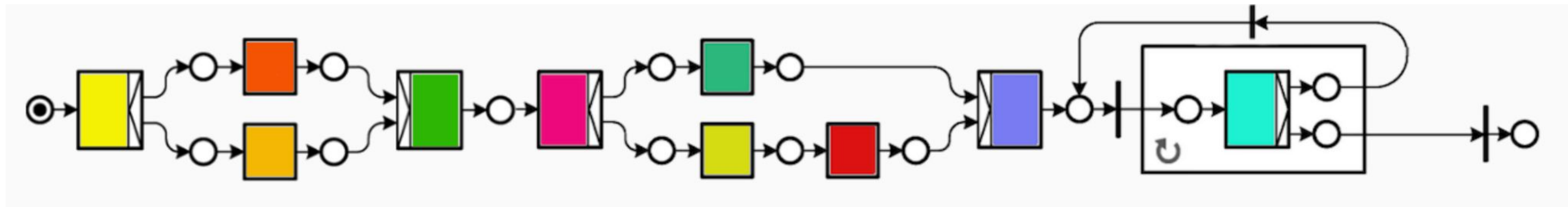
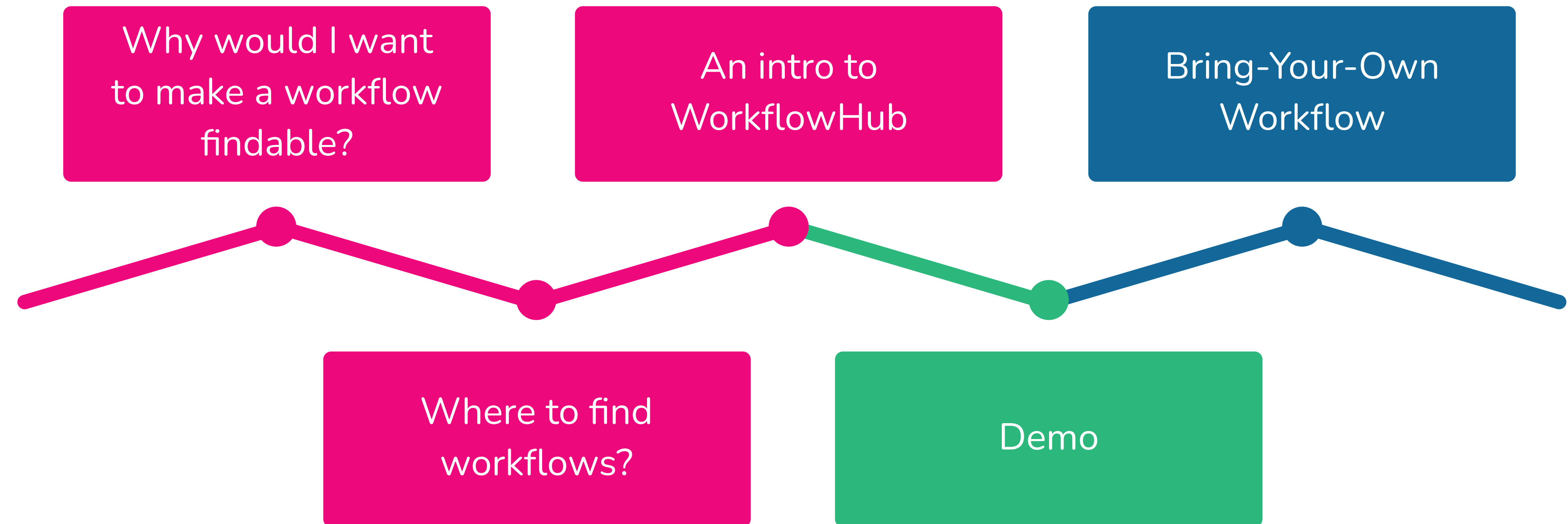


Make your bioinformatics workflows findable and citable



Johan Gustafsson, Australian BioCommons
Georgina Samaha, Sydney Informatics Hub

Today



Learning outcomes

Register with
WorkflowHub

Create and join
teams and spaces

Register
workflows

Maintain your
workflows

Create a DOI

Cite workflows



**Why make
workflows
findable?**

1

Workflows are important

GalaxyProject SARS-CoV-2 analysis effort

<https://galaxyproject.org/projects/covid19/>

PLOS PATHOGENS

OPINION
No more business as usual: Agile and effective responses to emerging pathogen threats require open data and open analytics

Damon Baker¹, Marius van den Beek², Daniel Blankenberg³, Dave Bouvier⁴, John Chilton⁵, Nate Connor⁶, Frederik Coppens^{6,7}, Isaac Eguinol^{8,9}, Simon Gladman¹⁰, Björn Grönberg¹¹, Nicholas Keaner¹², Delphine Larivière¹³, Andrew Linton¹⁴, Sergei Kosakovsky Pond¹⁵, Wolfgang Maier¹⁶, Anton Nekrutko¹⁷, James Taylor¹⁸, Steven Weiser¹⁹

Abstract
The current state of much of the Wuhan pneumonia virus (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]) research shows a regrettable lack of data sharing and considerable analytical obfuscation. This impedes global research cooperation, which is essential for tackling public health emergencies and requires unimpeded access to data, analysis tools, and computational infrastructure. Here, we show that community efforts in developing open analytical software tools over the past 10 years, combined with national investments into scientific computational infrastructure, can overcome these deficiencies and provide an accessible platform for tackling global health emergencies in an open and transparent manner. Specifically, we use all SARS-CoV-2 genomic data available in the public domain so far to (1) underscore the importance of access to raw data and (2) demonstrate that existing community efforts in curation and deployment of bioinformatic software can reliably support rapid, reproducible research during global health crises. All our analyses are fully documented at <https://github.com/galaxyproject/SARS-CoV-2>.

Introduction
The initial publications describing genomic features of SARS-CoV-2 [1–4] used Illumina and Oxford nanopore data to elucidate the sequence composition of patient specimens (although only Wu and colleagues [1] explicitly provided the accession numbers for their raw short-read sequencing data). However, their approaches to processing, assembly, and analysis of raw data differed widely (Table 1) and ranged from transparent [1] to entirely opaque [4]. Such lack of analytical transparency sets a dangerous precedent. Infectious disease outbreaks often occur

Cell Article

The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages

Graphical Abstract

Abstract
The emergence of the SARS-CoV-2 N501Y lineages is a key event in the evolution of the virus. We analyze the emergence and ongoing convergent evolution of these lineages. We find that the N501Y mutation is a key driver of the evolution of these lineages. We also find that the N501Y mutation is highly conserved across different lineages. This suggests that the N501Y mutation is a key driver of the evolution of these lineages. We also find that the N501Y mutation is highly conserved across different lineages. This suggests that the N501Y mutation is a key driver of the evolution of these lineages.

Highlights

- Detected a major global shift in the SARS-CoV-2 selective landscape in late 2020
- Identified ongoing convergent evolution between the alpha, beta, and gamma lineages
- Defined the mutational meta-signature upon which these lineages are converging

correspondence

Ready-to-use public infrastructure for global SARS-CoV-2 monitoring

To the Editor—The COVID-19 pandemic in the first health crisis characterized by large amounts of genomic data. Computational infrastructure can be a bottleneck for data analysis, amplifying global inequalities in ability to track SARS-CoV-2 evolution. This is an issue even in developed countries, as computational infrastructure requires expertise in resource procurement, configuration and maintenance. Commercial computational clouds do not fully address the problem because these resources must still be configured and funded. Furthermore, commercial clouds are predominantly US-based and many countries have policies making payments to foreign providers impractical. In developing countries, research computing infrastructure is rare and researchers often cannot afford commercial cloud-based computing. Here, we present the COVID-19 effort by the Galaxy Project, which provides worldwide public computational infrastructure, making the analysis of deep sequencing data accessible to anyone while also providing an analytical framework for global pathogen genomic surveillance based on raw sequencing read data.

Abstract
The programmed frameshift element (PFE) reinitiating translation from ORF1a to ORF1b is essential for the propagation of coronaviruses. The combination of genomic features that make up PFE—the overlap between the two reading frames, a slippery sequence, as well as an ensemble of complex secondary structure elements—places severe constraints on this region as most possible nucleotide substitution may disrupt one or more of these elements. The vast amount of SARS-CoV-2 sequencing data generated within the past year provides an opportunity to assess the evolutionary dynamics of PFE in great detail. Here, we performed a comparative analysis of all available coronavirus genomic data available to date. We show that the overlap between ORF1a and ORF1b evolved as a set of discrete 7, 16, 22, 25, and 31 nucleotide stretches with a well-defined phylogenetic specificity. We further examined sequencing data from over 1,500,000 complete genomes and 55,000 raw read data sets to demonstrate exceptional conservation and detect signatures of selection within the PFE region.

Key words: SARS-CoV-2, frameshift, conservation.

Stepwise Evolution and Exceptional Conservation of ORF1a/b Overlap in Coronaviruses

Han Mei¹, Sergei Kosakovsky Pond², and Anton Nekrutko¹

Abstract
The programmed frameshift element (PFE) reinitiating translation from ORF1a to ORF1b is essential for the propagation of coronaviruses. The combination of genomic features that make up PFE—the overlap between the two reading frames, a slippery sequence, as well as an ensemble of complex secondary structure elements—places severe constraints on this region as most possible nucleotide substitution may disrupt one or more of these elements. The vast amount of SARS-CoV-2 sequencing data generated within the past year provides an opportunity to assess the evolutionary dynamics of PFE in great detail. Here, we performed a comparative analysis of all available coronavirus genomic data available to date. We show that the overlap between ORF1a and ORF1b evolved as a set of discrete 7, 16, 22, 25, and 31 nucleotide stretches with a well-defined phylogenetic specificity. We further examined sequencing data from over 1,500,000 complete genomes and 55,000 raw read data sets to demonstrate exceptional conservation and detect signatures of selection within the PFE region.

Key words: SARS-CoV-2, frameshift, conservation.

Maier W, Bray S, et al. Ready-to-use public infrastructure for global SARS-CoV-2 monitoring. 10.1038/s41587-021-01069-1

Link	Workflow	Inputs	Outputs	Aligner	Caller
WorkFlowHub DockStore	ILLUMINA ARTIC: Variant analysis from ampliconic data produced with ARTIC protocol v1, v2, v3, or v4, or any alternative primer scheme. ILL-AMP	1. Paired reads [fastqsanger] 2. SARS-CoV-2 reference [fasta] 3. Primer coordinates [bed] 4. Primer pairs table [tsv]	Variants [vcf]	BWA MEM	Iofreq
WorkFlowHub DockStore	OXFORD NANOPORE ARTIC: Variant analysis from ampliconic data produced with ARTIC protocol v1, v2, v3, or v4, or any alternative primer scheme. ONT-AMP	1. Reads [fastqsanger] 2. SARS-CoV-2 reference [fasta] 3. Primer coordinates [bed]	Variants [vcf]	minimap2	medaka
WorkFlowHub DockStore	ILLUMINA METATRANSCRIPTOMIC PE: Variant analysis from metatranscriptomic data. ILL-MT-PE	1. Paired reads [fastqsanger] 2. SARS-CoV-2 reference [fasta]	Variants [vcf]	BWA MEM	Iofreq

Workflows are important

326

workflow languages

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

1.34 M

clones for nf-core workflows

<https://nf-co.re/stats>

Workflows are important

326

workflow languages

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

2,262

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

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

874

of workflows on
Galaxy Europe instance

<https://usegalaxy.eu/api/workflows>

1.34 M

clones for nf-core
workflows

<https://nf-co.re/stats>

298

workflows in WorkflowHub
(all public)

