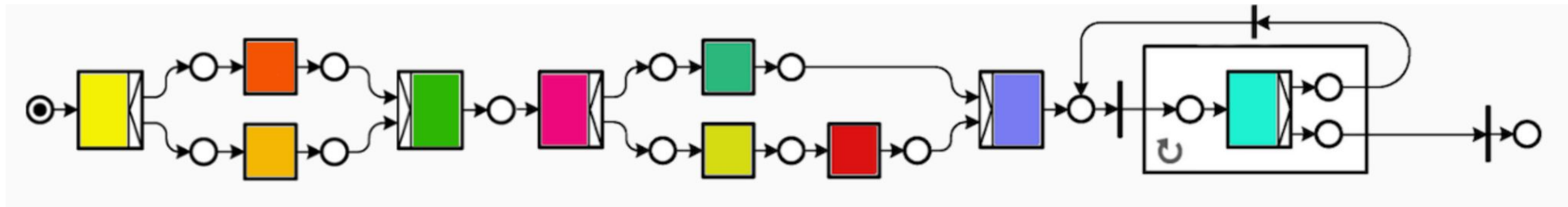


Make your computational workflows findable and citable

WorkflowHub BYOW workshop



Johan Gustafsson
Australian BioCommons

Gustafsson, J., & Samaha, G. (2023). WORKSHOP: Make your bioinformatics workflows findable and citable. Zenodo.
<https://doi.org/10.5281/zenodo.7787488>

Learning outcomes

Register with
WorkflowHub

Create and / or
join
WorkflowHub
Teams

Register
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 Baker^{1,2}, Marius van den Beek^{3,4}, Daniel Blankenberg^{5,6}, Dave Bouvier^{6,7}, John Chilton⁸, Nate Connor^{9,10}, Frederik Coppens^{11,12}, Ignacio Eguinoa^{13,14}, Simon Gladman^{15,16}, Björn Grünig¹⁷, Nicholas Kerner¹⁸, Delphine Lathière^{19,20}, Andrew Leslie²¹, Sergei Kosakovsky Pond^{22,23}, Wolfgang Maier²⁴, Anton Nekrutenko²⁵, James Taylor²⁶, Steven Weaver²⁷

1. Johns Hopkins University, Baltimore, Maryland, United States of America, 2. The Pennsylvania State University, University Park, Pennsylvania, United States of America, 3. Cleveland Clinic, Cleveland, Ohio, United States of America, 4. VIB Center for Plant Systems Biology, Ghent, Belgium, 5. Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, 6. University of Melbourne, Melbourne, Australia, 7. Queensland Cyber Infrastructure Foundation, St. Louis, Australia, 8. University of Freiburg, Freiburg im Breisgau, Germany, 9. Temple University, Philadelphia, Pennsylvania, United States of America

† Deceased.

* amartin@psu.edu (AM), spend@temple.edu (SKP)

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>.

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 in

OPEN ACCESS

Citation: Baker D, van den Beek M, Blankenberg D, Bouvier D, Chilton J, Connor N, et al. (2020) No more business as usual: Agile and effective responses to emerging pathogen threats require open data and open analytics. *PLoS Pathogens* 16(8): e1008643. <https://doi.org/10.1371/journal.ppat.1008643>

Editor: Carolyn B. Coyne, University of Pittsburgh, UNITED STATES

Published: August 13, 2020

Copyright: © 2020 Baker et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is funded by NIH Grant U41HG006620 and NSF AWR Grant 1661460 to AM and JT. Support for JT is supported by the German Federal Ministry of Education and Research grants 01KT1901C and 01KT1901B to BG. Galaxy and HUPPy integration is supported by NIH Grant R01HG006620 to AM. Integration into the Australian Government National Collaborative Research Infrastructure Strategy (HUPPy.org) development team is supported by NIH Grant R01HG006620 to AM. Integration into the Research Foundation Flanders (RVO) grant is

Check for updates

Cell

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

Graphical Abstract

Article

Authors

Darren P. Martin, Steven Weaver, Houriyah Tegally, ..., David L. Robertson, Tullio de Oliveira, Sergei L. Kosakovsky Pond

Correspondence

darrenpatrickmartin@gmail.com (D.P.M.), tuliodn@gmail.com (T.d.O.), spend@temple.edu (S.L.K.P.)

In brief

An analysis of synonymous and non-synonymous mutations in SARS-CoV-2 genomes since the inception of the COVID-19 pandemic provides insights into the emergence of a convergent mutational signature in the 501Y lineage (alpha, beta, and gamma variants) that is also likely present in other lineages that impacts host immune recognition.

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

Check for updates

correspondence

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

To the Editor — The COVID-19 pandemic is the first health crisis characterized by large amounts of genomic data. Computational infrastructure can be a bottleneck for data analysis, impeding global cooperation 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 computation. Here, we present the COVID-19 effort by the Galaxy Project, which pools free worldwide public computational infrastructure, making the analysis of deep sequencing data accessible to anyone while also providing an analytical framework for the global pathogen genomic surveillance based on raw sequencing read data.

Fig. 1. Analysis flow for calling SARS-CoV-2 variants using Galaxy. ONT, Oxford Nanopore Technologies; VCF, variant call format; TSV, tab-separated values; PE, paired-end; SE, single-end. For more information, see <https://covid19.galaxyproject.org>.

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

Han Mei,¹ Sergei Kosakovsky Pond,² and Anton Nekrutenko^{3,4}

¹Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA
²Department of Biology, Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, USA
³Corresponding author: amartin@psu.edu
Associate Editor: Aya Takahashi

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.

Coronaviruses have large 26–32 kbp positive-strand RNA genomes. The initial 1/3 of the genome is occupied by an open reading frame (ORF) ORF1ab encoding nonstructural proteins essential for the coronavirus life cycle. As the designation “ORF” suggests, it contains two reading frames with the 3’-end of ORF1a overlapping with the 5’-terminus of ORF1b. ORF1b is in a +1 phase relative to ORF1a and translated via the –1 programmed ribosomal frameshifting controlled by the PFE. As ORF1b encodes crucial components of coronavirus replication/translation, machines, including the RNA stretch of sequence from “YK” in the slippery heptamer to the stop codon of ORF1a. The position of the ORF1a stop codon determines overlap length. For example, in SARS-CoV-2, it is 16 bp, while in mouse hepatitis virus (MHV) it is 22 bp (Pang et al. 2016).

