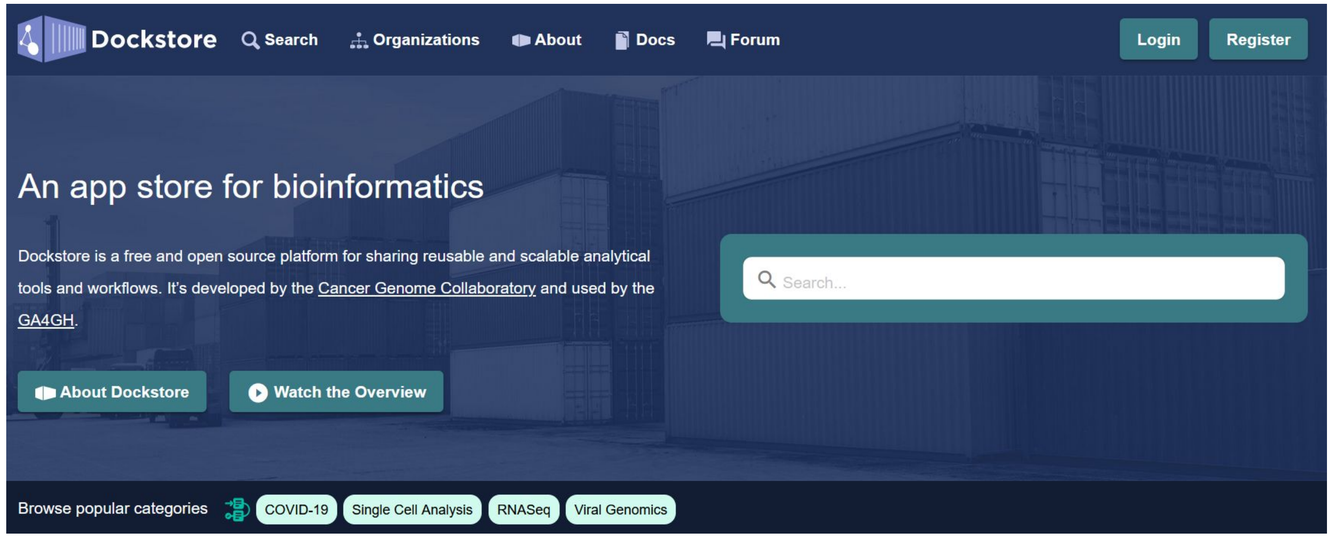
Going from idea to successful workflow isn't necessarily intuitive

How do I get started?