<https://workflowhub.eu/workflows>

329

of workflows on
Galaxy Australia instance

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

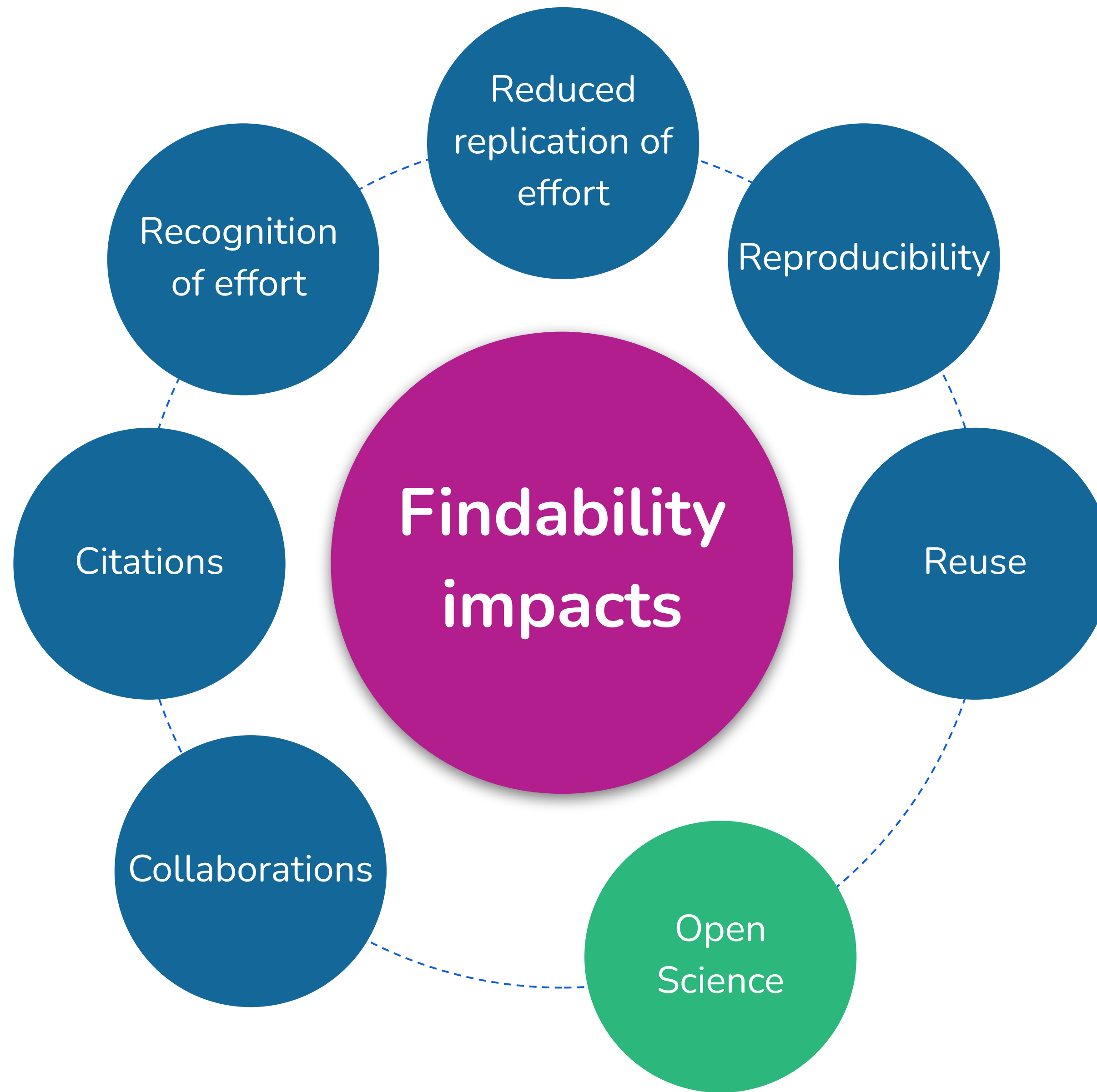
Workflows can also be ...

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

... and should be ...

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

- ☑ Findable
- ☑ Reusable
- ☑ Citable

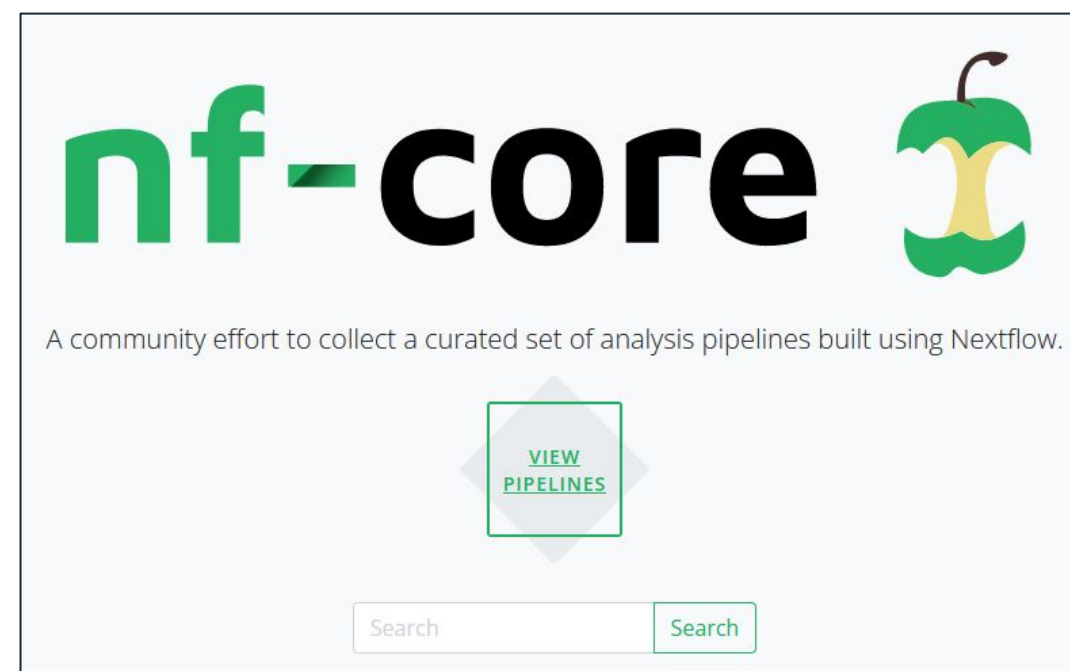




So where can I
find workflows? 2

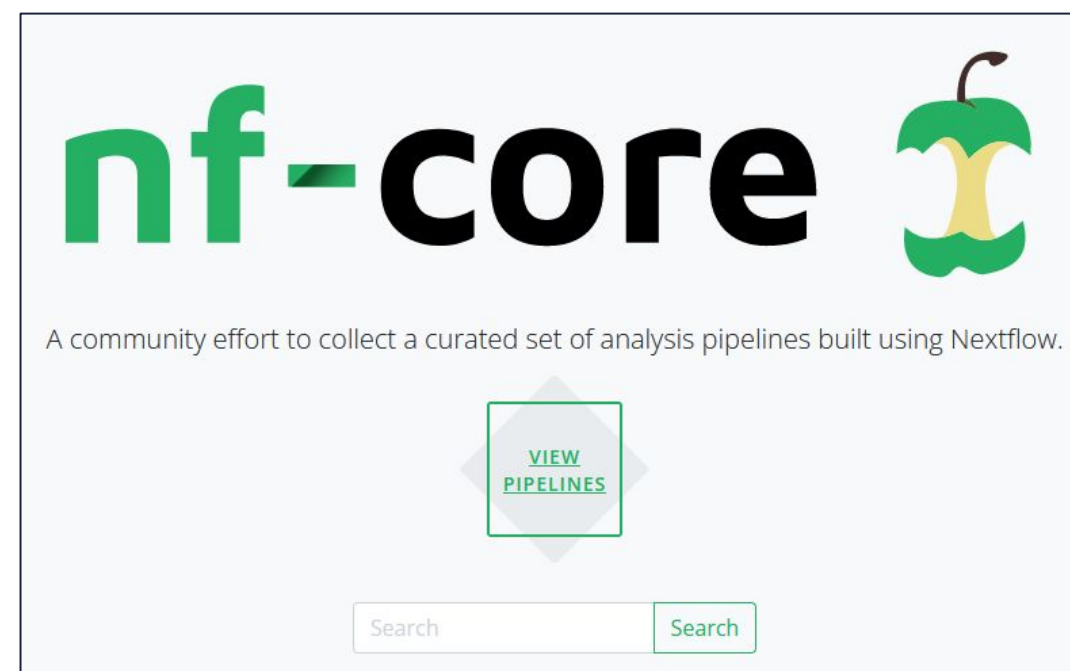
Multiple places you *could* find workflows