Our group has been interested in the evolutionary dynamics of overlapping coding regions (Nekrutenko et al. 2005; Chung et al. 2007; Solaryk et al. 2007). The vast amount of newly generated sequence and functional data—a result of the current SARS-CoV-2 pandemic—provides an

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	lofreq
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	lofreq

 Australian BioCommons

Workflows are important

362

workflow languages

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

756

forks for nf-core
workflow rnaseq

<https://github.com/nf-core/rnaseq>

Workflows are important

362

workflow languages

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

4,630

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

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

1,880

of workflows on
Galaxy Europe instance

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

756

forks for nf-core
workflow rnaseq

<https://github.com/nf-core/rnaseq>

916

workflows in WorkflowHub
(all public)

<https://workflowhub.eu/workflows>

908

of workflows on
Galaxy Australia instance

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

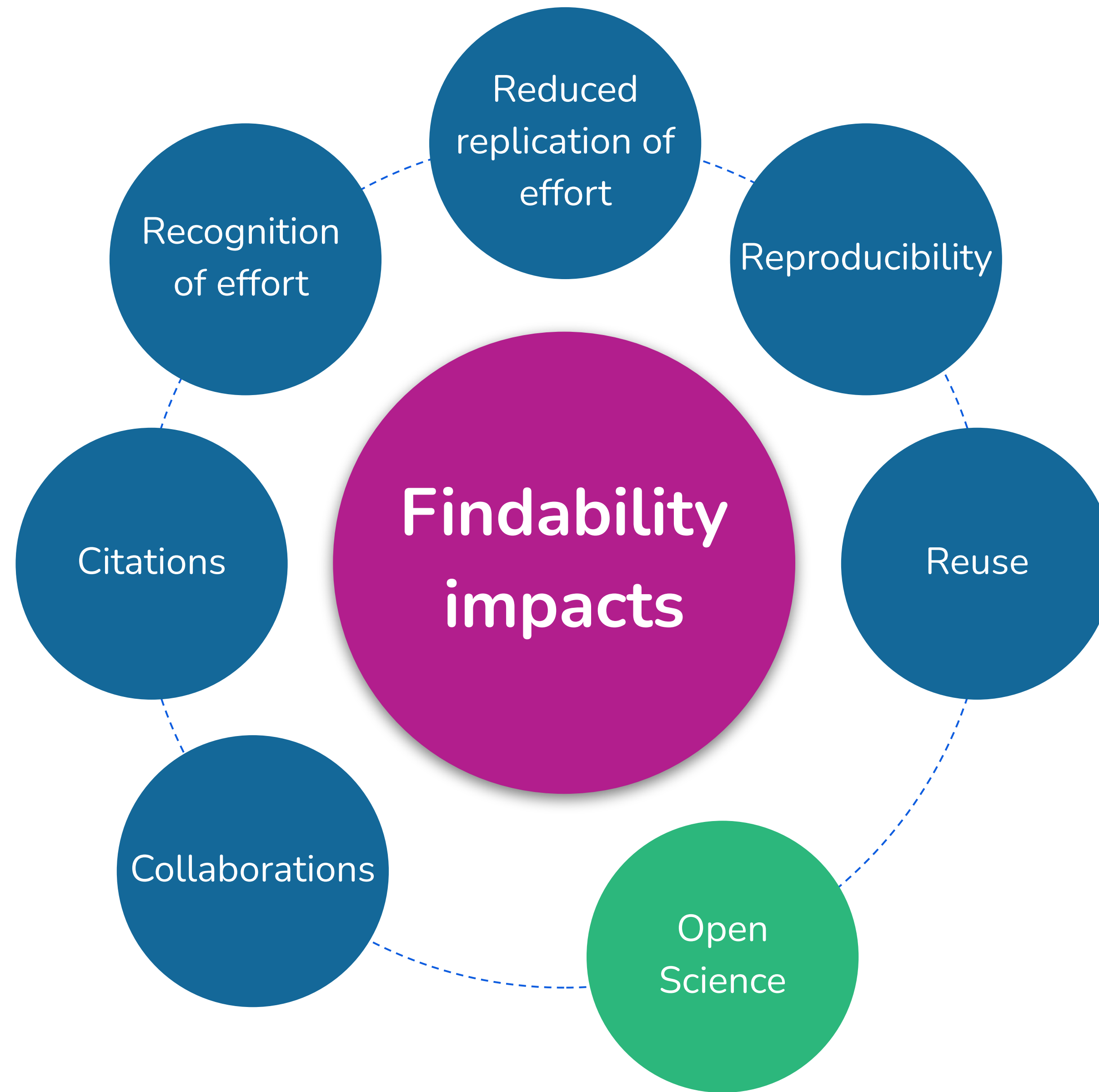
Workflows can also be ...

- ✓ 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



A global community effort to collect a curated set of
open-source analysis pipelines built using Nextflow.

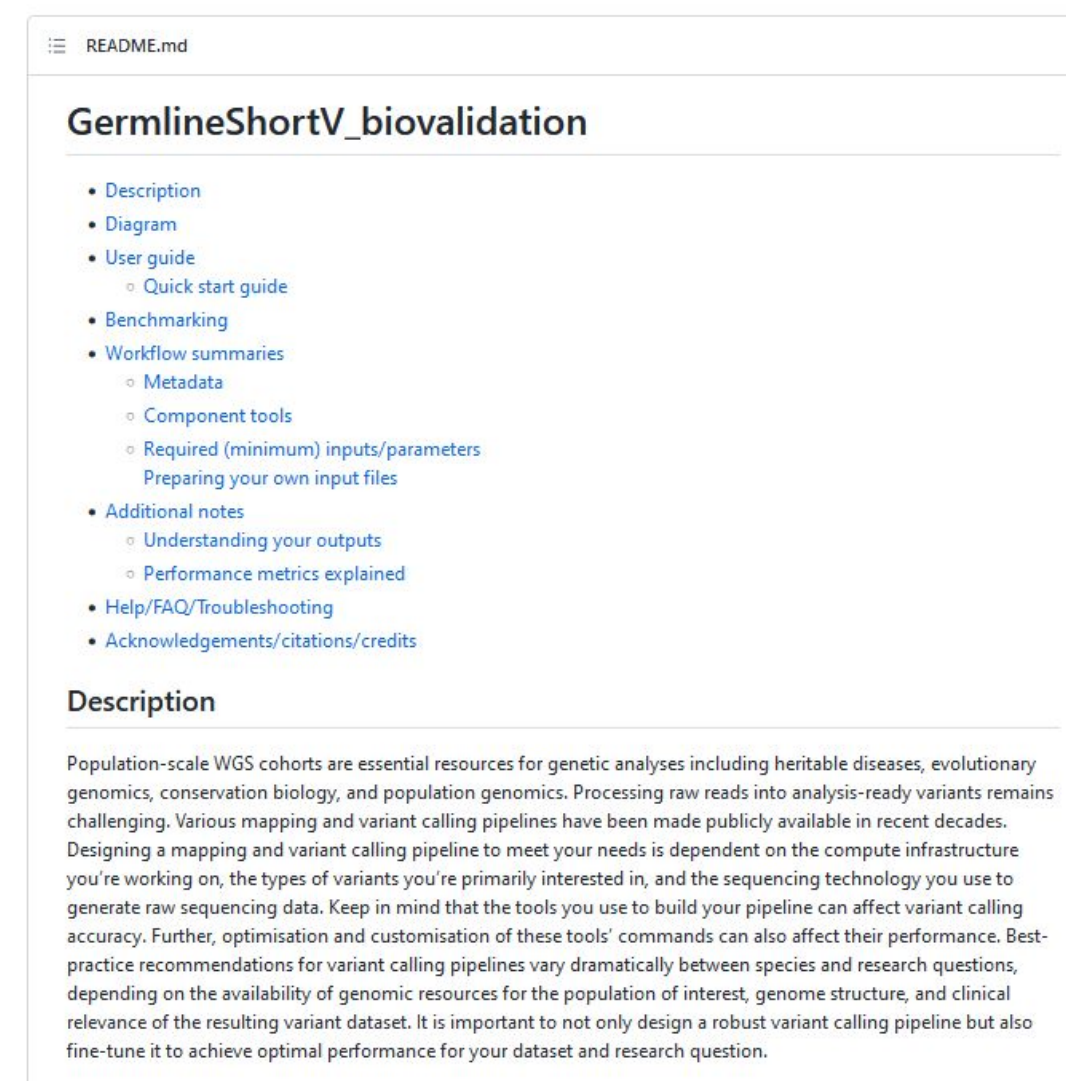
Multiple places you *could* find workflows

Platforms / community repositories



A global community effort to collect a curated set of open-source analysis pipelines built using Nextflow.