Option 1 might be trying to find a workflow to reuse

Two key international options

Dockstore

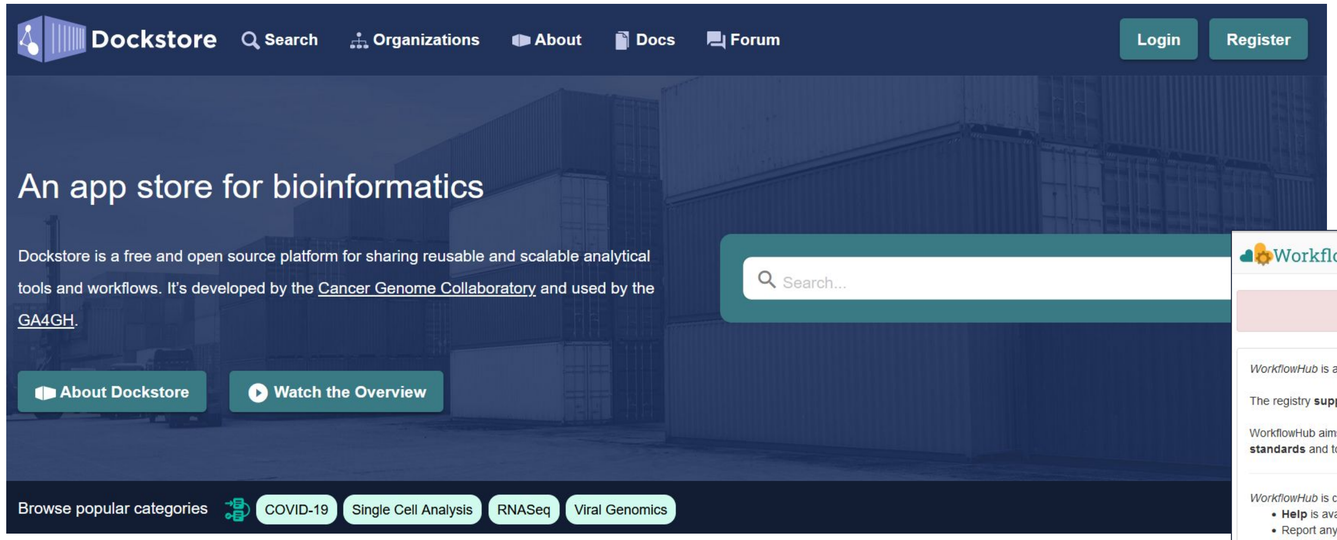


https://dockstore.org/

Australian BioCommons

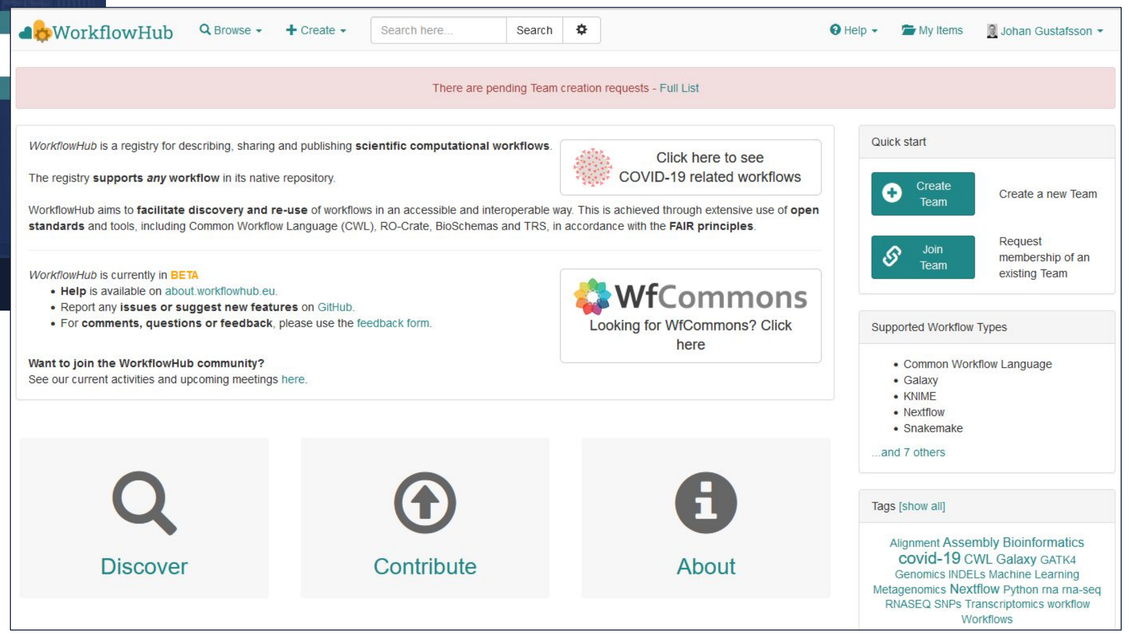
Two key international options

Dockstore



https://dockstore.org/

WorkflowHub



https://about.workflowhub.eu/docs/getting-started/

https://workflowhub.eu/

Australian BioCommons

WorkflowFinder

Copy CSV

Search: assembly

Workflow information					Citation & credits		Deployment	
Workflow name (+ links)	Topic(s)	Operation(s)	License	Updated	DOI	Contributing team(s)	User guide	Open workflow (if available)
Assembly polishing WorkflowHub	Sequence assembly	Sequence assembly	Apache-2.0	2022-09-16	10.48546/workflowhub.workflow.226.1	Galaxy Australia Australian BioCommons	See user guide	Open workflow on Galaxy Australia
Assembly with Flye WorkflowHub	Sequence assembly	De-novo assembly	Apache-2.0	2022-09-16	10.48546/workflowhub.workflow.225.1	Galaxy Australia Australian BioCommons	See user guide	Open workflow on Galaxy Australia
Assess genome quality WorkflowHub	Sequence assembly		Apache-2.0	2022-09-16	10.48546/workflowhub.workflow.229.1	Galaxy Australia Australian BioCommons	See user guide	Open workflow on Galaxy Australia
BAM to FASTQ + QC v1.0 WorkflowHub	Sequence assembly	Conversion Sequencing quality control	GPL-3.0	2022-10-17	10.48546/workflowhub.workflow.220.2	Australian BioCommons Galaxy Australia	See user guide	Open workflow on Galaxy Australia
Combined workflows for large genome assembly WorkflowHub			Apache-2.0	2022-09-16	10.48546/workflowhub.workflow.230.1	Galaxy Australia Australian BioCommons	See user guide	Open workflow on Galaxy Australia
Data QC WorkflowHub	Sequence assembly	Sequencing quality control	Apache-2.0	2022-09-16	10.48546/workflowhub.workflow.222.1	Galaxy Australia Australian BioCommons	See user guide	Open workflow on Galaxy Australia