Platforms / community repositories

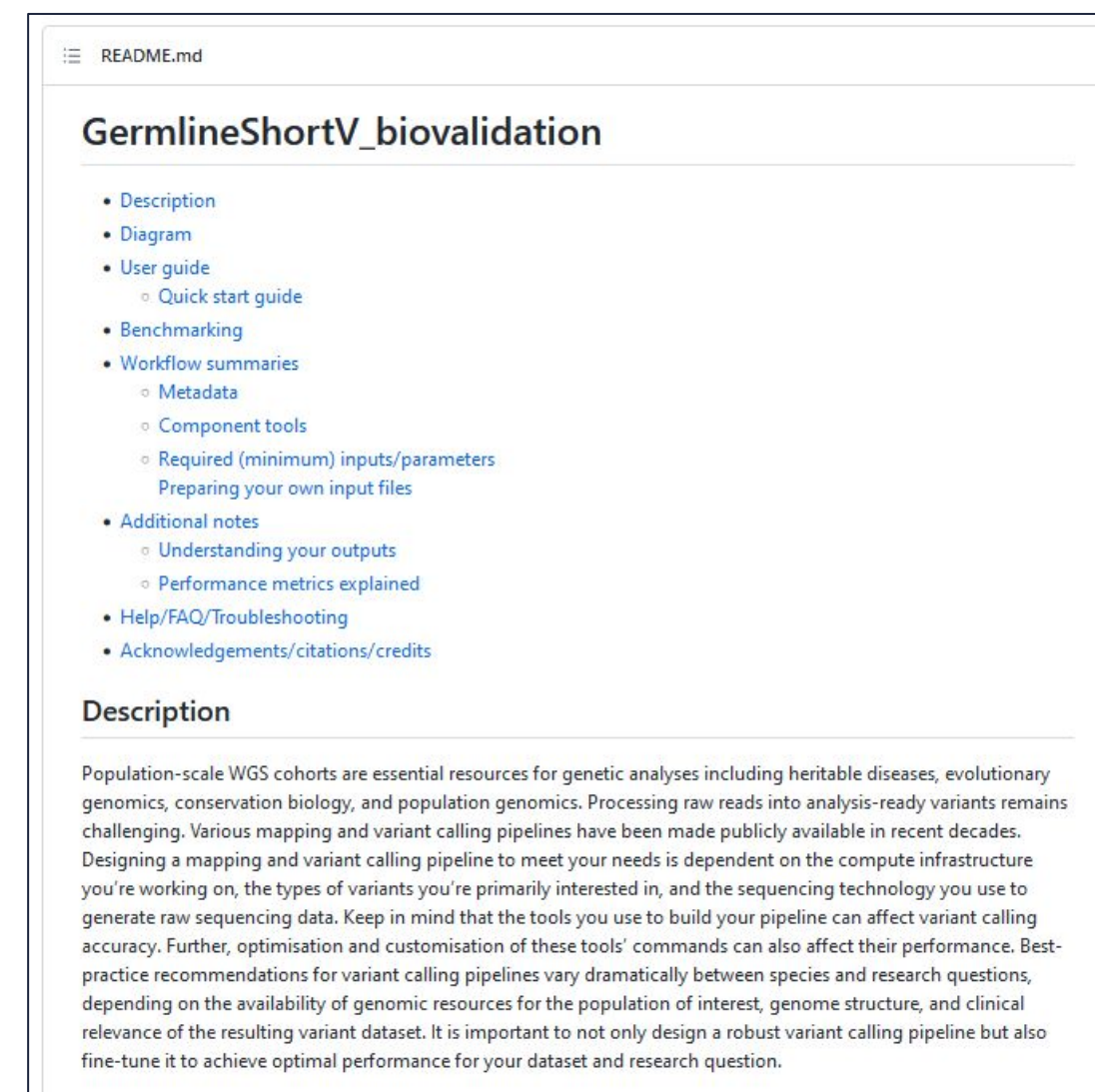


Multiple places you could find workflows

Platforms / community repositories

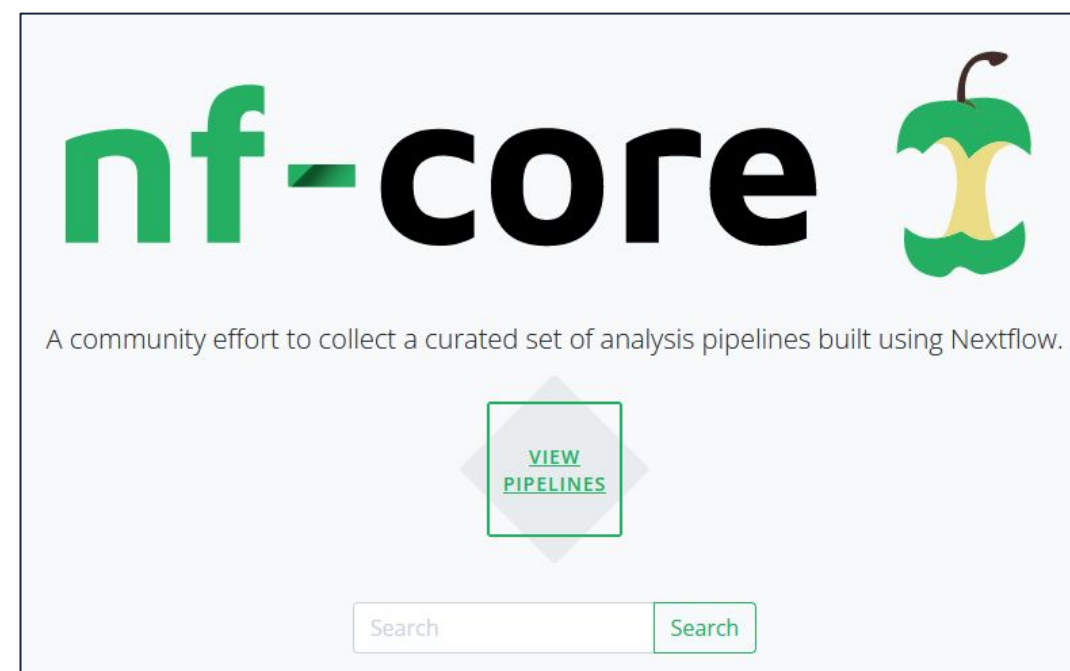


GitHub

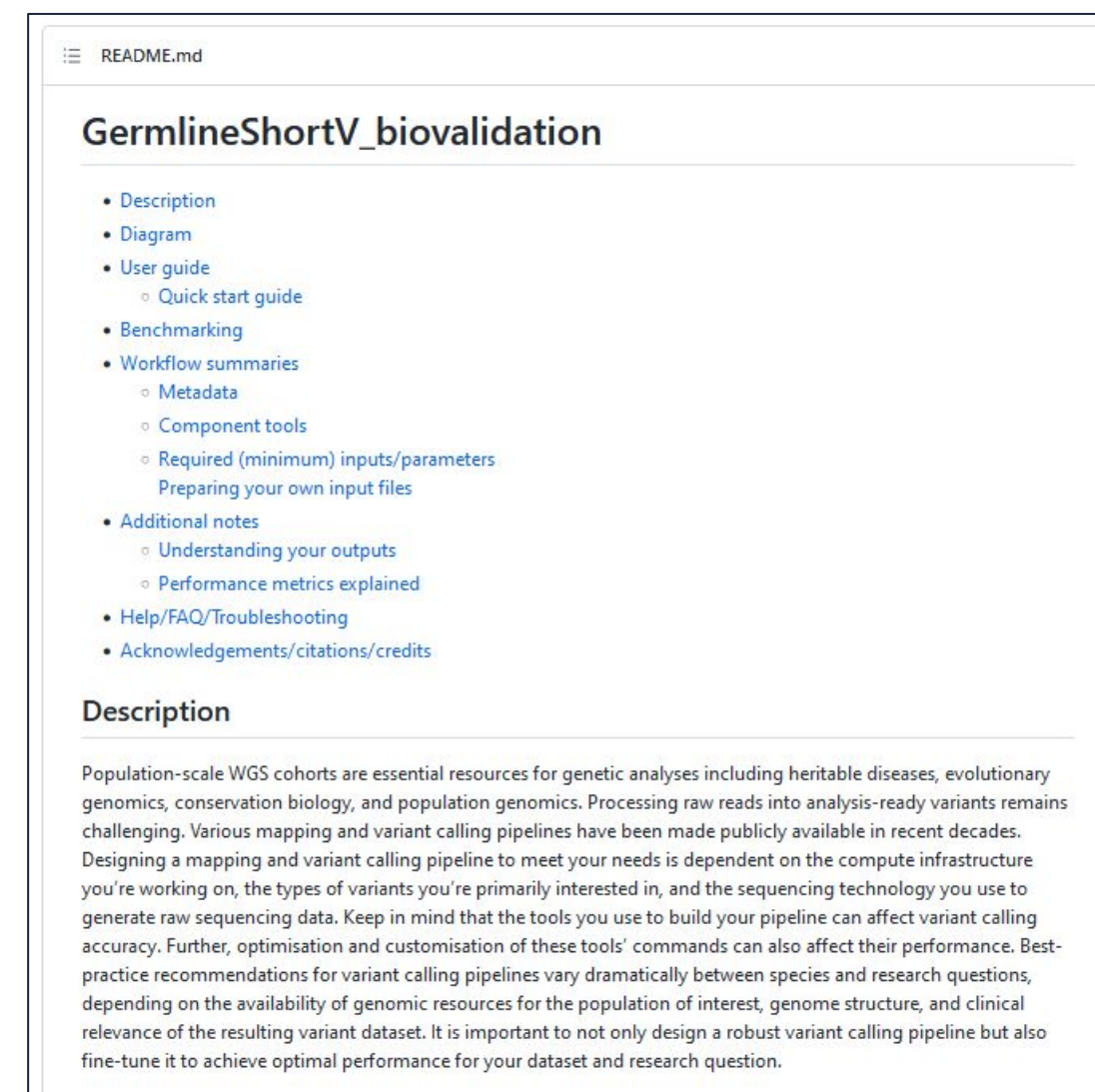


Multiple places you could find workflows

Platforms / community repositories



GitHub



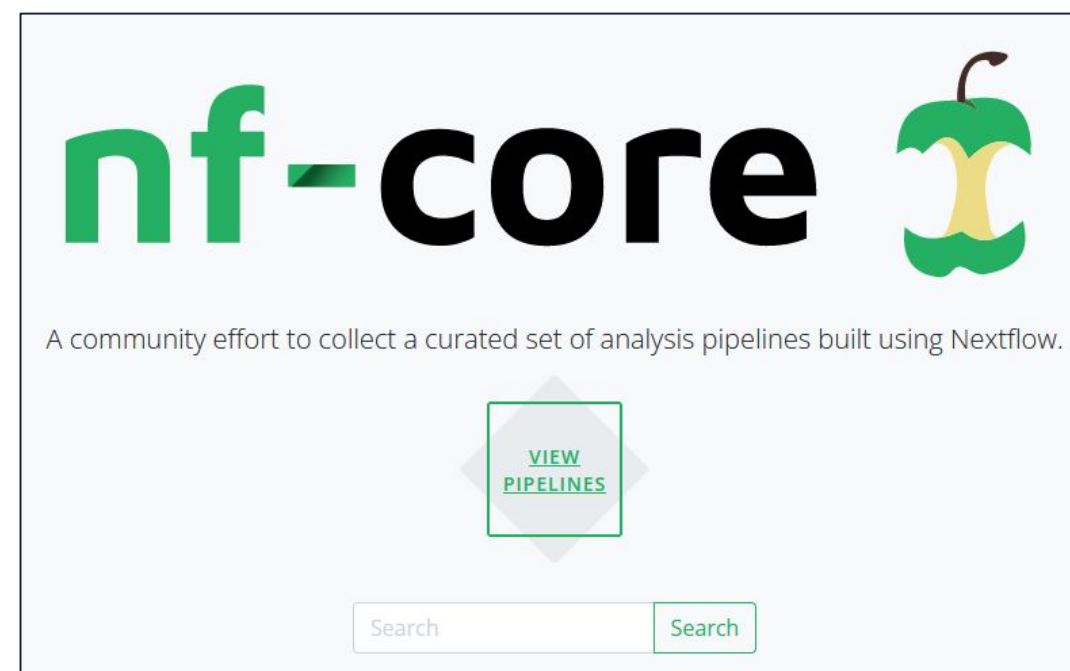
Publications

Murigneux, V., Roberts, L.W., Forde, B.M. et al. MicroPIPE: validating an end-to-end workflow for high-quality complete bacterial genome construction. *BMC Genomics* **22**, 474 (2021). <https://doi.org/10.1186/s12864-021-07767-z>

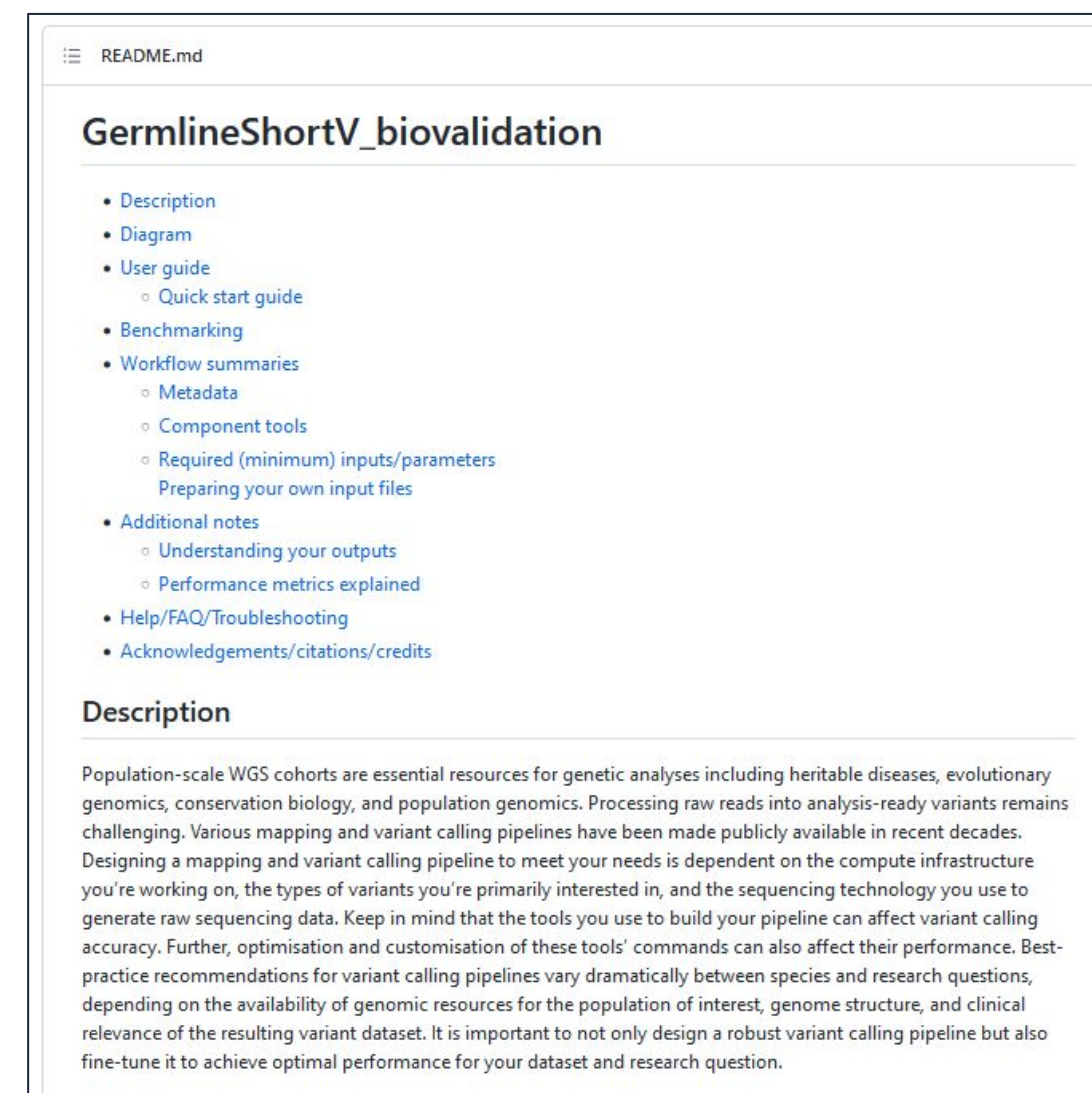
Lott, M. J., Wright, B. R., Neaves, L. E., Frankham, G. J., Dennison, S., Eldridge, M. D. B., Potter, S., Alquezar-Planas, D. E., Hogg, C. J., Belov, K., & Johnson, R. N. (2022). Future-proofing the koala: Synergising genomic and environmental data for effective species management. *Molecular Ecology*, 31, 3035– 3055. <https://doi.org/10.1111/mec.16446>

Multiple places you could find workflows

Platforms / community repositories



GitHub



Publications

Murigneux, V., Roberts, L.W., Forde, B.M. et al. MicroPIPE: validating an end-to-end workflow for high-quality complete bacterial genome construction. *BMC Genomics* **22**, 474 (2021). <https://doi.org/10.1186/s12864-021-07767-z>

Lott, M. J., Wright, B. R., Neaves, L. E., Frankham, G. J., Dennison, S., Eldridge, M. D. B., Potter, S., Alquezar-Planas, D. E., Hogg, C. J., Belov, K., & Johnson, R. N. (2022). Future-proofing the koala: Synergising genomic and environmental data for effective species management. *Molecular Ecology*, 31, 3035– 3055. <https://doi.org/10.1111/mec.16446>

Google

Consider a registry

- ✓ central
- ✓ searchable
- ✓ integrated
- ✓ standardised
- ✓ citable
- ✓ interoperable

Publications
Infrastructure efforts
Workflow organisations



Single GitHub repositories
Workflow language specific registries
Research consortia

Workflows

What is a Workflow?