GitHub



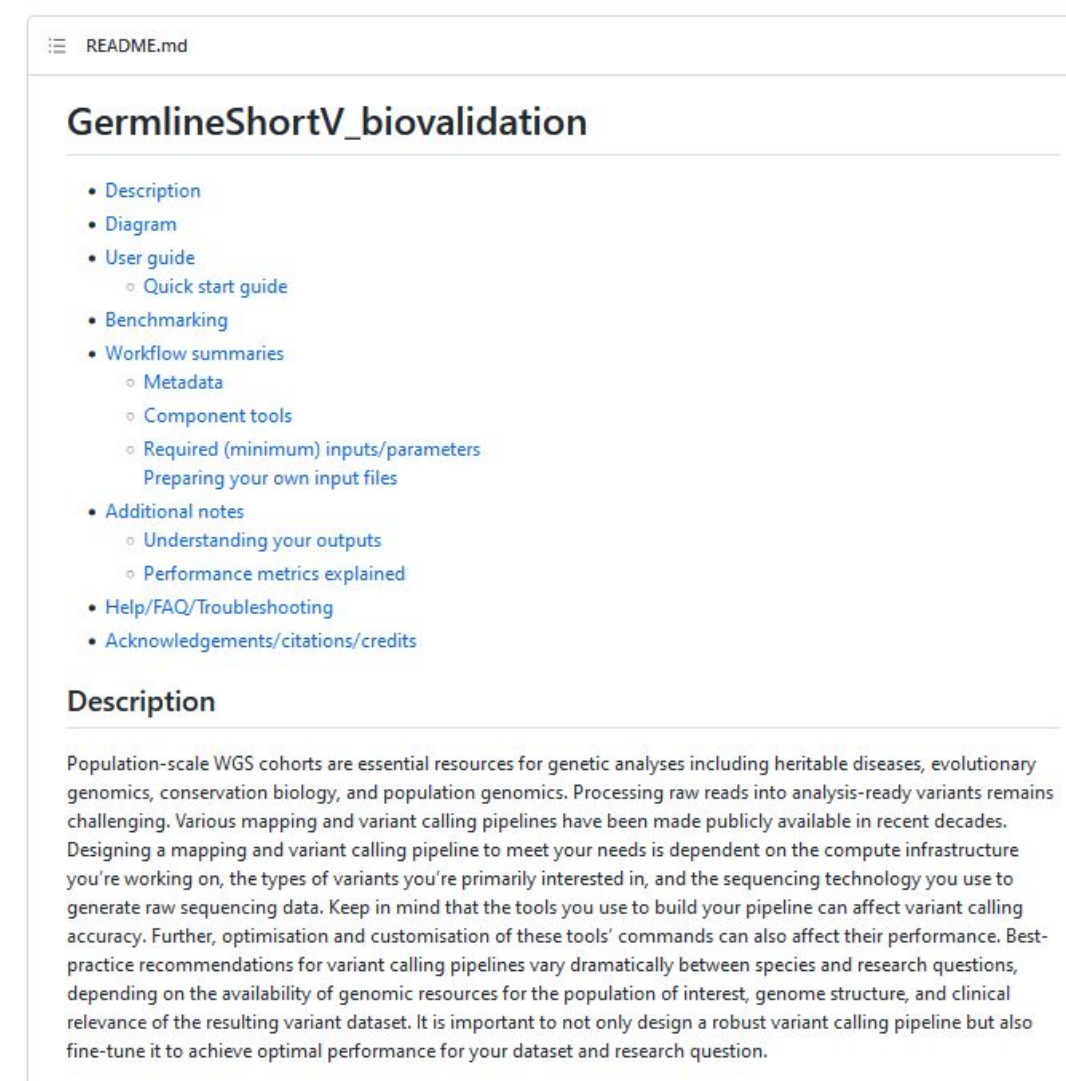
Multiple places you *could* find workflows

Platforms / community repositories



A global community effort to collect a curated set of open-source analysis pipelines built using Nextflow.

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>

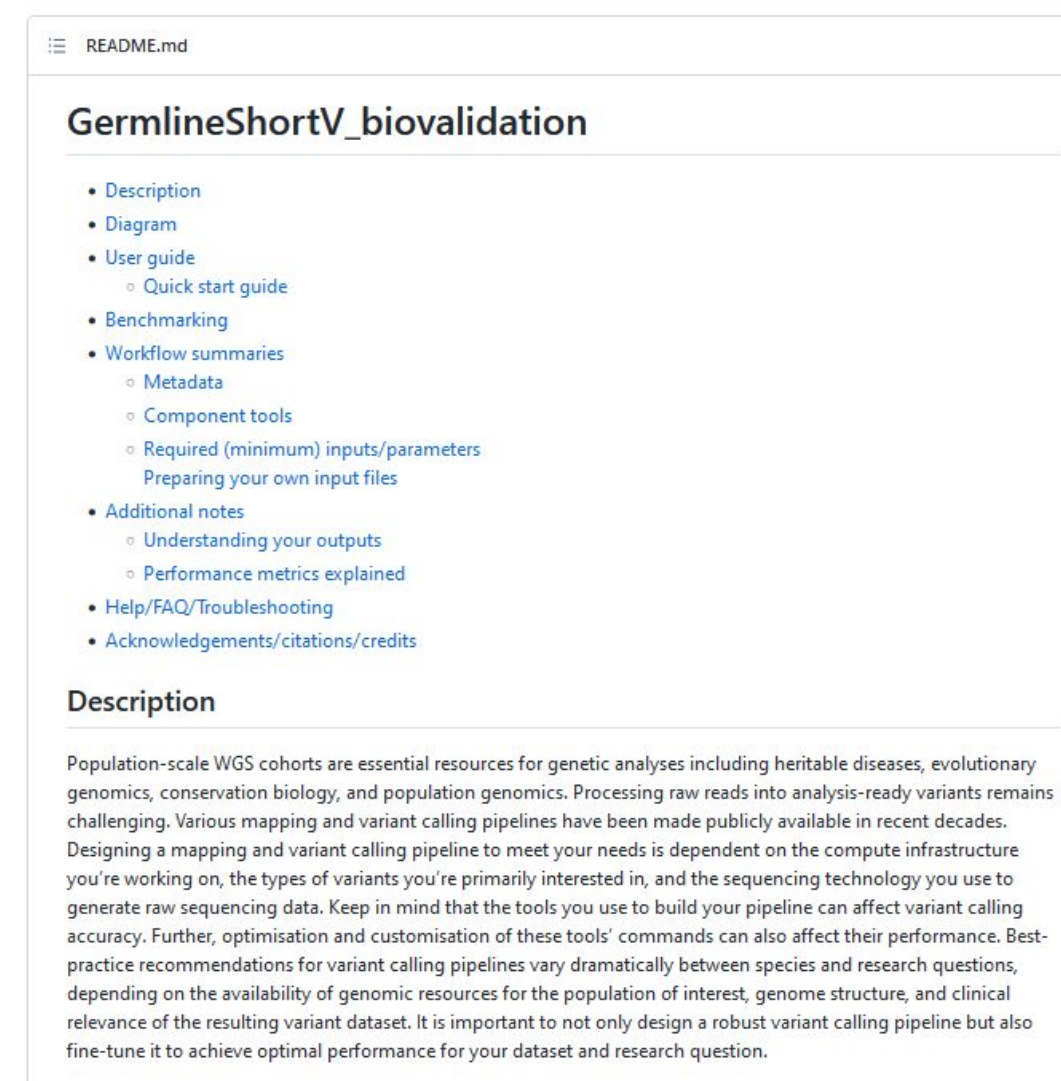
Multiple places you *could* find workflows

Platforms / community repositories



A global community effort to collect a curated set of open-source analysis pipelines built using Nextflow.