https://australianbiocommons.github.io/2_1_workflows.html

Australian BioCommons

From within Galaxy Australia

Tool Registry Server

GA4GH Tool Registry Server (TRS) Workflow Search

TRS Server: workflowhub.eu

hifi genome assembly

Name	Description	Organization
PacBio HiFi genome assembly using hifiasm v2.1	# PacBio HiFi genome assembly using hifiasm v2.1 ## General usage recommendations Please see the [Genome assembly with hifiasm on Galaxy Australia](https://australianbiocommons.github.io/how-to-guides/genome_assembly/hifi_assembly) guide. ## See [change log](./change_log.md) ## Acknowledgements The workflow & the [doc_guidelines template used](https://github.com/AustralianBioCommons/doc_guidelines) are supported by the Australian BioCommons via Bioplatforms Australia funding, the Australian Research Data Commons (https://doi.org/10.47486/PL105) and the Queensland Government RICF programme. Bioplatforms Australia and the Australian Research Data Commons are enabled by the National Collaborative Research Infrastructure Strategy (NCRIS).	

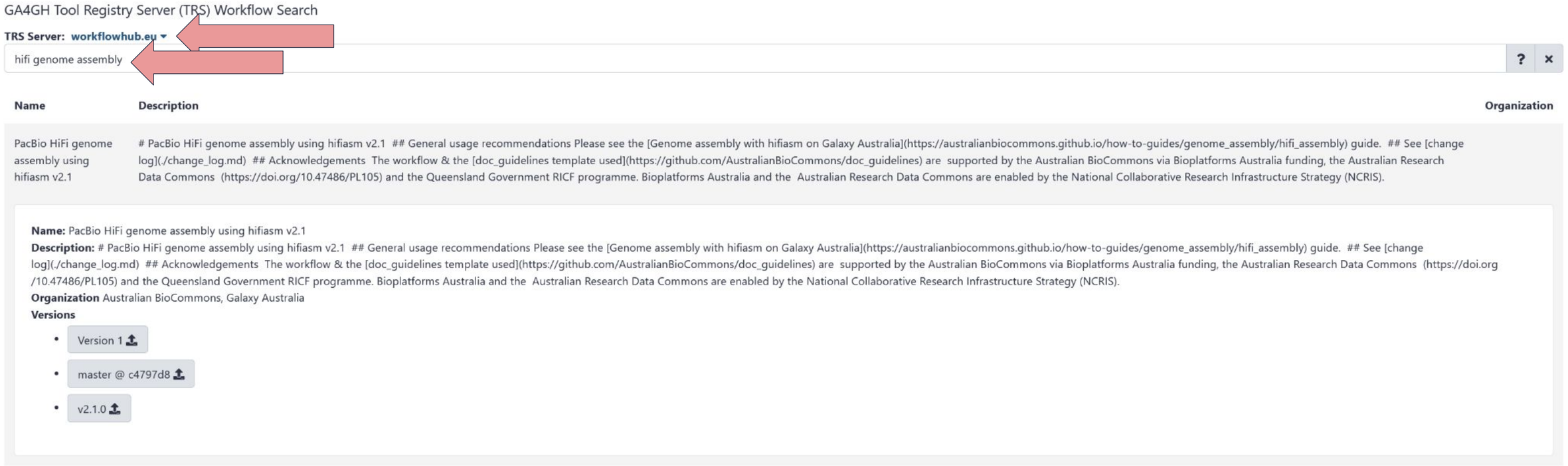
Name: PacBio HiFi genome assembly using hifiasm v2.1
Description: # PacBio HiFi genome assembly using hifiasm v2.1 ## General usage recommendations Please see the [Genome assembly with hifiasm on Galaxy Australia](https://australianbiocommons.github.io/how-to-guides/genome_assembly/hifi_assembly) guide. ## See [change log](./change_log.md) ## Acknowledgements The workflow & the [doc_guidelines template used](https://github.com/AustralianBioCommons/doc_guidelines) are supported by the Australian BioCommons via Bioplatforms Australia funding, the Australian Research Data Commons (https://doi.org/10.47486/PL105) and the Queensland Government RICF programme. Bioplatforms Australia and the Australian Research Data Commons are enabled by the National Collaborative Research Infrastructure Strategy (NCRIS).
Organization Australian BioCommons, Galaxy Australia
Versions