21 Workflows matching the given criteria: (Clear all filters)

Search here... Go **Tag: Assembly**

Sort: Last update date (Descending)

Navigation: Previous 1 2 Next

Created At: Any time

Updated At: Any time

Tool:

- BLAST: 5
- antiSMASH: 4
- fastp: 4
- MultiQC: 4
- SAMtools: 4
- SRA Software Toolkit: 4
- More...

Workflow type:

- Galaxy: 12
- Common Workflow Language: 3
- Shell Script: 3
- Nextflow: 2

Workflow Name	Creators	Teams	Version
2: Plant virus confirmation	None	Integrated and Urban Plant Pathology Laboratory	1
LongRead Quality Control and Filtering	Bart Nijse, Jasper Koehorst, Germán Royval	UNLOCK	1
(Hybrid) Metagenomics workflow	Bart Nijse		
V-pipe (main multi-virus version)	Ivan Topol, Jędrzej Jablonski		
Purge duplicates from hifiasm assembly v1.0	Gareth P		
Shotgun-Metagenomics-Analysis	Cali Wilk, Sadsad, Tracy Che		

Search: genome assembly nextflow workflow

Tools: All Images Videos News Books More

About 105,000 results (0.42 seconds)

Scholarly articles for genome assembly nextflow workflow

- ... , Containerized **Workflow** for **Genome-to-Genome** Lift ... - Talenti - Cited by 6
- ... pipeline for complete analysis of bacterial **genomes** - Petit III - Cited by 39
- Metagenomics **workflow** for hybrid **assembly**, ... - Van Damme - Cited by 15

<https://academic.oup.com/gbe/article/evad020>

polishCLR: A Nextflow Workflow for Polishing PacBio CLR ...

by J Chang · 2023 — Throughout the polishCLR **workflow**, de novo evaluation reports are automatically generated to assess **genome assembly** quality, including k-mer ...

<https://github.com/fmalmeida/MpGAP>

fmalmeida/MpGAP: Multi-platform genome assembly pipeline ...

MpGAP is built using **Nextflow**, a **workflow** tool to run tasks across multiple compute infrastructures in a very portable manner. It uses Docker/Singularity ...

<https://nf-co.re/genomeassembler>

nf-core/genomeassembler

The pipeline is built using **Nextflow**, a **workflow** tool to run tasks across multiple compute infrastructures in a very portable manner. It uses Docker/Singularity ...

<https://workflowhub.eu/workflows>

HiFi de novo genome assembly workflow - WorkflowHub

10 May 2022 — HiFi-assembly-**workflow** is a bioinformatics pipeline that can be used to analyse Pacbio CCS reads for de novo **genome assembly** using PacBio ...

<https://workflowhub.eu/projects/workflows>

Australian BioCommons - Workflows

Workflows · PacBio HiFi **genome assembly** using hifiasm v2.1 · Purge-duplicates-from-hifiasm-assembly · BAM-to-FASTQ-QC · IndexReferenceFasta-nf · **workflow-partial** ...

central

searchable

Upload/Import Files

Import Git Repository

Upload/Import Workflow
RO-Crate

URL *

URL to the git repository. Should end in ".git".

`https://github.com/example/repository.git`

Register or **Cancel**

- ✓ central
- ✓ searchable
- ✓ **integrated**



README.md

GermlineShortV_biovalidation





- Description
- Diagram
- User guide
 - Quick start guide
- Benchmarking
- Workflow summaries
 - Metadata
 - Component tools
 - Required (minimum) inputs/parameters
 - Preparing your own input files
- Additional notes
 - Understanding your outputs
 - Performance metrics explained
- Help/FAQ/Troubleshooting
- Acknowledgements/citations/credits

Description

Population-scale WGS cohorts are essential resources for genetic analyses including heritable diseases, evolutionary genomics, conservation biology, and population genomics. Processing raw reads into analysis-ready variants remains challenging. Various mapping and variant calling pipelines have been made publicly available in recent decades. Designing a mapping and variant calling pipeline to meet your needs is dependent on the compute infrastructure you're working on, the types of variants you're primarily interested in, and the sequencing technology you use to generate raw sequencing data. Keep in mind that the tools you use to build your pipeline can affect variant calling accuracy. Further, optimisation and customisation of these tools' commands can also affect their performance. Best-practice recommendations for variant calling pipelines vary dramatically between species and research questions, depending on the availability of genomic resources for the population of interest, genome structure, and clinical relevance of the resulting variant dataset. It is important to not only design a robust variant calling pipeline but also fine-tune it to achieve optimal performance for your dataset and research question.

Metadata list

Name	Description	Mandatory
Title*	This field is mandatory and is with some workflow types pre-filled with the title of the workflow.	Yes
Description	If a CWL (abstract) file is given, the description will be parsed automatically out of the doc attribute. In any other case this field can be used to write some documentation that will be shown on the workflow page.	No
Source	If the workflow came from an external repository (i.e. GitHub), you can include its original URL here.	No
Maturity	This field can be used to specify in which maturity state the workflow is. The two available options are: <ul style="list-style-type: none">• <code>work-in-progress</code>• <code>stable</code>	No
Teams*	Every workflow registration is linked to one or more teams. If you can not select the correct team, please go to Joining a team .	Yes
Licence	The standard licence is Apache Software Licence 2.0. If you did not make the workflow yourself, be sure that the licence corresponds to the licence where you took the workflow from (for example github licences).	No
Sharing	Specify who can view the summary, get access to the content, and edit the Workflow. This is possibly already filled in according to the selected project.	No
Tags	Choose an appropriate tag for your workflow. Please check if your tag is already available and use the existing one if so. If you make a new tag, keep it simple without capitals or spaces. For example all new covid-19 workflows need to be tagged with covid-19.	No
Creators	This is an important section where all the people that were involved in making / publishing this workflow are listed. These creators will be added to the metadata in the RO-crate. 3 sections are used to specify the contributors or creators of the workflow. So far you have specified the following creators	No

-  central
-  searchable
-  integrated
-  **standardised**

Based on computational workflow profile for BioSchemas

<https://bioschemas.org/profiles/ComputationalWorkflow/1.0-RELEASE>

10.48546/WORKFLOWHUB.WORKFLOW

About 15 results (0.04 sec)

CroMaSt: A workflow for domain family curation through cross-mapping of structural instances between protein domain databases
 H Dhondge, IC de Beauchêne... - ECCB2022-21st European ..., 2022 - hal.inria.fr
 Protein domains can be viewed as building blocks, essential for understanding structure-function relationships in proteins. However, each domain database classifies protein domains ...
 ☆ Save Cite Related articles All 6 versions

[PDF] **Incrementally building FAIR Digital Objects with Specimen Data Refinery workflows**
 O Woolland, P Brack, S Soiland-Reyes... - Research Ideas and ..., 2022 - riojournal.com
 Specimen Data Refinery (SDR) is a developing platform for automating transcription of specimens from natural history collections (Hardisty et al. 2022). SDR is based on computational ...
 ☆ Save Cite Related articles All 6 versions

Making canonical workflow building blocks interoperable across workflow languages
 S Soiland-Reyes, G Bayarri, P Andrio, R...
 We introduce the concept of Canonical Workflow Building Blocks (CWB) for describing and wrapping computational workflows...
 ☆ Save Cite Cited by 1 Related

[PDF] **Creating lightweight FAIR Digital Objects for workflow provenance**
 S Soiland-Reyes, P Sefton, LJ Castro...
 RO-Crate (Soiland-Reyes et al. 2022) is a standard for packaging software along with their metadata, based on LinkML...
 ☆ Save Cite Related articles All

[PDF] **Species-specific exudates influence phytoplankton community composition**
 E Briand, M Lassudrie, C Noël, C Jauzein...
 Phytoplankton-bacteria interactions are crucial for the cycling of organic matter, exchange of metabolites...
 ☆ Save Cite

Citation Copy

Price, G. & Farquharson, K. PacBio HiFi genome assembly using hifiasm v2.1. (2022)
 doi:10.48546/WORKFLOWHUB.WORKFLOW.221.3

- ✓ central
- ✓ searchable
- ✓ integrated
- ✓ standardised
- ✓ **citable**

Price, G. & Farquharson, K. PacBio HiFi genome assembly using hifiasm v2.1. (2022)
 doi: <https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3>

Assembly workflow

⚠ Important: This workflow includes an adapter filtering step that makes use of `HiFiAdapterFilt`. Adapter contamination in the reads could cause misassemblies, and will result in NCBI rejecting the upload of affected genome assemblies. `HiFiAdapterFilt` identifies `.ccs` reads containing adapter sequences using the same method as NCBI and removes the entire read prior to genome assembly to avoid such misassemblies.

1. Make sure you are logged into Galaxy Australia

2. Visit this link to:

- retrieve v2.1 of the assembly workflow,
- add it to your Galaxy Australia workflows list, and
- open your workflow (which can also be reached by clicking the `Workflow` tab [highlighted by a red box in Fig 6] in the Galaxy interface)

Tip: Did you know? version 2.1 of

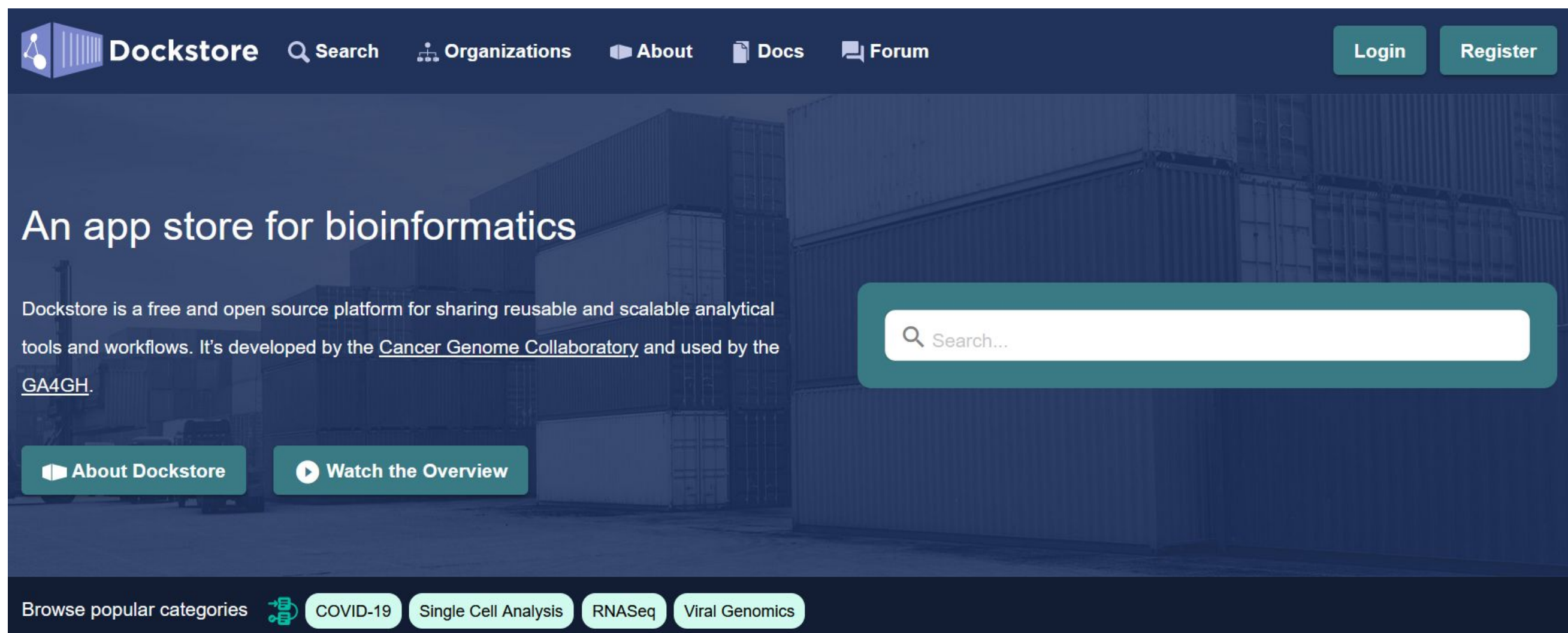
3. Select the play button (highlighted in Fig 12) (the workflow is shown in Fig 12).