GitHub



Publications

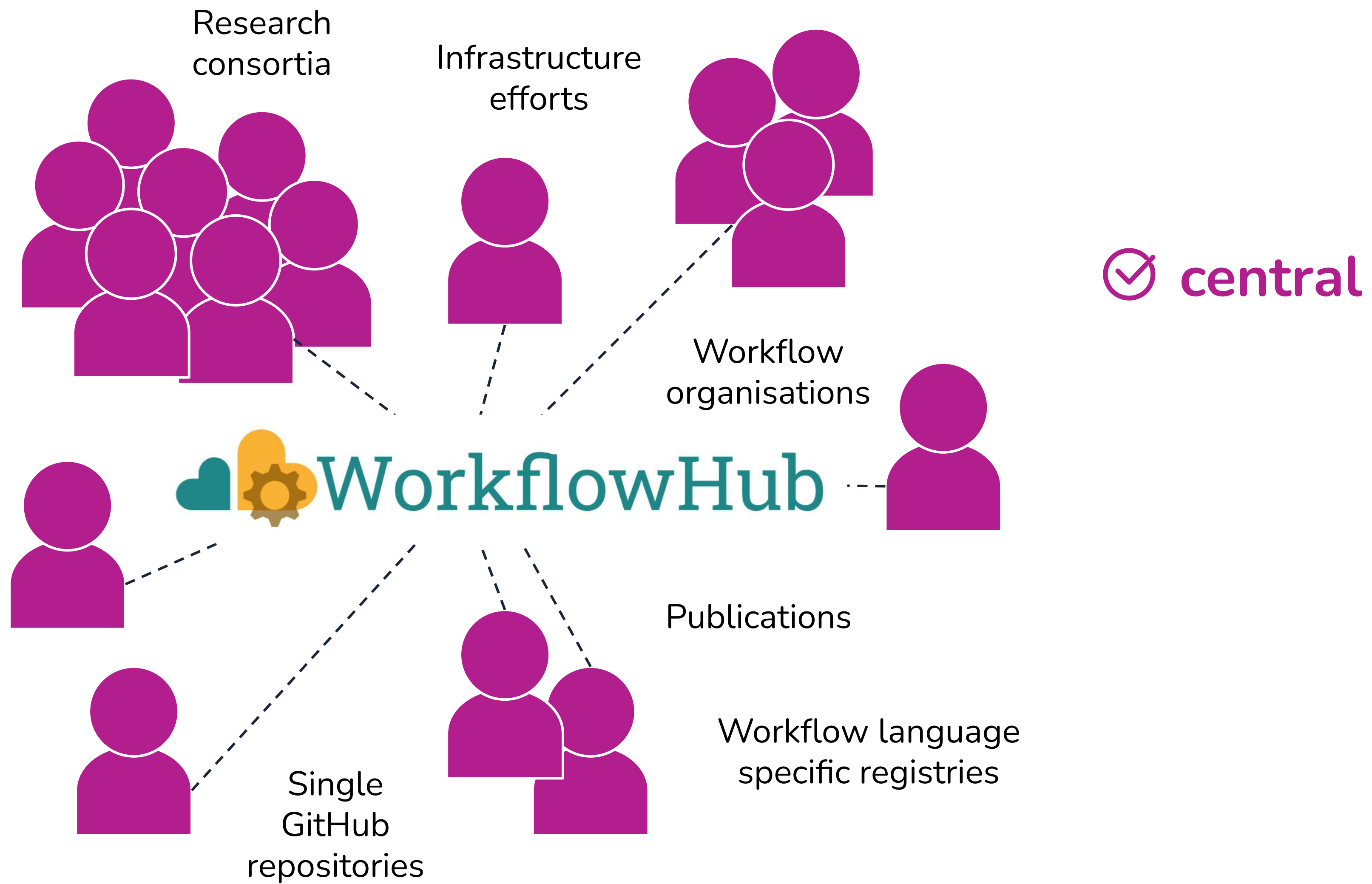
Murigneux, V., Roberts, L.W., Forde, B.M. et al. MicroPIPE: validating an end-to-end workflow for high-quality complete bacterial genome construction. *BMC Genomics* 22, 474 (2021). <https://doi.org/10.1186/s12864-021-07767-z>

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>

Search engines

Consider a registry

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



Workflows

What is a Workflow?

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

Tag: Assembly ✕

Search here... Go

Last update date (Descending)

Created At

Any time

Updated At

Any time

Tool

BLAST

antiSMASH

fastp

MultiQC

SAMtools

SRA Software Toolkit

More...

Workflow type

Galaxy

Common Workflow Language

Shell Script

Nextflow

Default

Condensed

Table

2: Plant virus confirmation

Creators: None

Teams: Integrated and Urban Plant Pathology Laboratory

Version: 1

LongRead Quality Control and Filtering

Creators: Bart Nijse, Jasper Koehorst, Germán Royval

Teams: UNLOCK

Version: 1

(Hybrid) Metagenomics workflow

Creators: Bart Nijse

V-pipe (main multi-virus version)

Creators: Ivan Topolinski

Purge duplicates from hifiasm assembly v1.0

Creators: Gareth P

Shotgun-Metagenomics-Analysis

Creators: Cali Wilk, Sadsad, Tracy Che

genome assembly nextflow workflow

All Images Videos News Books More Tools

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>



Bioschemas

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, P...
 We introduce the concept of Canonical describing and wrapping computational
 ☆ Save Cite Cited by 1 Related

[PDF] Creating lightweight FAIR
 S Soiland-Reyes, P Sefton, LJ Castro...
 RO-Crate (Soiland-Reyes et al. 2022) is along with their metadata, based on Lin
 ☆ Save Cite Related articles A

[PDF] Species-specific exudates community composition
 E Briand, M Lassudrie, C Noël, C Jauze...
 Phytoplankton-bacteria interactions are 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.WORK

- ✓ 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 2.1.0

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

Workflows

My workflows

Workflows shared by others

hifi

Sort by:

▼ Name

Update time

 Filter:

Show deleted

Show bookmarks

edited over 2 years ago

never run

☆

▼

PacBio HiFi genome assembly using hifiasm v2.1

Assembly, visualisation and quality control workflow... ▼

fastq

hifiasm

HiFi

genome_assembly

Add Tags

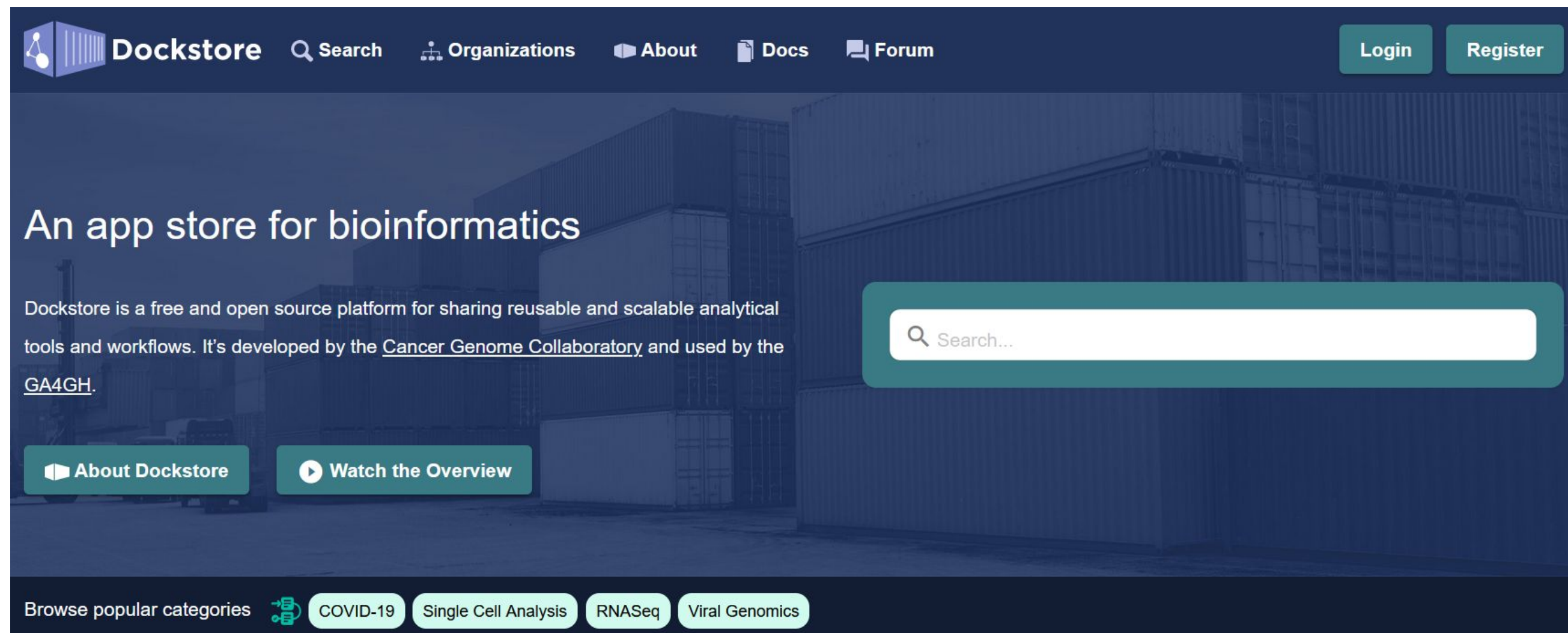
Edit

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

 **Galaxy**
PROJECT

Two key options

Dockstore



<https://dockstore.org/>

O'Connor BD, Yuen D 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>)

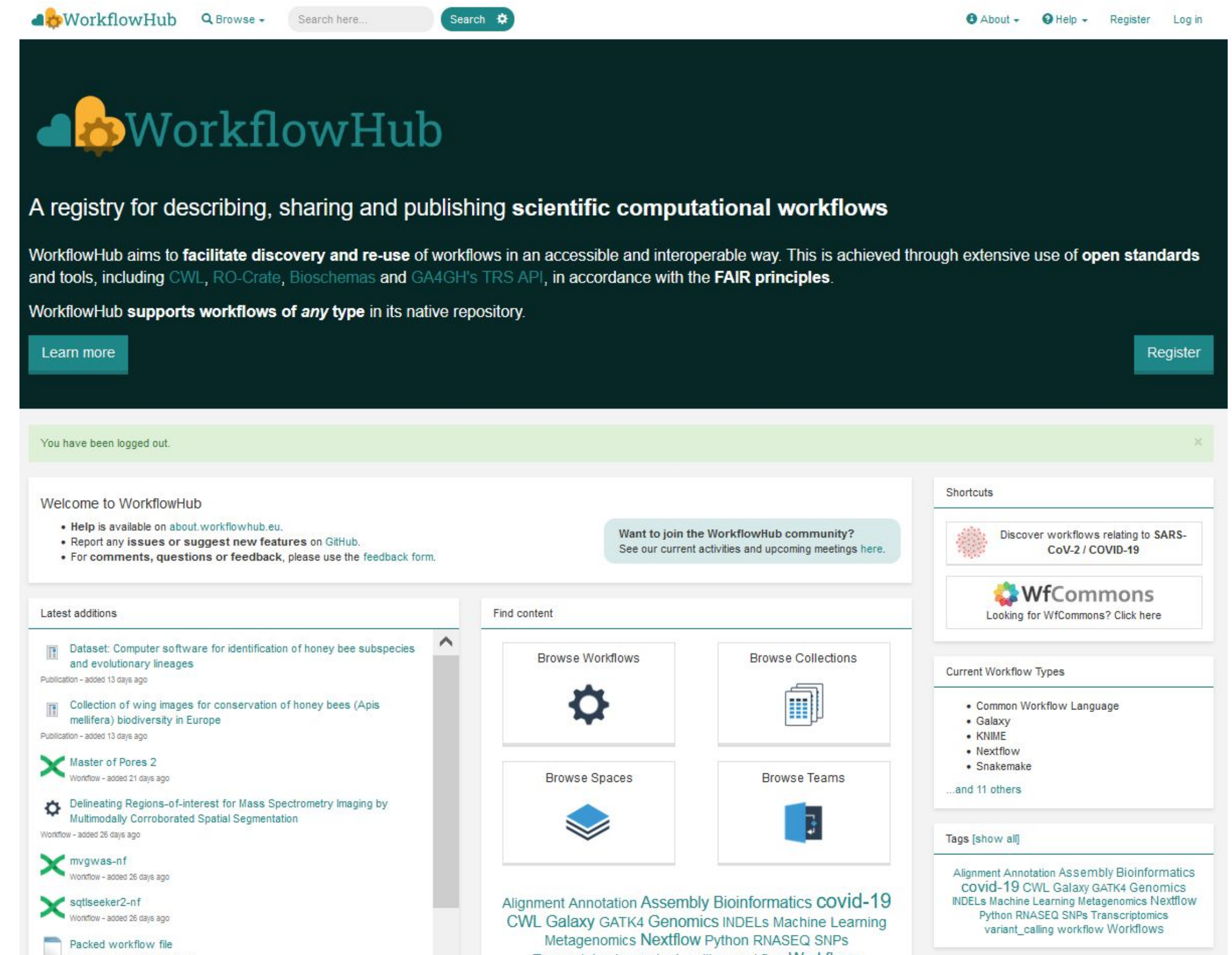
Two key options



<https://workflowhub.eu/>

Gustafsson, O. J. R., Wilkinson, S. R. et al. (2024). WorkflowHub: a registry for computational workflows. arXiv preprint arXiv:2410.06941.