- Version 1
- master @ c4797d8
- v2.1.0

Australian BioCommons

From within Galaxy Australia

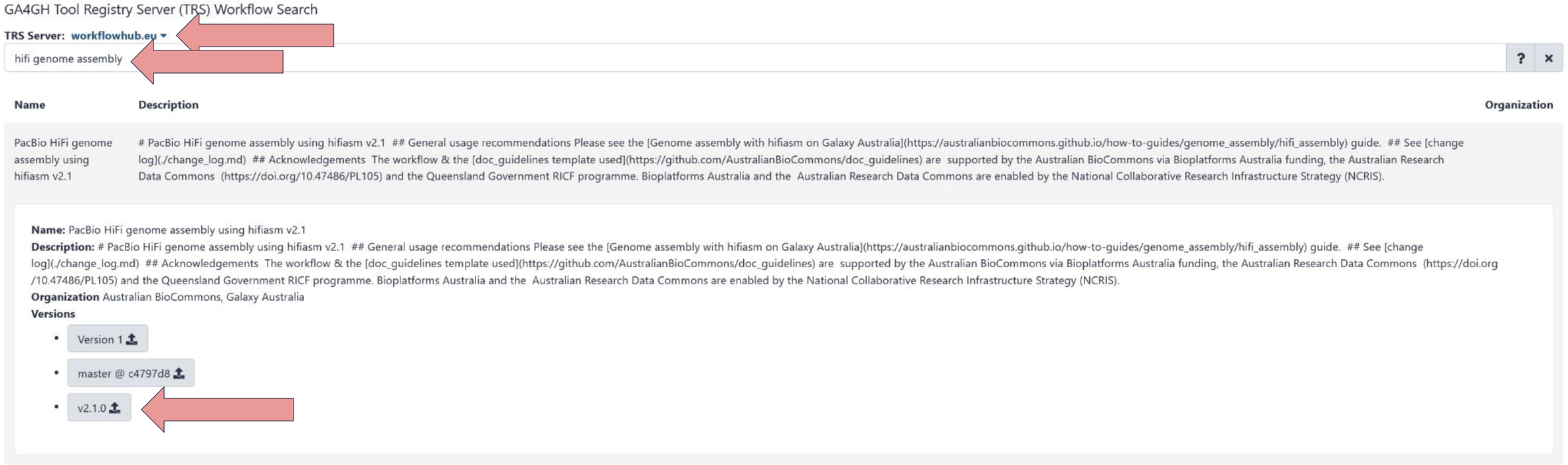
Tool Registry Server



Australian BioCommons

From within Galaxy Australia

Tool Registry Server



Australian BioCommons

Option 2
What if you need to *create* a workflow?