Name	Tags	Updated	Sharing	Bookmarked
▼ PacBio HiFi genome assembly using hifiasm v2.1 ✓	fastq × hifiasm × HiFi × genome_assembly ×	in less than a minute		☆

- ✓ central
- ✓ searchable
- ✓ integrated
- ✓ standardised
- ✓ citable
- ✓ **interoperable**

Two key options

Dockstore



<https://dockstore.org/>

O'Connor BD, Yuen D, Chung V *et al.* The Dockstore: enabling modular, community-focused sharing of Docker-based genomics tools and workflows [version 1; peer review: 2 approved]. *F1000Research* 2017, **6**:52 (<https://doi.org/10.12688/f1000research.10137.1>)

Two key options

WorkflowHub

The screenshot shows the WorkflowHub homepage. At the top, there is a navigation bar with the WorkflowHub logo, a search bar, and links for 'About', 'Help', 'Register', and 'Log in'. Below the navigation bar is a large dark green banner with the WorkflowHub logo and the text: 'A registry for describing, sharing and publishing scientific computational workflows'. It also states: 'WorkflowHub aims to facilitate discovery and re-use of workflows in an accessible and interoperable way. This is achieved through extensive use of open standards and tools, including CWL, RO-Crate, Bioschemas and GA4GH's TRS API, in accordance with the FAIR principles. WorkflowHub supports workflows of any type in its native repository.' There are 'Learn more' and 'Register' buttons. Below the banner is a light green notification bar: 'You have been logged out.' The main content area is divided into several sections: 'Welcome to WorkflowHub' with links to help, issues, and feedback; 'Latest additions' listing recent publications and workflows; 'Find content' with buttons for 'Browse Workflows', 'Browse Collections', 'Browse Spaces', and 'Browse Teams'; 'Shortcuts' for SARS-CoV-2/COVID-19 and WfCommons; 'Current Workflow Types' listing various tools like Galaxy, KNIME, Nextflow, and Snakemake; and 'Tags' for various bioinformatics topics.

<https://workflowhub.eu/>

Carole Goble, Stian Soiland-Reyes, Finn Bacall, Stuart Owen, Alan Williams, Ignacio Eguinoa, Bert Driesbeke, Simone Leo, Luca Pireddu, Laura Rodríguez-Navas, José M^a Fernández, Salvador Capella-Gutierrez, Hervé Ménager, Björn Grüning, Beatriz Serrano-Solano, Philip Ewels, & Frederik Coppens. (2021). Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory. Zenodo. <https://doi.org/10.5281/zenodo.4605654>

Two key options

Dockstore

Register tools and workflows

Standard metadata

Git integration

DOI minting capability

Organisations

WDL, CWL, Nextflow, Galaxy

WorkflowHub

Register workflows

Standard metadata

Git integration

DOI minting capability

Spaces & teams

All workflow types



WorkflowHub

step-by-step

3



A registry for describing, sharing and publishing **scientific computational workflows**

WorkflowHub aims to **facilitate discovery and re-use** of workflows in an accessible and interoperable way. This is achieved through extensive use of **open standards** and tools, including [CWL](#), [RO-Crate](#), [Bioschemas](#) and [GA4GH's TRS API](#), in accordance with the **FAIR principles**.

WorkflowHub **supports workflows of any type** in its native repository.

[Learn more](#)[Register](#)

Welcome to WorkflowHub

- **Help** is available on about.workflowhub.eu.
- Report any **issues or suggest new features** on [GitHub](#).
- For **comments, questions or feedback**, please use the [feedback form](#).

Want to join the WorkflowHub community?
See our current activities and upcoming meetings [here](#).

Latest additions

 [Dataset: Computer software for identification of honey bee subspecies and evolutionary lineages](#)

Publication - added 20 days ago

 [Collection of wing images for conservation of honey bees \(*Apis mellifera*\) biodiversity in Europe](#)

Publication - added 20 days ago

 [Master of Pores 2](#)
Workflow - added 28 days ago

Find content

Browse Workflows



Browse Collections



Browse Spaces

Browse Teams

Shortcuts



Discover workflows relating to **SARS-CoV-2 / COVID-19**

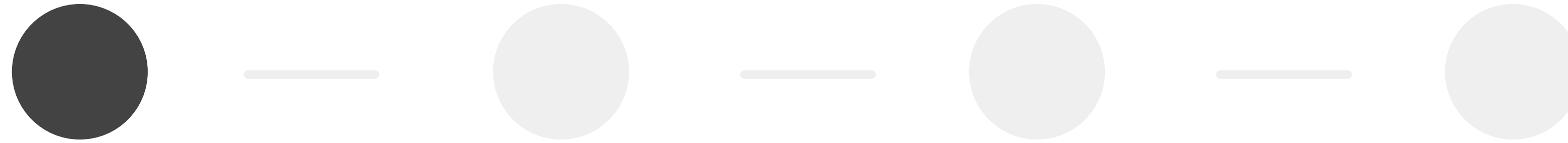
 **WfCommons**

Looking for WfCommons? [Click here](#)

Current Workflow Types

- Common Workflow Language
- Galaxy
- KNIME
- Nextflow
- Snakemake

Register



Register an account for WorkflowHub

Username

Username

The username should contain a minimum of 3 characters.

Email address

Email address

Password

Password

Password should contain a minimum of 10 characters.

Confirm Password

Password

I have read and agree to the [Terms and Conditions](#) and the [Data Policy](#)

Register

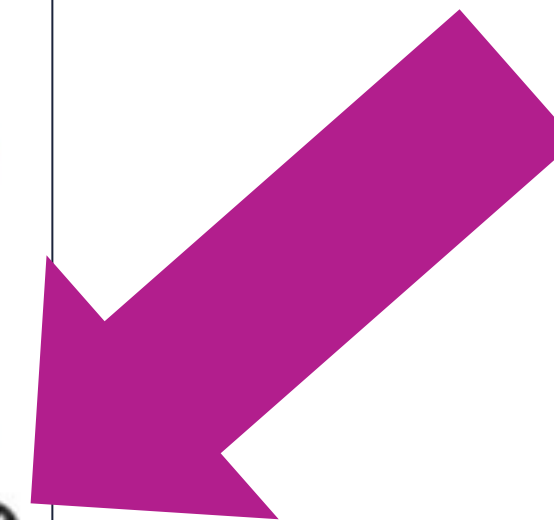
[Already registered? - goto Login](#)

Alternatively...

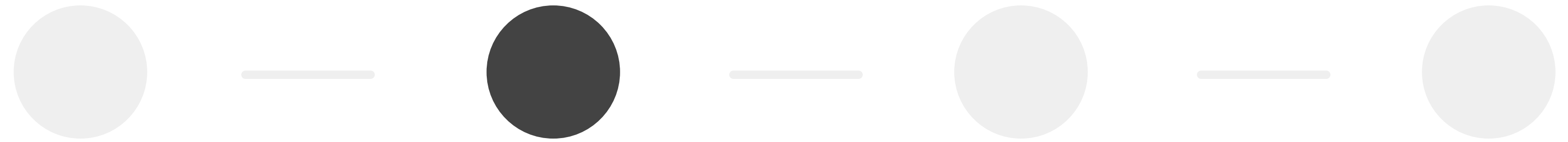
- [Log in using LS Login](#)




- [Log in using GitHub](#) 



Choose a space



Space

 **Australian BioCommons**

[Overview](#) [Related items](#)

The Australian BioCommons enhances digital life science research through world class collaborative distributed infrastructure. It aims to ensure that Australian life science research remains globally competitive, through sustained strategic leadership, research community engagement, digital service provision, training and support.

Web page: <https://www.biocommons.org.au/>

Funding details:
Core funding for the Australian BioCommons comes from the National Collaborative Research Infrastructure Strategy (NCRIS) via Bioplatforms Australia, which is subcontracted to The University of Melbourne as the lead agent. This core funding is amplified through coinvestment from BioCommons partners <https://www.biocommons.org.au/funding>

 **Independent Teams**

[Overview](#) [Related items](#)

A space managed by WorkflowHub administrators for teams that don't want/need to manage their own space.

Web page: *Not specified*

Funding details:
No funding details specified




Workflow Hub

Join a team

or create
one!

Space

Team

 Galaxy Australia

[Overview](#) [Related items](#)

Galaxy is an **open, web-based** platform for accessible, reproducible, and transparent computational biological research.

- **Accessible:** Users can easily run tools without writing code or using the CLI; all via a user-friendly web interface.
- **Reproducible:** Galaxy captures all the metadata from an analysis, making it completely reproducible.
- **Transparent:** Users share and publish analyses via interactive pages that can enhance analyses with user annotations.
- **Scalable:** Galaxy can run on anything, from a laptop, to large clusters, to the cloud

Space: [Australian BioCommons](#) **WorkflowHub PALS:** *No PALS for this Team*

SEEK ID: <https://workflowhub.eu/projects/54> **Team created:** 9th Aug 2021

Public web page: <https://usegalaxy.org.au/>

Organisms: *No Organisms specified*

Join a team

or create one!

Space

Team

Space

Select the Space you wish to associate the new Team with, out of the following list that you administer.

Australian BioCommons

Alternatively you can choose to create a new Space, which your new Team will be associated with.

Create a new Space?

Team

Please provide some basic details about the Team that will be created. You will be able to update and add additional information once it has been created.

Title *

Description

Web page

Organization

This is the Organization you are associated with, and will be used when defining your membership with the Team.

Type the name of the Organization, or part of the name or the city it may be associated with. From the list of the results found, select the Organization you are looking for. If you cannot find it in the list, continue to type the full name and select the top *new item*; you will then be able to provide the additional details about the new Organization.