<https://doi.org/10.48550/arXiv.2410.06941>





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**



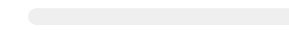
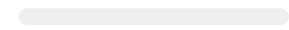
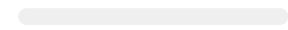
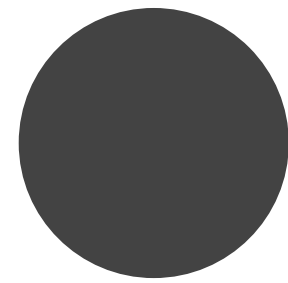
WfCommons

Looking for WfCommons? [Click here](#)

Current Workflow Types

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

Register



Register an account for WorkflowHub

Username

The username should contain a minimum of 3 characters.

Email address

Password

Password should contain a minimum of 10 characters.

Confirm Password

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

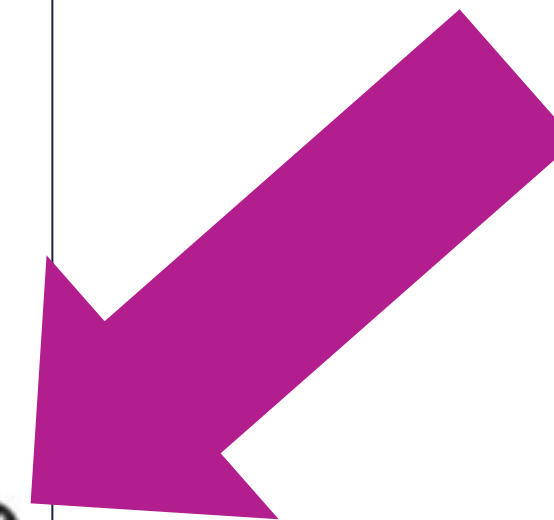
[Already registered? - goto Login](#)

Alternatively...

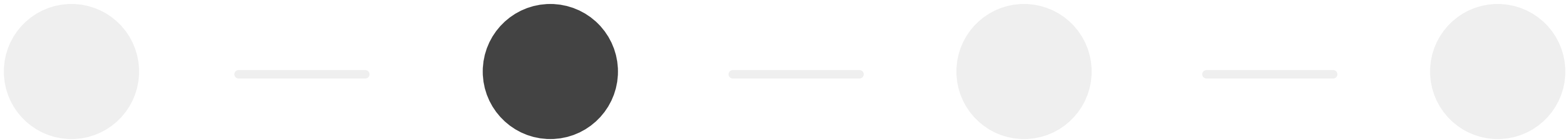
- Log in using LS Login




- 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

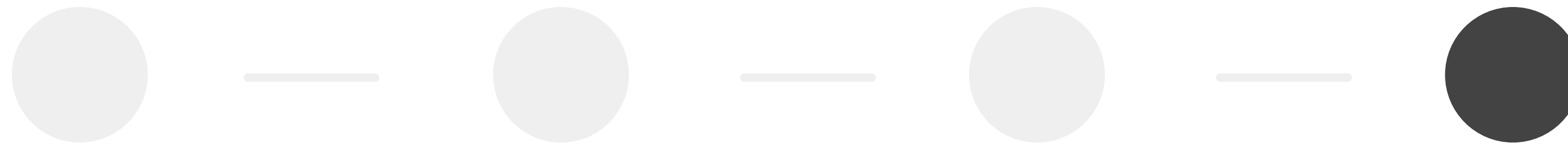
Country

Select a country

Submit

One wizard
accommodates
these processes

Add items



Space

Team

Workflow

Publication

Workflow

Document

Workflow

SOP

Collection

Add items

Space

Team

QCIF Bioinformatics

Overview

Related items

Related items

People (5)

Spaces (1)

Organizations (1)

Publications (1)

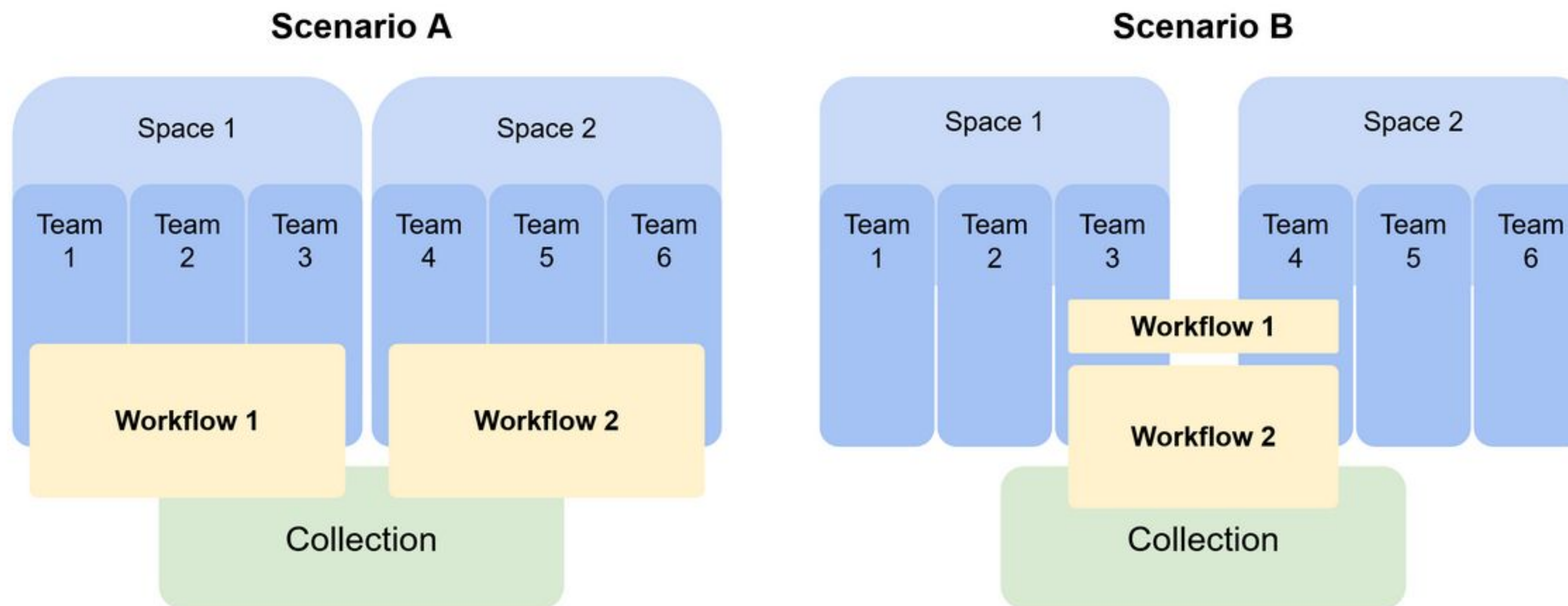
Documents (1)

Workflows (20)

Collections (2)