Australian BioCommons

Using Galaxy to build workflows

1

De novo

Add tools sequentially to create a workflow

Using Galaxy to build workflows

1

De novo

Add tools sequentially to create a workflow

2

From a history

Use a set of tools sequentially

Use history to create, curate and finalise a workflow file

Using Galaxy to build workflows

1

De novo

Add tools sequentially to create a workflow

2

From a history

Use a set of tools sequentially

Use history to create, curate and finalise a workflow file

3

From a publication

Read publication

Find / request tools for Galaxy

Create workflow using *de novo* build approach

Australian BioCommons

Using Galaxy to build workflows

1

De novo

Add tools sequentially to create a workflow

Paint the sistine chapel from scratch

2

From a history

Use a set of tools sequentially

Use history to create, curate and finalise a workflow file

“Re-paint” the sistine chapel from previous experience

3

From a publication

Read publication

Find / request tools for Galaxy

Create workflow using de novo build approach

Paint the sistine chapel based on someone else painting an equivalent chapel

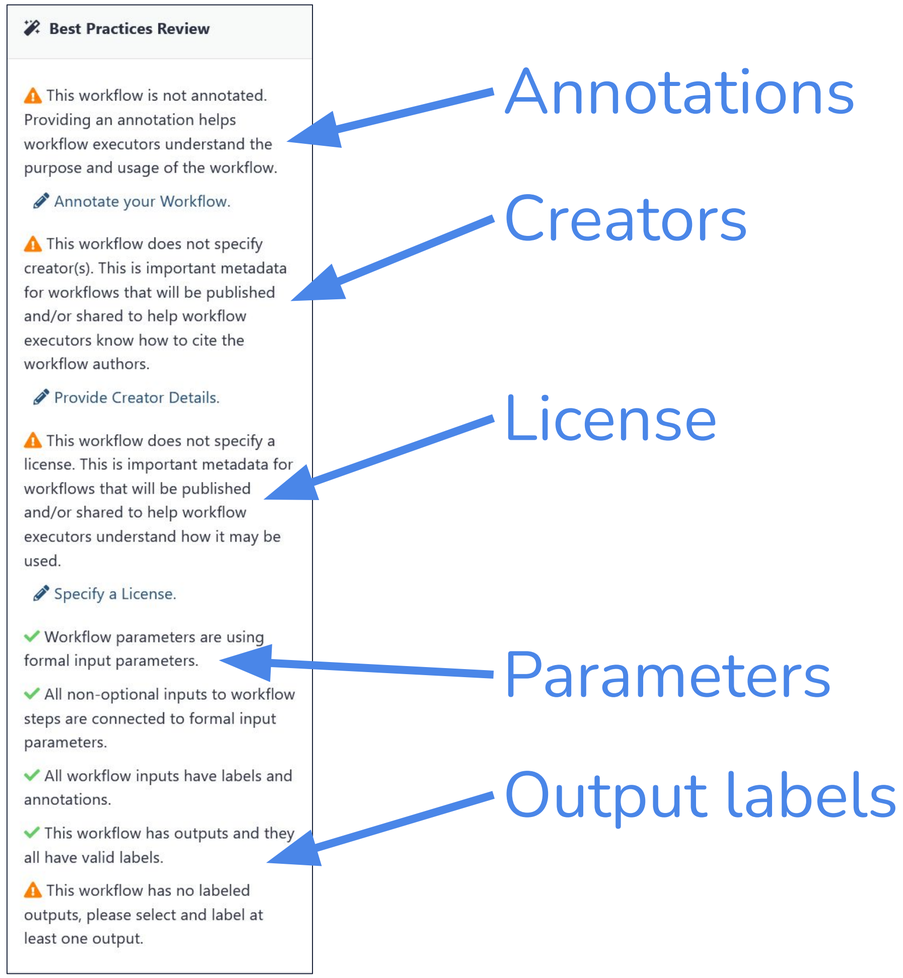
Australian BioCommons

Time for some demos!

Now for some final suggestions

Australian BioCommons

Follow workflow best practices



Annotations

Creators

License

Parameters

Output labels

Australian BioCommons

Follow workflow best practices

Best Practices Review

⚠ This workflow is not annotated. Providing an annotation helps workflow executors understand the purpose and usage of the workflow.

Annotate your Workflow.

⚠ This workflow does not specify creator(s). This is important metadata for workflows that will be published and/or shared to help workflow executors know how to cite the workflow authors.

Provide Creator Details.

⚠ This workflow does not specify a license. This is important metadata for workflows that will be published and/or shared to help workflow executors understand how it may be used.

Specify a License.

✔ Workflow parameters are using formal input parameters.

✔ All non-optional inputs to workflow steps are connected to formal input parameters.

✔ All workflow inputs have labels and annotations.

✔ This workflow has outputs and they all have valid labels.

⚠ This workflow has no labeled outputs, please select and label at least one output.