Type the name of the Organization *

Website

City

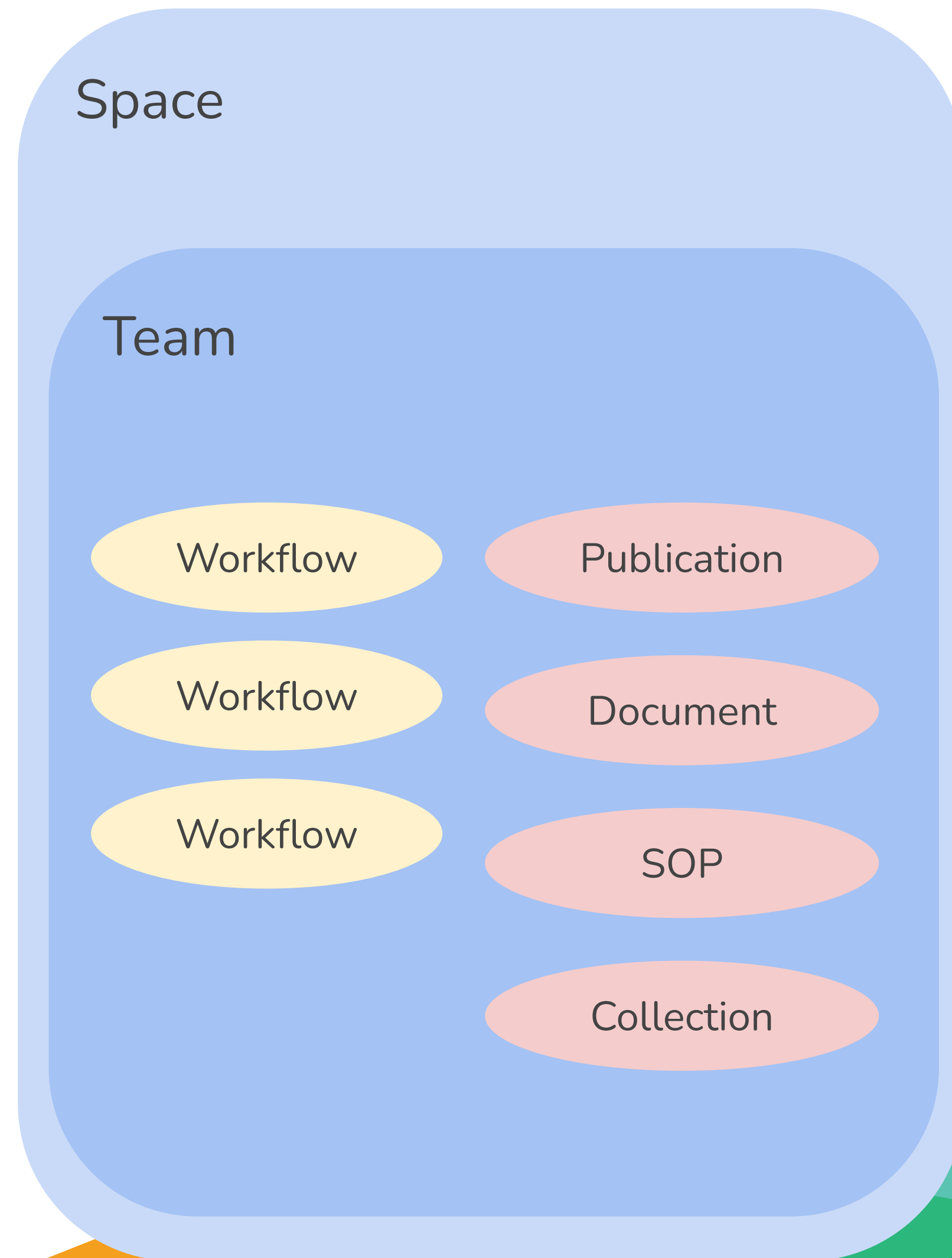
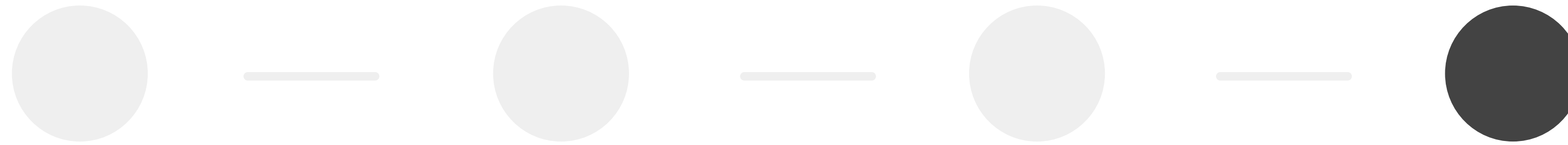
Country

Select a country

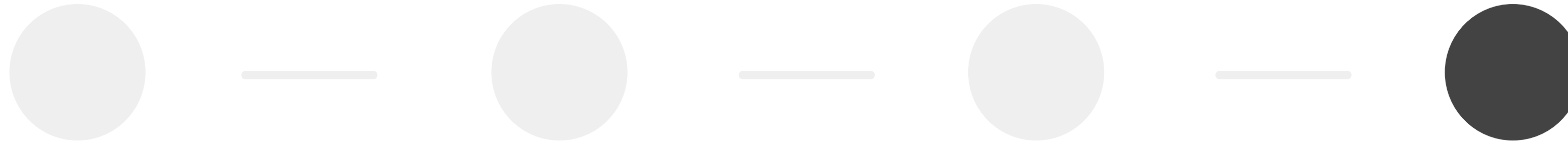
Submit

One wizard accommodates these processes

Add items



Add items

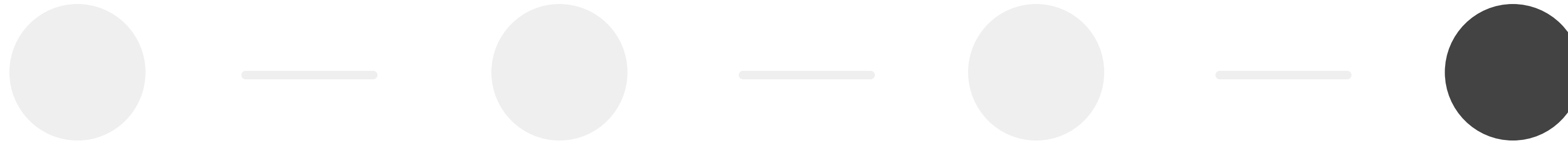


Space

Team

The screenshot shows the WorkflowHub interface for the 'QCIF Bioinformatics' space. At the top, there is a navigation bar with the WorkflowHub logo, a 'Browse' dropdown, a search bar with the placeholder 'Search here...', and a 'Search' button with a gear icon. On the right side of the navigation bar are links for 'About', 'Help', 'Register', and 'Log in'. Below the navigation bar, the main header displays 'QCIF Bioinformatics' with a small icon to the left and an 'Overview' link to the right. Underneath the header, there are two tabs: 'Overview' and 'Related items', with 'Related items' being the active tab. The main content area is titled 'Related items' and features a horizontal list of filters: 'People (3)', 'Spaces (1)', 'Organizations (1)', 'Publications (1)', and 'Workflows (3)'. The 'People (3)' filter is currently selected, highlighted with a dark teal background.

Add items



Space

Team

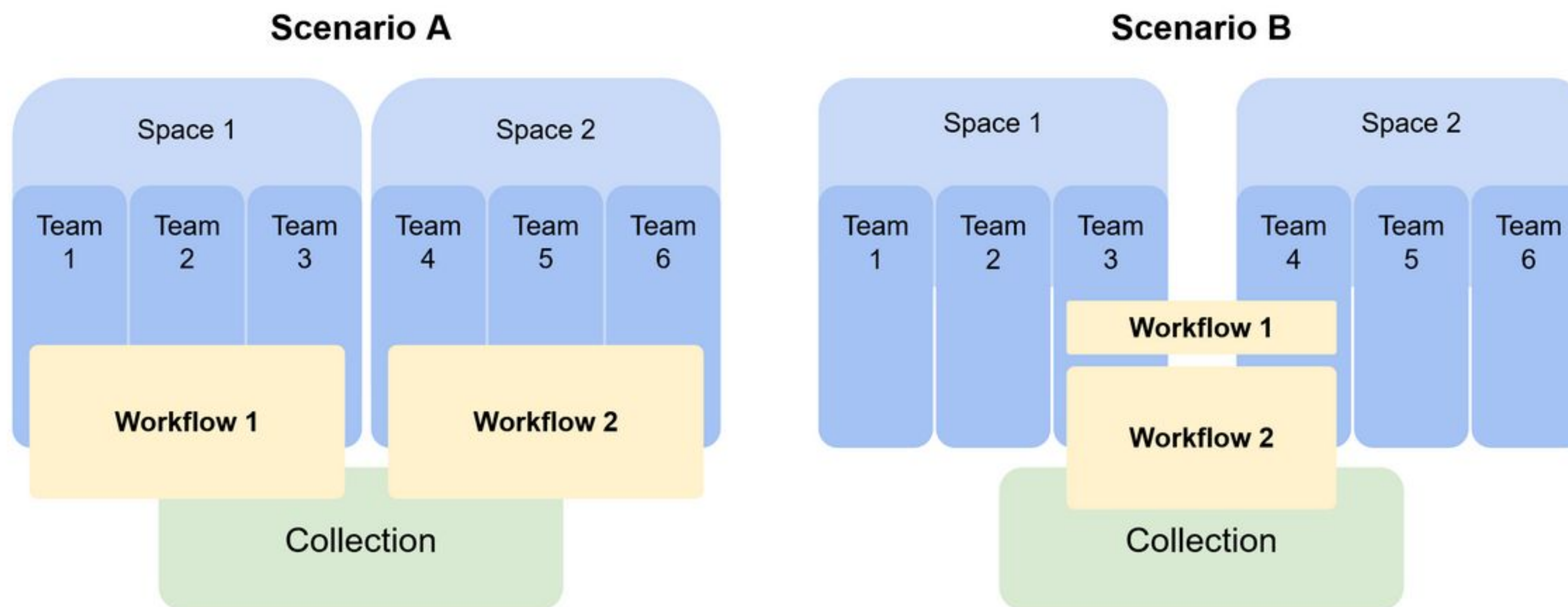
Workflow

Workflow

Workflow

The screenshot shows the WorkflowHub interface. At the top, there is a navigation bar with the WorkflowHub logo, a search bar, and links for 'About', 'Help', 'Register', and 'Log in'. Below the navigation bar, the page title is 'QCIF Bioinformatics' with an 'Overview' link. The main content area is titled 'Related items' and has tabs for 'People (3)', 'Spaces (1)', 'Organizations (1)', 'Publications (1)', and 'Workflows (3)'. The 'Workflows (3)' tab is selected, showing a list of items under the 'Australian BioCommons' organization. The items are: 'Australian BioCommons', 'QCIF Bioinformatics', 'MicroPIPE: validating an end-to-end workflow for high-quality complete bacterial genome construction.', 'microPIPE: a pipeline for high-quality bacterial genome construction using ONT and Illumina sequencing' (marked with a green X), '16S_biodiversity_for_overlap_paired_end', and '16S_biodiversity_for_nonoverlap_paired_end' (both marked with a Galaxy icon). Below this list, there are links to 'Pawsey Supercomputing Research Centre', 'Sydney Informatics Hub', and 'Janis'.

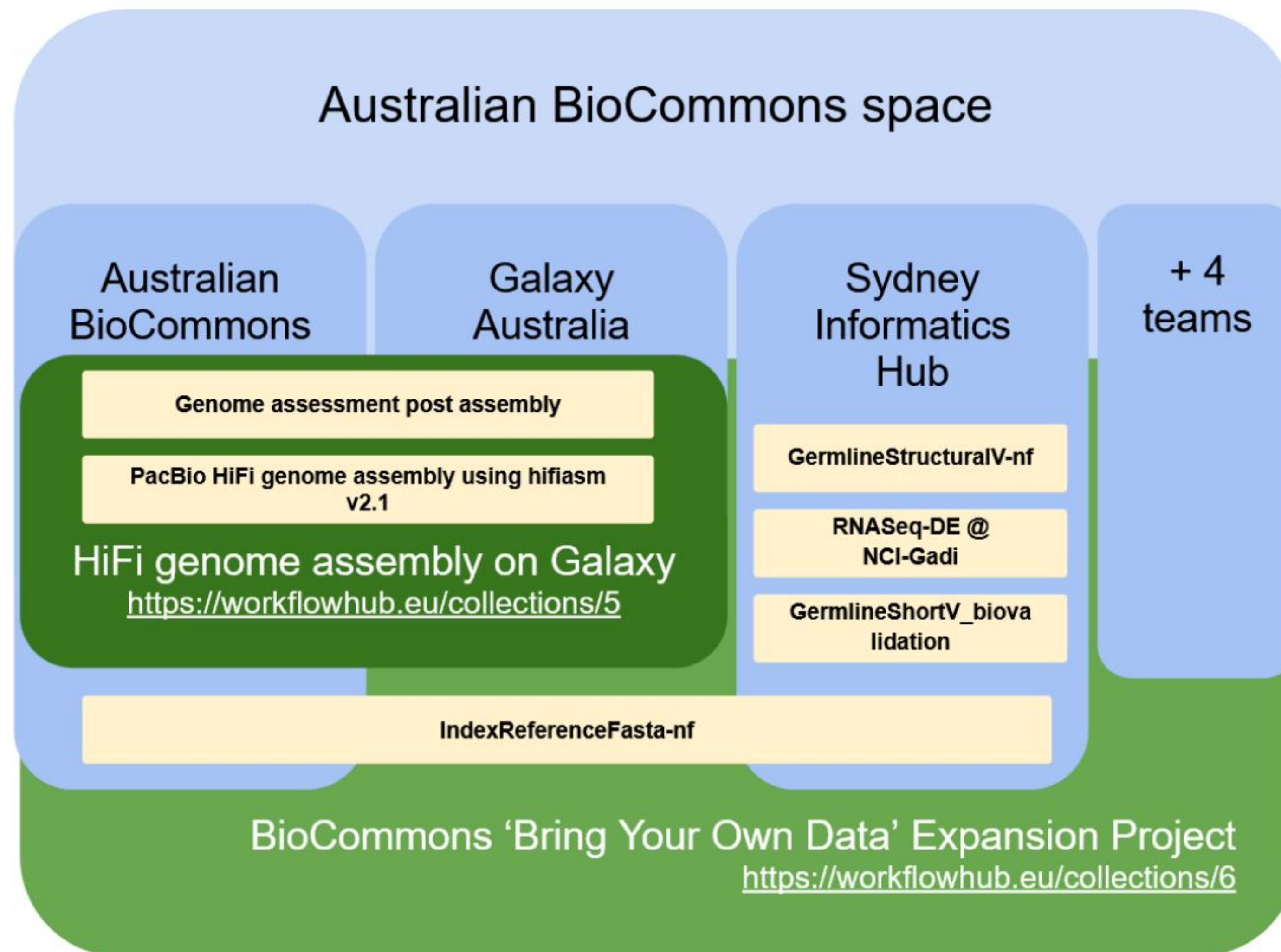
Note: these structures are flexible



- 3 *Teams* from space one, and 3 *Teams* from space two have created a *workflow* each
- These are both part of the same *collection*
- *Spaces* are not sharing a *workflow*, but do contribute to the same *Collection*

- Team #3 from space one, and team #4 from space two have collaborated to create 2x *workflows*
- *Workflow 2* is contributed to a collection
- *Spaces* are sharing *workflows* via their component *Teams*, but do not directly share the *workflows*

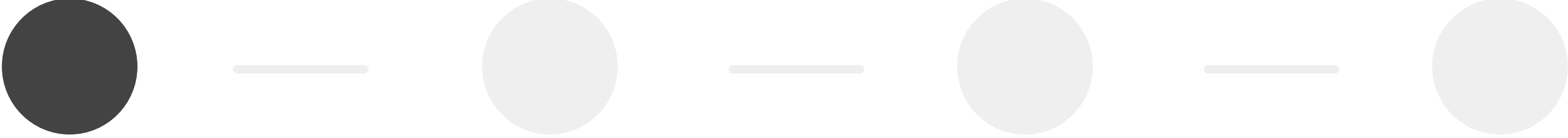
Australian BioCommons example



Let's register a workflow!