- Australian BioCommons
 - Australian BioCommons
 - QCIF Bioinformatics
 - MicroPIPE: validating an end-to-end workflow for high-quality complete bacterial genome construction.
 - 16S Microbial Analysis with mother - howto
 - microPIPE: a pipeline for high-quality bacterial genome construction using ONT and Illumina sequencing
 - 16S_biodiversity_for_overlap_paired_end
 - 16S_biodiversity_for_nonoverlap_paired_end
 - scRNAseq Single Sample Processing STARSolo
 - scRNAseq_CellQC
 - scRNAseq_QCtoBasicProcessing
 - scRNAseq: Load counts matrix
 - scRNAseq: Count and Load with starSOLO
 - scRNAseq Single Sample Processing Counts Matrix
 - Analyses of shotgun metagenomics data with MetaPhlAn2
 - scRNAseq: Count and Load with Cell Ranger
 - scRNAseq Single Sample Processing Cell Ranger
 - Workflow 1: Further Quality Control [16S Microbial Analysis With Mothur]
 - Workflow 2: Data Cleaning And Chimera Removal [16S Microbial Analysis With Mothur]
 - Workflow 3: Classification [Galaxy Training: 16S Microbial Analysis With Mothur]
 - Workflow 5: OTU Clustering [16S Microbial Analysis With Mothur]
 - Workflow 6: Alpha Diversity [16S Microbial Analysis With Mothur]
 - Workflow 7 : Beta Diversity [16S Microbial Analysis With Mothur]
 - Taxonomy classification using Kraken2 and Bracken
 - ONTViSc (ONT-based Viral Screening for Biosecurity)
 - scRNAseq processing in galaxy
 - 16S Microbial Analysis with mothur (on Galaxy Australia)

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*

Let's register a workflow!

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](#).



Discover



Contribute



About

My WorkflowHub

Shortcuts



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



WfCommons

Looking for WfCommons? [Click here](#)

Quick start



Create
Team

Create a new Team

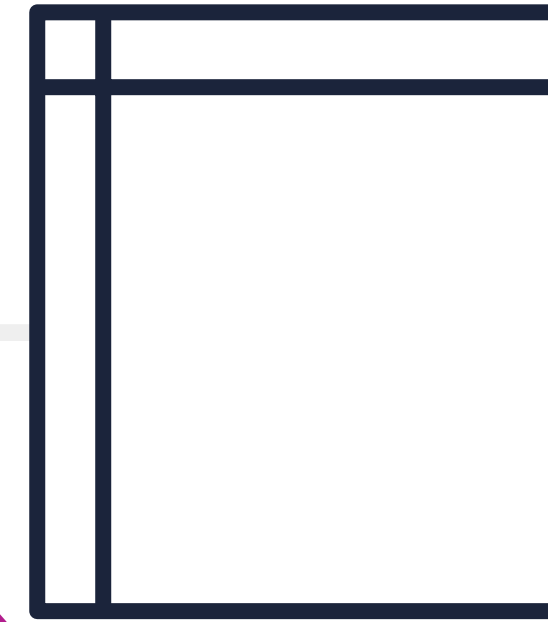
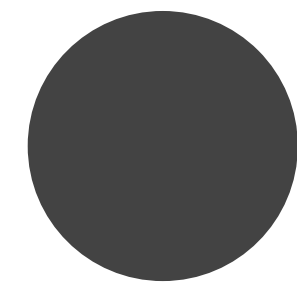


Join
Team

Request
membership of an
existing Team

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.

Upload/Import Files

Import Git Repository

Upload/Import Workflow RO-Crate

URL *

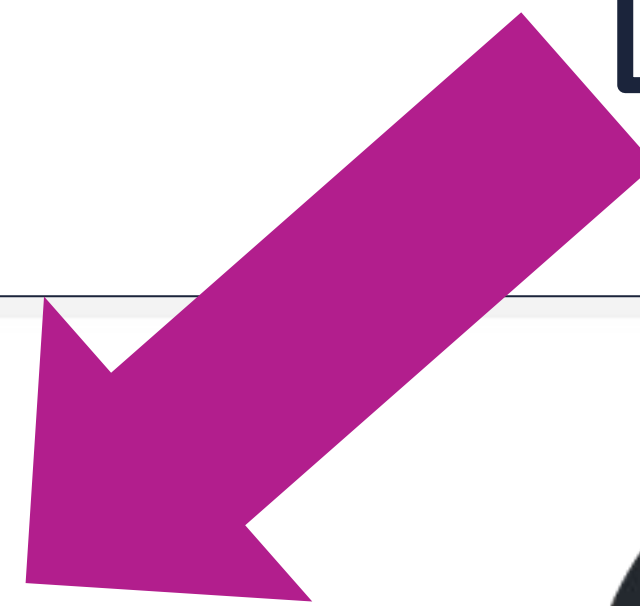
URL to the git repository. Should end in ".git".

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

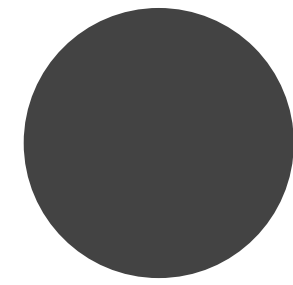
Register

or

Cancel



Select your file



import method



Upload/Import Files

Import Git Repository

Upload/Import Workflow RO-Crate

Workflow RO-Crate *

The zipped Workflow RO-Crate.

Local file

Remote URL

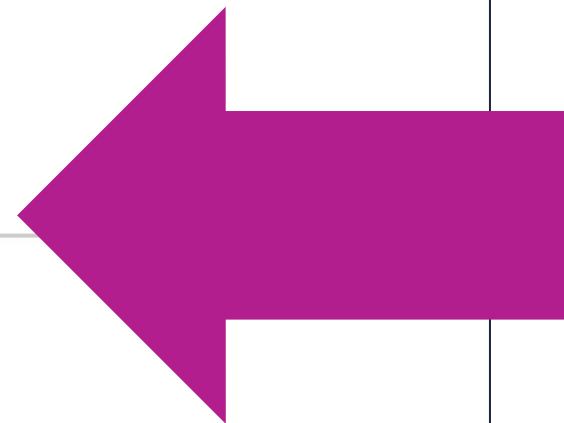
Browse...

No file selected.

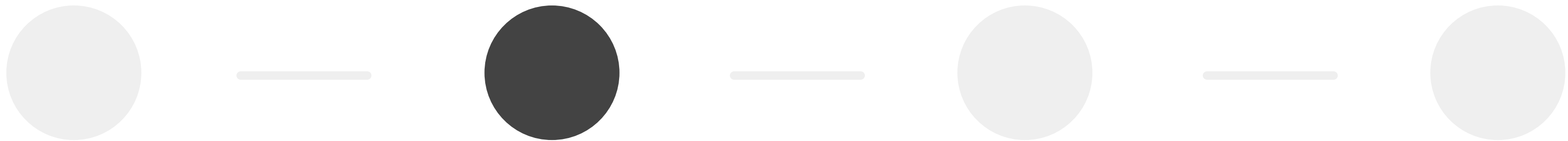


<https://www.researchobject.org/ro-crate/>

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



Import



Upload/Import Files

Import Git Repository

Upload/Import Workflow
RO-Crate

URL *

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

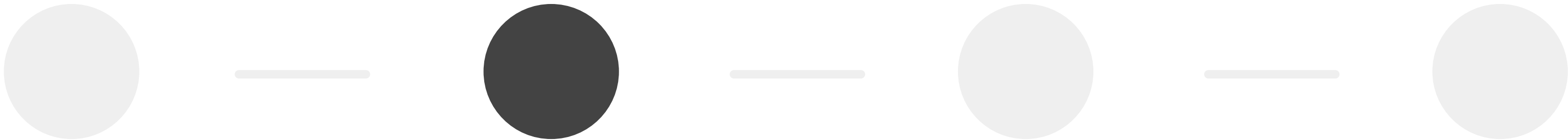
https://github.com/example/repository.git

Register

 or

Cancel

Import



Select Target

Branches

Tags

☐ v1.0.0

☐ v2.0.0

☒ v2.1.0

Register

or

Cancel

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

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

Galaxy-Workflow-PacBio_HiFi_genome_assembly_using_hifiasm_v2.1.ga

README.md

change_log.md

infrastructure_optimisation.md


workflows.md

Select

Cancel

Add metadata

New Workflow










Workflow Type* 

Galaxy

Title*

PacBio HiFi genome assembly using hifiasm v2.1

Description

B I H         

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

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

Select a team...