Annotations

Creators

License

Parameters

Output labels

Publish your workflows

Published Workflows

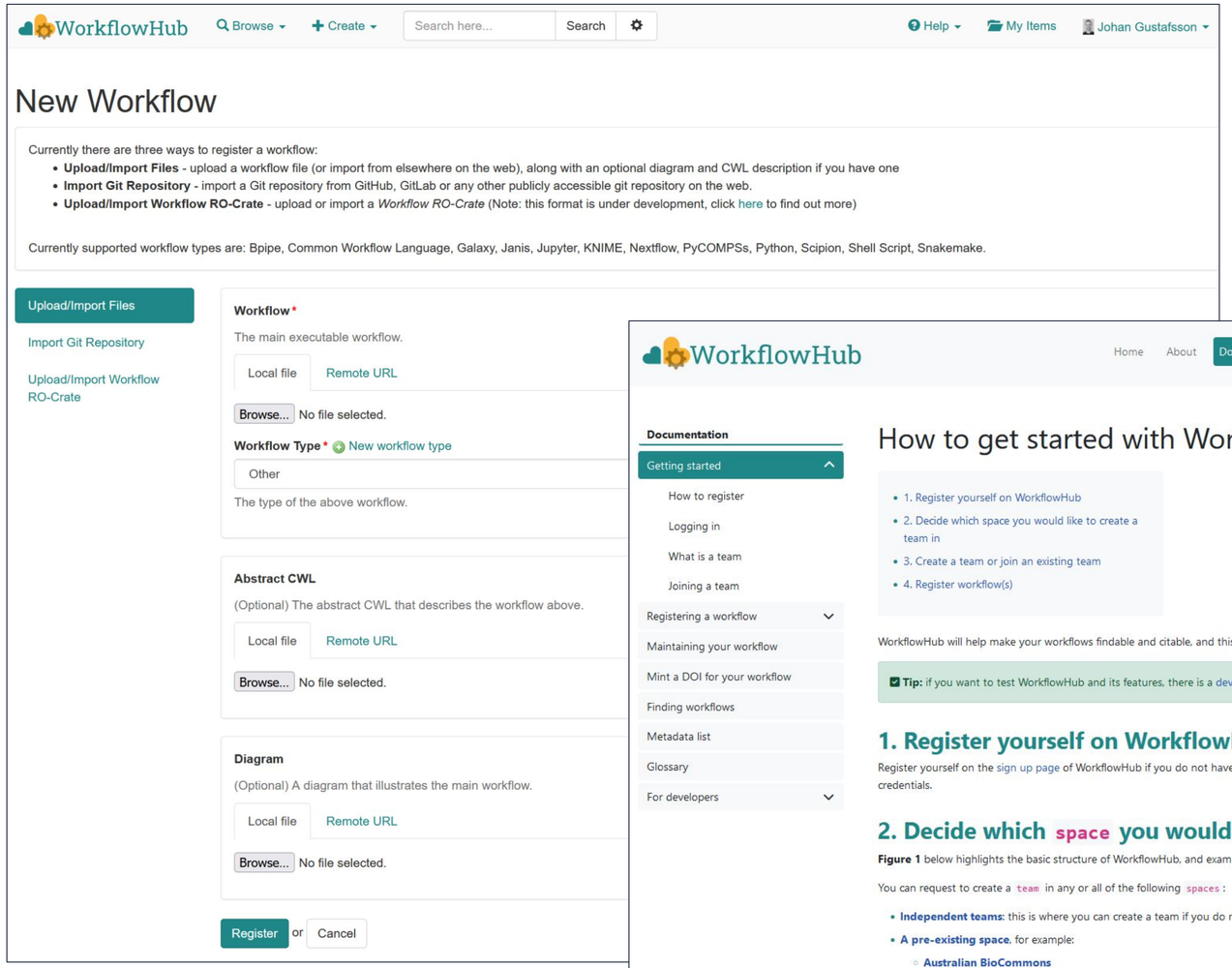
search name, annotation, owner, and

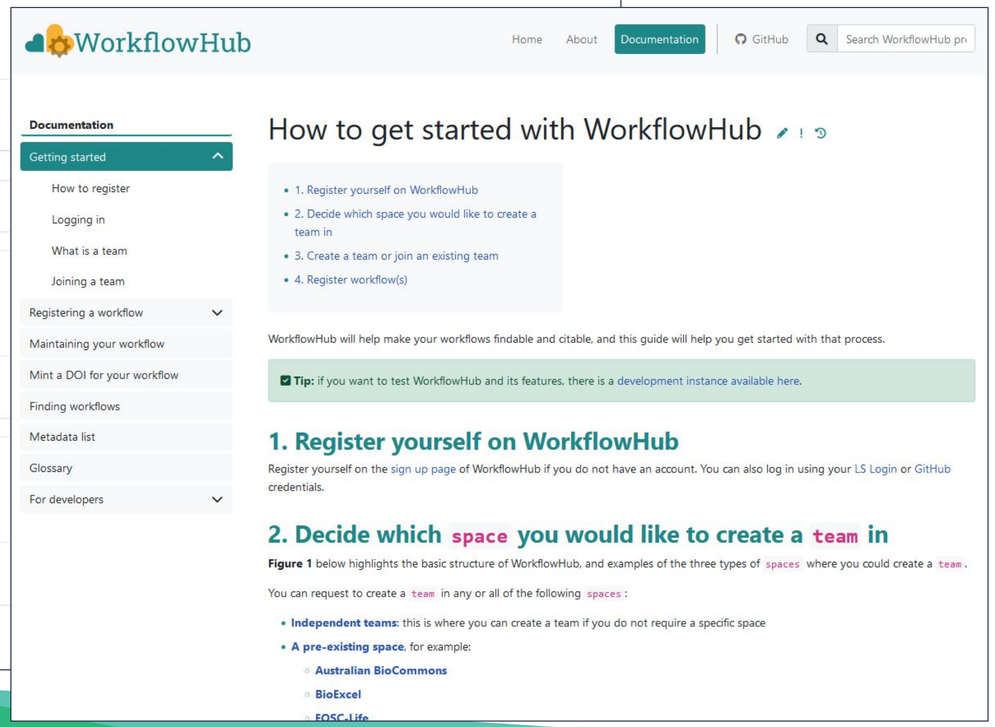
Advanced Search

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated
exploring iris dataset with stats and plots		delaney_burnard			Aug 18, 2022
VGP assembly: training workflow (imported from GTN materials August 4 2022)	VGP assembly tutorial	anna		assembly	Aug 04, 2022
PacBio HiFi genome assembly using hifiasm v1.0	Assembly, visualisation and quality control workflow for high fidelity reads built from circular consensus sequence (PacBio HiFi) data.	johan		fastq genome_assembly hifi hifiasm	Jul 19, 2022
Generic Assembly	This workflow has been designed to take paired end reads from an Illumina platform, perform some QC and then assemble them into contigs. It is mainly for...	simongladman		assembly bacterial	Jul 18, 2022
GA-virReport-Stats (Public)	Downstream workflow for positive detections from GA-virReport to compute mapping statistics to a reference genome and to generate reference guided genome...	lelwala		ga-virreport-confidence	Jun 28, 2022
GA-VirReport-Single_input_files (Public)	Bioinformatics workflow for detecting viruses and viroids from host plant small RNA sequencing data	lelwala		ga-virreport	Jun 28, 2022
GA-VirReport-Multiple_input_files_(Public)	Bioinformatics workflow for detecting viruses and viroids from host plant small RNA sequencing data	lelwala		ga-virreport	Jun 28, 2022
Generic consensus construction from VCF calls	Build a consensus sequence from FILTER PASS variants with intrasample allele-frequency above a configurable consensus threshold. Hard-mask regions with...	galaxy-australia		generic mbv	Jun 06, 2022
Genetic variation analysis reporting	This workflow takes a VCF dataset of variants produced by any of the variant calling workflows in...	galaxy-australia		generic mpvx	Jun 06, 2022