Select your file

import method



Upload/Import Files

Import Git Repository

Upload/Import Workflow RO-Crate

Workflow *

The main executable workflow.

Local file Remote URL

Browse... No file selected.

Workflow Type * [+ New workflow type](#)

Other

The type of the above workflow.

URL *

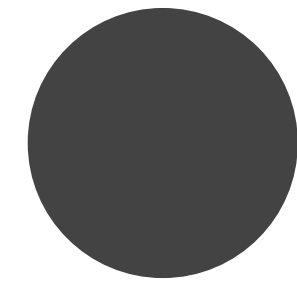
URL to the git repository. Should end in ".git".

`https://github.com/example/repository.git`

Register or Cancel

*Note: Two purple arrows point to the 'Workflow *' and 'Workflow Type *' sections.*

Select your file



import method



Upload/Import Files

Import Git Repository

Upload/Import Workflow RO-Crate

URL *
URL to the git repository. Should end in ".git".

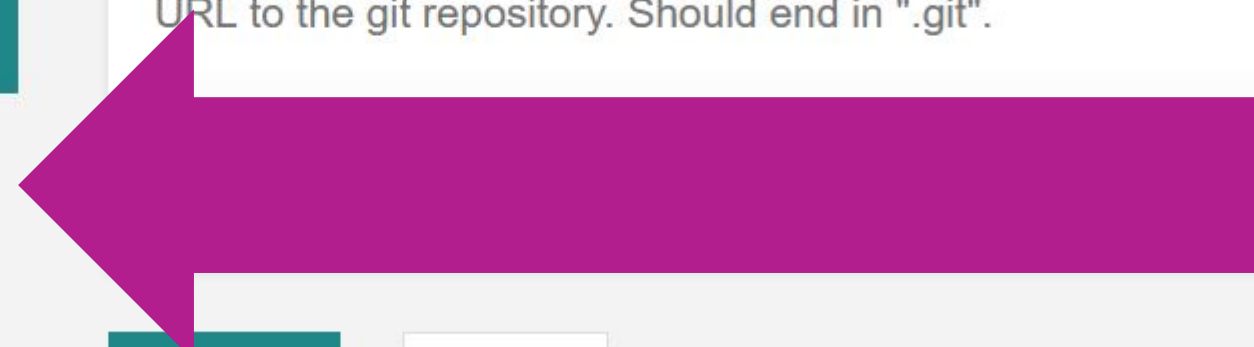
<https://www.researchobject.org/ro-crate/>

<https://w3id.org/workflowhub/workflow-ro-crate/>

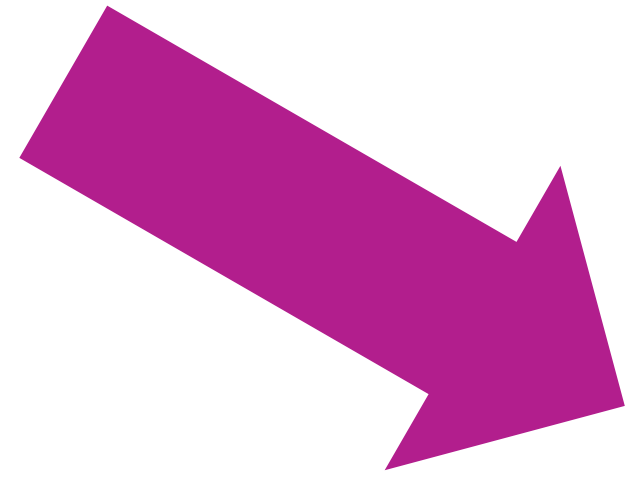
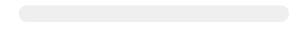
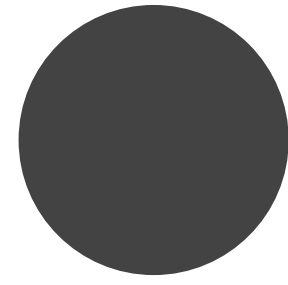
Register or Cancel



Workflow
RO-Crate



Import



Select Target

Branches Tags

v1.0.0

v2.0.0

v2.1.0

Register or Cancel

Import

Select Target

Branches **Tags**

v1.0.0
 v2.0.0
 v2.1.0

Register or **Cancel**

New Workflow

Files

Workflow *
The main executable workflow.

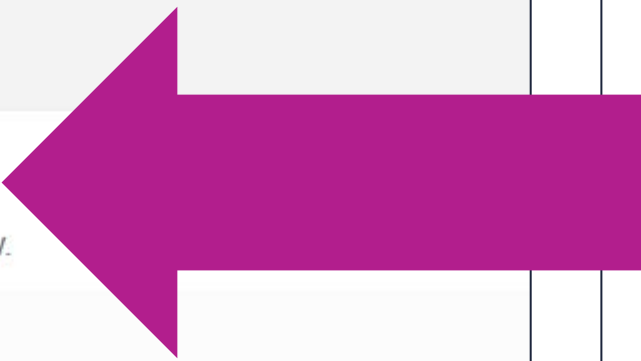
Workflow Type * [New workflow type](#)

Other
The type of the above workflow.

Abstract CWL
(Optional) The abstract CWL that describes the main workflow.

Diagram
(Optional) A diagram that illustrates the main workflow.

Continue



Select path

test_data

- CITATION.cff
- Galaxy-Workflow-PacBio_HiFi_genome_assembly_using_hifiasm_v2.1.ga
- LICENSE.md
- README.md
- change_log.md
- infrastructure_optimisation.md
- workflows.md

Select **Cancel**

Add metadata

New Workflow

Workflow Type* **Workflow type**

Galaxy

Title*

PacBio HiFi genome assembly using hifiasm v2.1

Description

Assembly, visualisation and analysis of PacBio HiFi genome assembly using hifiasm v2.1

Update description if needed

Source

<https://github.com/AustralianBioCommons/PacBio-HiFi-genome-assembly-using-hifiasm>

If this workflow came from an external repository (i.e. GitHub).

Topic annotations

Operation annotations

Add EDAM ontology terms

Maturity

Not specified

This field is used to indicate to users what level of stability they can expect from the workflow.

Teams*

The following teams are associated with this Workflow:

Australian BioCommons

Select a team...

Teams

License

Apache Software License 2.0

For more information on this license, please visit <https://opensource.org/licenses/Apache-2.0>

Add license (auto if license in GitHub)

Discussion Channels

Sharing

Here you can specify who can view the summary of, get access to the content of, and edit the Workflow.

	No Access	View	Download	Edit	Manage
Public	<input type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Australian BioCommons Dev	<input type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Share with a Person | Share with a Team / Organization | Share with a Space

Tags

fastq x hifiasm x HiFi x genome_assembly x

Custom tags

Creators

Type to search for creators.. or Add new creator

- Gareth Price No affiliation specified
- Katherine Farquharson No affiliation specified

Edit creators

Additional credit

A free-text field to specify additional credit for the creation of this Workflow.

Add metadata

<https://edamontology.github.io/edam-browser/>

New Workflow

Workflow Type + New workflow type

Galaxy

Title *

PacBio HiFi genome assembly using hifiasm v2.1

Description

Assembly, visualisation and quality control workflow for high fidelity reads built from circular consensus sequence (PacBio HiFi) data.

Source

<https://github.com/AustralianBioCommons/PacBio-HiFi-genome-assembly-using-hifiasm>

If this workflow came from an external repository (i.e. GitHub).

Topic annotations

Operation annotations

Maturity

Not specified

This field is used to indicate to users what level of stability they can expect from the workflow.

Teams *

The following teams are associated with this Workflow:

Australian BioCommons Dev

Select a team...

EDAM EDAM ontology

EDAM is a simple ontology of well established, familiar concepts that are prevalent within bioinformatics.

1.25 (stable) - All EDAM -

Search for concepts, enter at least two letters

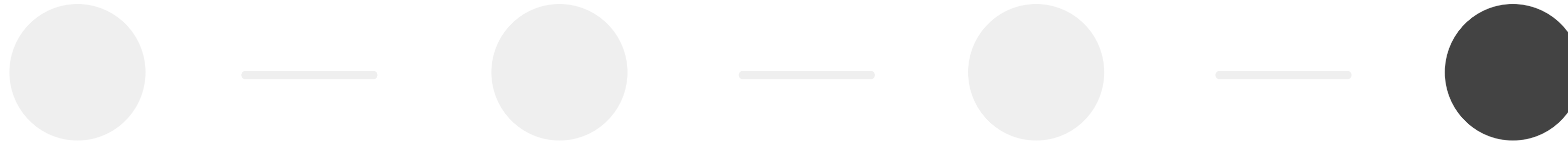
Details of term "Bioinformatics"

Term	Bioinformatics
Definition	The archival, curation, processing and analysis of complex biological data.
Comment	This includes data processing in general, including basic handling of files and databases, datatypes, workflows and annotation.
Exact synonyms	
Narrow synonyms	
URI	topic_0091
Parents	Informatics
Community usage	
bio.tools	571 times
Biosphere	22 times by appliances, 15 times by tools.
TeSS	10 times
Links	Open in AberOWL , BioPortal , OLS or WebProtégé .

Add EDAM ontology terms

Ison, J., Kalaš, M., Jonassen, I., Bolser, D., Uludag, M., McWilliam, H., Malone, J., Lopez, R., Pettifer, S. and Rice, P. (2013). EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics*, 29(10): 1325-1332. 10.1093/bioinformatics/btt113 PMID: 23479348 Open access

Review



Workflow was successfully uploaded and saved.

PacBio HiFi genome assembly using hifiasm

v2.1 Version 1

[View on GitHub](#) [Subscribe](#) [Download RO Crate](#) [Add new](#) [Actions](#)

[Overview](#) [Files](#) [Related items](#)

Workflow Type: Galaxy

Assembly, visualisation and quality control workflow for high fidelity reads built from circular consensus sequence (PacBio HiFi) data.

SEEK ID: <https://dev.workflowhub.eu/workflows/584?version=1>

Inputs

ID	Name	Description	Type
FASTQ input	FASTQ input	The FASTQ input to hifiasm can be sourced from the "BAM to FASTQ + QC" sub workflow.	File

Steps

ID	Name	Description
1	HiFi Adapter Filter	toolshed.g2.bx.psu.edu/repos/galaxy-australia/hifiadapterfilt/hifiadapterfilt/2.0.0+galaxy0
2	hifiasm assembly	toolshed.g2.bx.psu.edu/repos/bgruening/hifiasm/hifiasm/0.16.1+galaxy3
3	Haplotype resolved Bandage info	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_info/0.8.1+galaxy1
4	Haplotype resolved Bandage image	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_image/0.8.1+galaxy3
5	Processed unitig Bandage info	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_info/0.8.1+galaxy1
6	Processed unitig Bandage image	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_image/0.8.1+galaxy3
7	Alternate assembly Bandage info	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_info/0.8.1+galaxy1
8	Primary assembly Bandage info	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_info/0.8.1+galaxy1
9	Convert primary assembly GFA to FASTA	toolshed.g2.bx.psu.edu/repos/iuc/gfa_to_fa/gfa_to_fa/0.1.2
10	Alternate assembly Bandage image	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_image/0.8.1+galaxy3
11	Primary assembly Bandage image	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_image/0.8.1+galaxy3
12	Primary assembly Fasta Statistics	toolshed.g2.bx.psu.edu/repos/iuc/fasta_stats/fasta_stats/1.0.3

Outputs

ID	Name	Description	Type
HiFi Adapter Filter on input dataset(s): clean reads	HiFi Adapter Filter on input dataset(s): clean reads	n/a	File

Creators and Submitter

Creators
Gareth Price, Katherine Farquharson

Submitter
Johan Gustafsson

Citation

Make your Workflow easily citable by generating a DOI for it.

Your Workflow is published and eligible for a DOI.

[Generate a DOI](#)

License

Apache Software License 2.0

Activity

Views: 0
Created: 17th Mar 2023 at 05:45

Tags

fastq genome_assembly HiFi hifiasm

[Update your tags](#)

Attributions

None

Look at a best practice example

Structured documentation imported from GitHub



Complete author list (in correct order)



DOI



License



Annotations



Diagram





Demo

4

Hints for maintaining your workflow!

Include a CFF
file in your
GitHub
repository

13 lines (13 sloc) | 417 Bytes

```
1  cff-version: 2.1.0
2  message: "If you use this workflow, please cite it as below."
3  authors:
4    - family-names: "Price"
5      given-names: "Gareth"
6      orcid: "https://orcid.org/0000-0003-2439-8650"
7    - family-names: "Farquharson"
8      given-names: "Katherine"
9      orcid: "https://orcid.org/0000-0002-9009-7453"
10 title: "PacBio-HiFi-genome-assembly-using-hifiasm"
11 version: 2.1.0
12 doi: [DOI goes here]
13 date-released: 2022-10-21
```

<https://citation-file-format.github.io/>

Druskat, Stephan, Spaaks, Jurriaan H., Chue Hong, Neil, Haines, Robert, Baker, James, Bliven, Spencer, Willighagen, Egon, Pérez-Suárez, David, & Konovalov, Alexander. (2021). Citation File Format (1.2.0). Zenodo. <https://doi.org/10.5281/zenodo.5171937>

Hints for maintaining your workflow!

Include a CFF file in your GitHub repository

```
13 lines (13 sloc) | 417 Bytes
1  cff-version: 2.1.0
2  message: "If you use this workflow, please cite it as below."
3  authors:
4    - family-names: "Price"
5      given-names: "Gareth"
6      orcid: "https://orcid.org/0000-0003-2439-8650"
7    - family-names: "Farquharson"
8      given-names: "Katherine"
9      orcid: "https://orcid.org/0000-0002-9009-7453"
10 title: "PacBio-HiFi-genome-assembly-using-hifiasm"
11 version: 2.1.0
12 doi: [DOI goes here]
13 date-released: 2022-10-21
```

GitHub citation option

The screenshot shows a GitHub repository page for a workflow. The 'About' section describes it as an assembly, visualization, and quality control workflow for high fidelity reads built from circular consensus sequence (PacBio HiFi) data. The commit history shows 8 commits, with the most recent one from October 21, 2022. A modal window titled 'Cite this repository' is open, providing citation metadata and options for APA and BibTeX. The citation text is: 'Price, G., & Farquharson, K. (2022). PacBio-HiF'. A 'View citation file' button is also visible.

+ import of creators to WorkflowHub

The screenshot shows the 'Creators' management interface in WorkflowHub. It features a search bar with the placeholder text 'Type to search for creators..' and an 'Add new creator' button. Below the search bar, two creators are listed: Gareth Price (No affiliation shown) and Katherine Farquharson (with an ORCID icon and No affiliation shown). There is also an 'Additional credit' section with a text field for specifying additional credit for the creation of this Workflow.

Hints for maintaining your workflow!

Create a DOI
when metadata
is completed

PacBio HiFi genome assembly using hifiasm v2.1


General usage recommendations

Please see the [Genome assembly with hifiasm on Galaxy Australia](#) guide.

See [change log](#)

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/221?version=3>
DOI: [10.48546/workflowhub.workflow.221.3](https://doi.org/10.48546/workflowhub.workflow.221.3) 


Inputs

ID	Name	Description	Type
FASTQ input	FASTQ input	The FASTQ input to hifiasm can be sourced from the "BAM to FASTQ + QC" sub workflow.	File


Creators


 Gareth Price,  Katherine Farquharson

Submitter


 Johan Gustafsson

Tools


HiFiAdapterFilt 

Bandage 

Citation

 Copy

Price, G., & Farquharson, K. (2022). *PacBio HiFi genome assembly using hifiasm v2.1*. WorkflowHub. <https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3>

American Psychological Association 


Hints for maintaining your workflow!

Create a DOI
when metadata
is completed

Price, G., & Farquharson, K. (2022). *PacBio HiFi genome assembly using hifiasm v2.1*.
WorkflowHub. <https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3>

Hints for maintaining your workflow!

Use GitHub app
to automate
registry
updates!



GitHub App
LifeMonitor

The **LifeMonitor GitHub App** is a companion to the LifeMonitor service for supporting the sustainability and reusability of published computational scientific data analysis workflows.

The LifeMonitor GitHub App does the following things:

- Examines the repositories on which it is installed and applies a series of checks.
- Suggests pull requests to make changes or additions to bring the workflow repository closer to conforming to best practices.
- Opens issues to let you know about problems detected by the checks and to request new metadata.
- Registers new releases/versions of the workflow with both the LifeMonitor service and the WorkflowHub workflow registry.

Together, the LifeMonitor service and GitHub App help you keep your workflow functional over time and help you follow the community-accepted conventions for your specific workflow type, so that it may be more easily adopted and re-used by others in your scientific community.

Additional information is available in [the documentation](#).

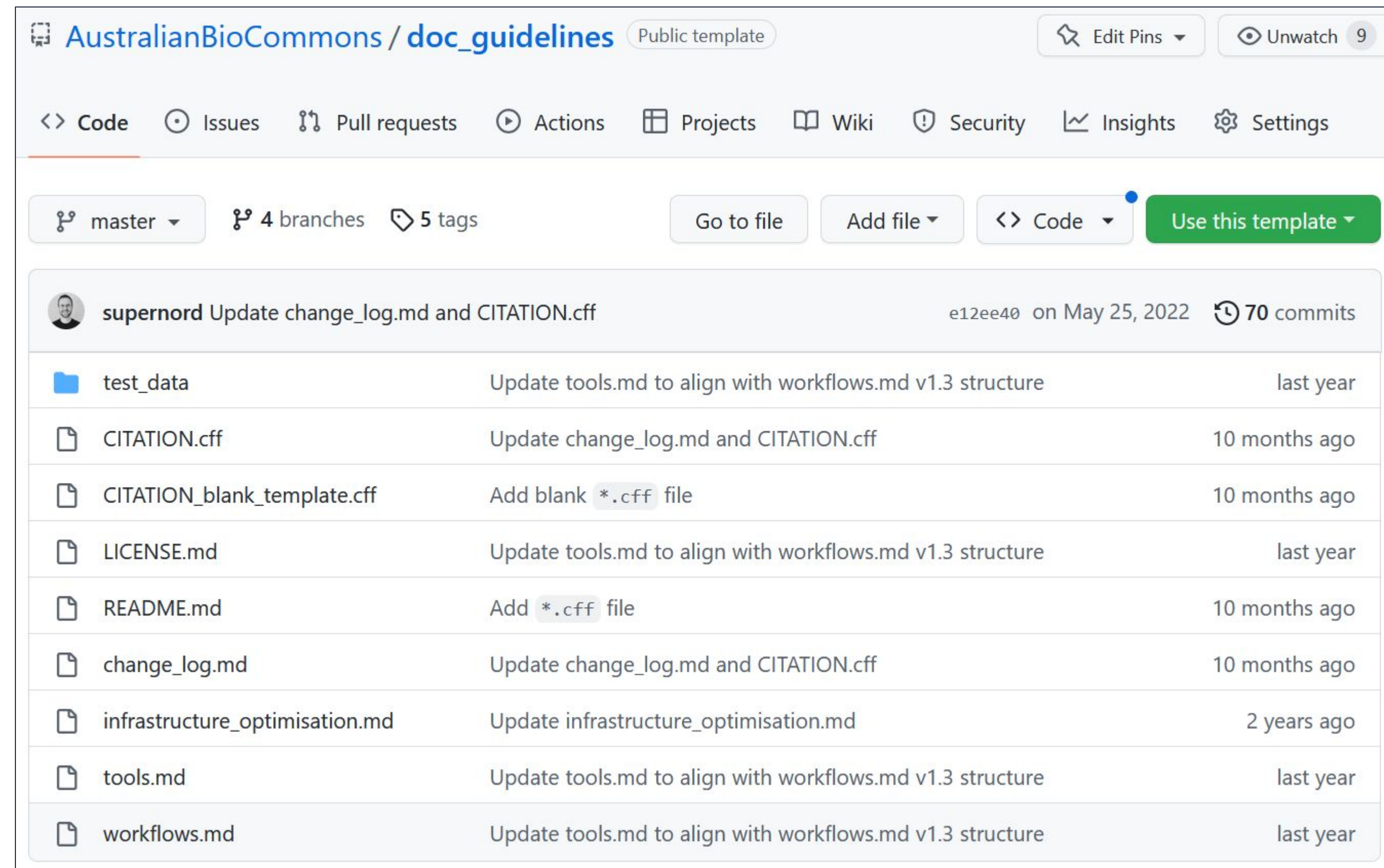
Useful links

- [Docs](#)
- [API Explorer](#)
- [Web Application](#)
- [Workflow Hub](#)

https://crs4.github.io/life_monitor/lm_wft_best_practices_github_app#installation

Hints for maintaining your workflow!

Use
documentation
templates



AustralianBioCommons / doc_guidelines Public template

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 4 branches 5 tags Go to file Add file Code Use this template

supernord Update change_log.md and CITATION.cff e12ee40 on May 25, 2022 70 commits

test_data	Update tools.md to align with workflows.md v1.3 structure	last year
CITATION.cff	Update change_log.md and CITATION.cff	10 months ago
CITATION_blank_template.cff	Add blank *.cff file	10 months ago
LICENSE.md	Update tools.md to align with workflows.md v1.3 structure	last year
README.md	Add *.cff file	10 months ago
change_log.md	Update change_log.md and CITATION.cff	10 months ago
infrastructure_optimisation.md	Update infrastructure_optimisation.md	2 years ago
tools.md	Update tools.md to align with workflows.md v1.3 structure	last year
workflows.md	Update tools.md to align with workflows.md v1.3 structure	last year

https://github.com/AustralianBioCommons/doc_guidelines

Hints for maintaining your workflow!

Use
documentation
templates

Should I contribute to an existing community workflow effort instead?

The BioCommons guideline is based on learnings from multiple community efforts including the Galaxy Intergalactic Workflow Commission (IWC), nf-core and Snakemake.

If you are using Galaxy, Nextflow or Snakemake workflow languages and would like to contribute to their community workflow efforts, **you could be using their specific guidelines**, which are linked to below.

Community effort	Workflow language documentation	Workflow guideline information	How to create / contribute new workflows
IWC	Galaxy	IWC GitHub	IWC adding workflows
nf-core	Nextflow	nf-core developer guidelines	nf-core adding_pipelines
Snakemake	Snakemake	Snakemake-workflows guidelines	Snakemake-workflows contribute

<https://australianbiocommons.github.io/how-to-guides/documentation/DocumentationGuidelines>

Acknowledgements

The community of workflow developers



BioCommons 'Bring Your Own Data' Expansion Project

<https://www.biocommons.org.au/byo-data-platform-expansion>



Thanks!

Any questions?

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

Tell us what you thought ...

Feedback survey

surveymonkey.com/r/findable-workflows

NEXT ...

Unlocking nf-core: customising workflows for your research

18-19 May 2023

biocommons.org.au/webinars-workshops



Bring-Your-Own Workflow 5