Generic variation analysis on WGS PE data	This workflows performs paired end read mapping with bwa-mem followed by sensitive variant calling across a wide range of AFs with lofreq	galaxy-australia		mpxv generic	Jun 06, 2022
Di Marzio Workflow		dimarziogabriel			Jun 04, 2022
Stacks RAD-seq de novo workflow	Version 1	anna			May 31, 2022
Stacks RAD-seq reference-guided workflow	Version 1	anna			May 31, 2022
Partial ref-guided workflow - bwa mem only	Version 1	anna			May 19, 2022
Partial ref-guided workflow - gstacks and pops	Version 1	anna			May 18, 2022
Partial de novo workflow: c-s-g-pops only	Version 1	anna			May 18, 2022
Partial de novo workflow: ustacks only	Version 1	anna			May 18, 2022
QC of RADseq reads	Version 1	anna			May 18, 2022

Australian BioCommons

Register your workflows





Australian BioCommons

Reference your workflows

Purge duplicates from hifiasm assembly v1.0 v1.0.0 (latest)

View on GitHub | Request Contact | Unsubscribe | Download RO Crate | Run on usegalaxy.eu | Add new | Actions

Overview | Files | Related items

Workflow Type: Galaxy

Stable

Purge-duplicates-from-hifiasm-assembly

General recommendations for using `Purge-duplicates-from-hifiasm-assembly`

Please see the `Genome assembly with hifiasm on Galaxy Australia` guide.

Acknowledgements

The workflow & the doc_guidelines template used are supported by the Australian BioCommons via Bioplatforms Australia funding, the Australian Research Data Commons (https://doi.org/10.47486/PL105) and the Queensland Government RICF programme. Bioplatforms Australia and the Australian Research Data Commons are enabled by the National Collaborative Research Infrastructure Strategy (NCRIS).

SEEK ID: https://workflowhub.eu/workflows/237?version=2

DOI: 10.48546/workflowhub.workflow.237.2

Inputs

ID	Name	Description	Type
HiFiASM 1o assembly	HiFiASM 1o assembly	*n/a*	*n/a*
HiFi reads as FASTQ	HiFi reads as FASTQ	*n/a*	*n/a*

Creators and Submitter

Creator

Gareth Price

Additional credit

Gareth Price

Submitter

Johan Gustafsson

Citation Copy

Price, G. (2022). *Purge duplicates from hifiasm assembly v1.0*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.237.2

American Psychological Association 7th

License

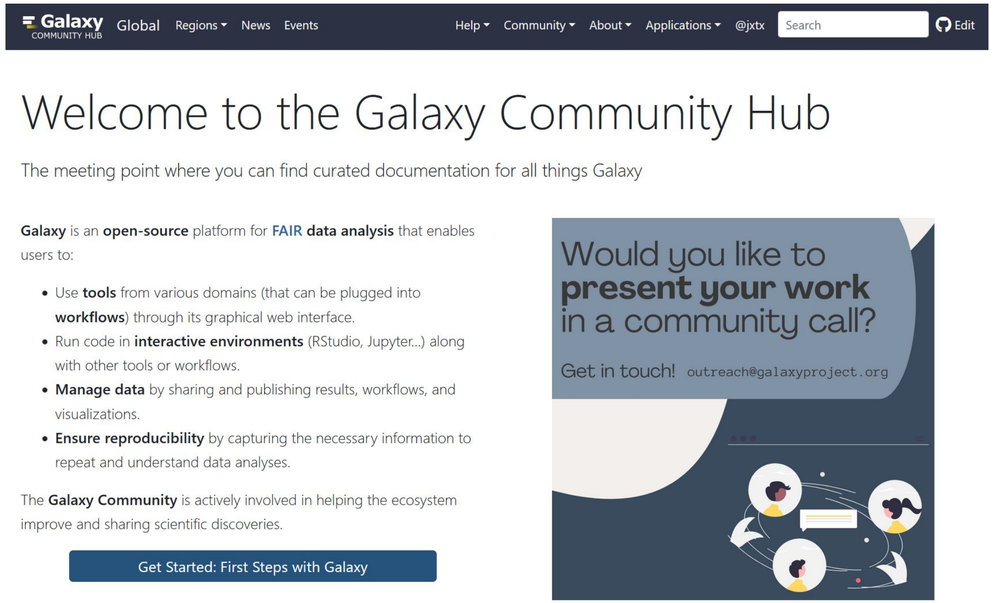
GNU General Public License 3.0

Citation Copy

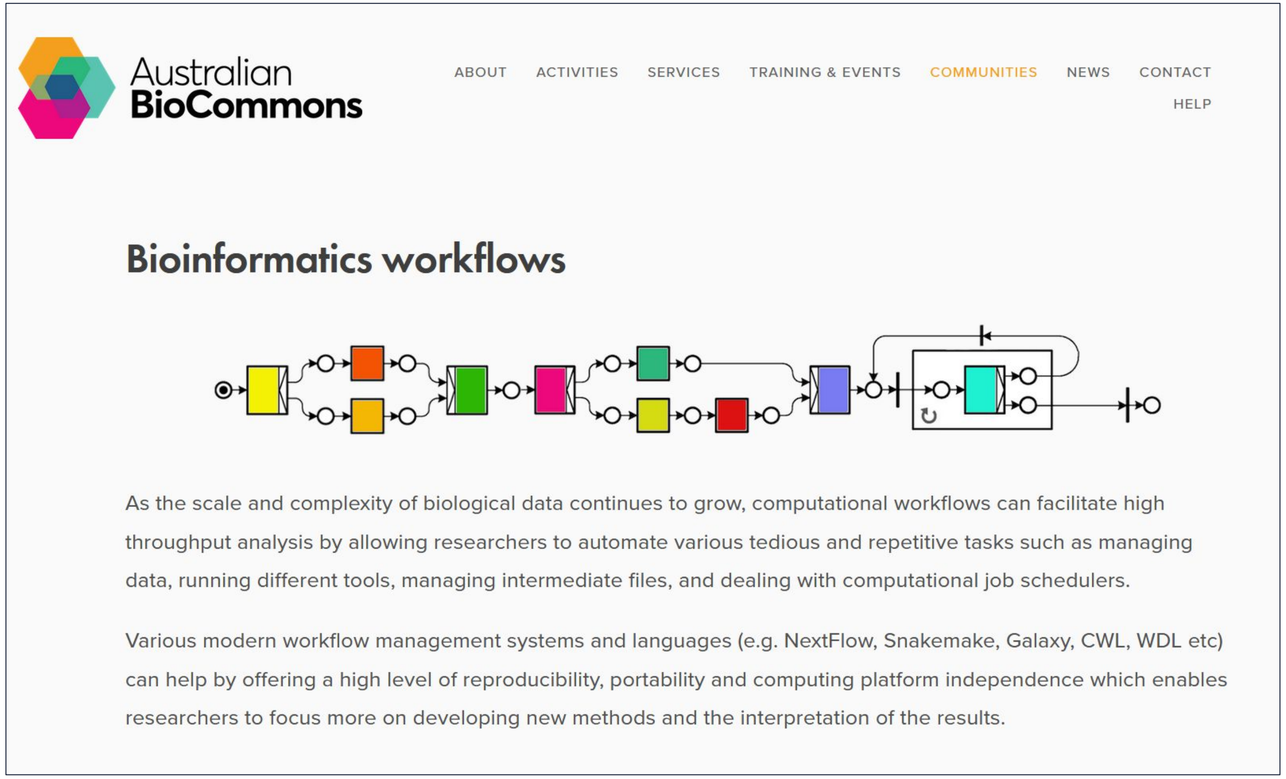
Price, G. (2022). *Purge duplicates from hifiasm assembly v1.0*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.237.2

Australian BioCommons

Staying Connected through Community



https://galaxyproject.org



https://www.biocommons.org.au/workflows

Australian BioCommons

Thanks!

Any questions?

You can email me at: johan@biocommons.org.au

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