Workflow type

Update description if needed

Add EDAM ontology terms

Teams

License



Apache Software License 2.0

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

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


View suggestions 


Creators

Type to search for creators., or Add new creator


1 Gareth Price

No affiliation specified



2 Katherine Farquharson 

No affiliation specified



Additional credit


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

Add license (auto if license in GitHub)

Set visibility

Custom tags

Edit creators



Australian
BioCommons

Add metadata

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

New Workflow

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.

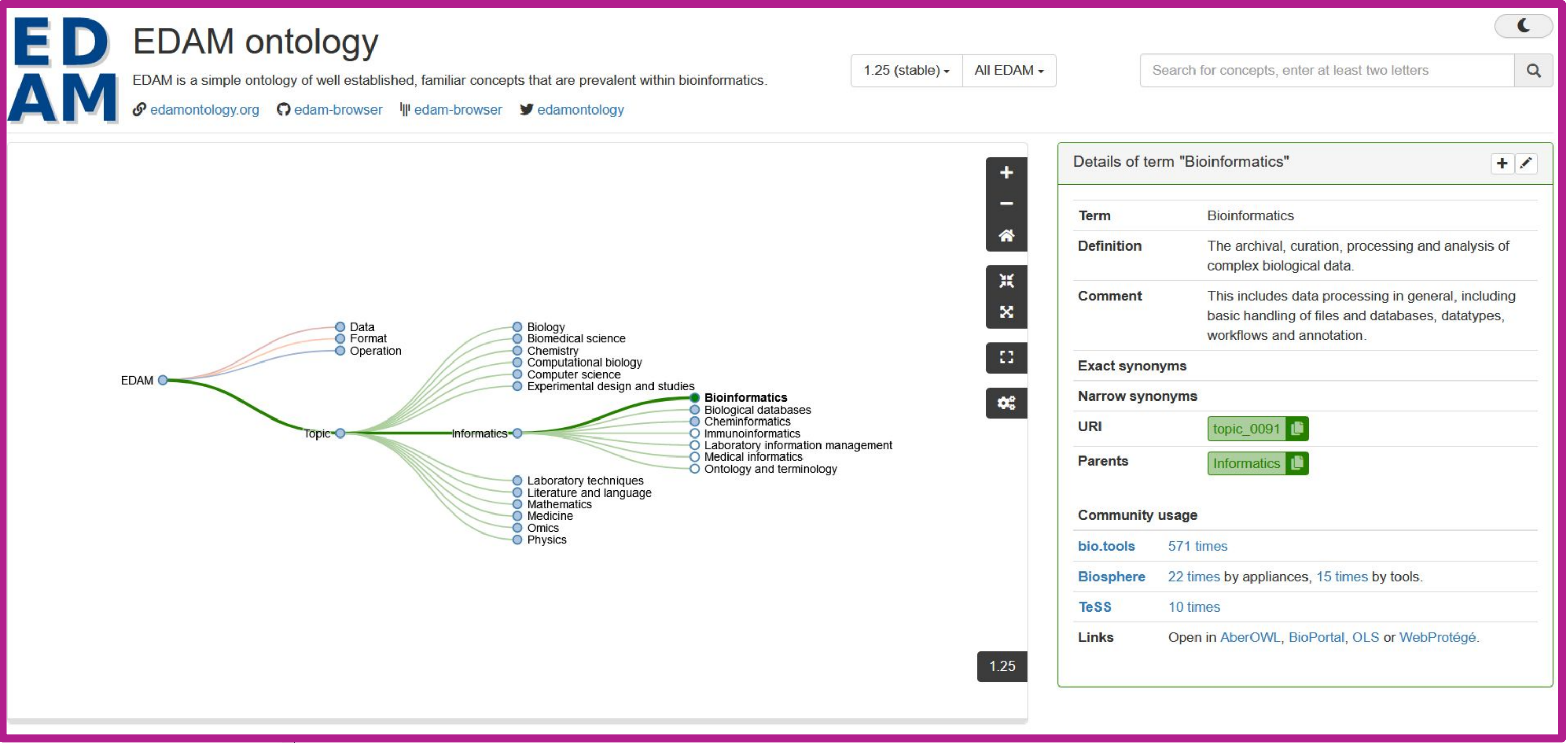
Teams

The following teams are associated with this Workflow:

Australian BioCommons Dev

Select a team...

Add EDAM
ontology terms



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 unig Haplotype info	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_info/0.8.1+galaxy1
6	Processed unig Haplotype 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_fa/gfa_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

Diagram

The screenshot displays the WorkflowHub interface for the 'GermlineShortV_biovalidation' workflow (Version 1). The page is organized into several sections:

- Overview:** Includes tabs for Overview, Files, and Related items. The workflow type is 'Shell Script'.
- Structured documentation:** A sidebar on the left lists various documentation sections: Description, Diagram, User guide (with sub-items like Quick start guide), Benchmarking, Workflow summaries (with sub-items like Metadata, Component tools, Required inputs/parameters), Additional notes (with sub-items like Understanding outputs, Performance metrics), Help/FAQ/Troubleshooting, and Acknowledgements/citations/credits.
- Diagram:** A decision tree diagram titled 'What type of organism are you working with?' branches into three categories: Human, Non-human model organism, and Non-model organism. Each category has a corresponding 'Sample-based metrics' box, followed by 'Variant-based metrics' and 'Known variant concordance' boxes. A 'Biological' box is at the bottom. A dashed box labeled 'vcfstat.sh' encloses the 'Variant-based metrics' and 'Known variant concordance' boxes for the Human and Non-human model organism categories.
- Metadata:** On the right, a sidebar contains: Creators and Submitter (listing Georgina Samaha, Tracy Chew, Cali Willet, and Nandan Deshpande), Citation (with a 'Copy' button), License (GNU General Public License 3.0), Activity (Views: 1155, Created: 5th May 2022 at 06:02), Annotated Properties (Topic annotations: Genetic variation, Bioinformatics, Whole genome sequencing; Operation annotations: SNP detection, Indel detection, Validation), Tags (with a note 'This item has not yet been tagged.' and an 'Add your tags' button), and Attributions (None).

Complete author list
(in correct order)

DOI

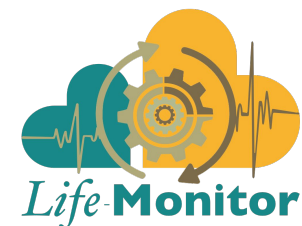
License

Annotations

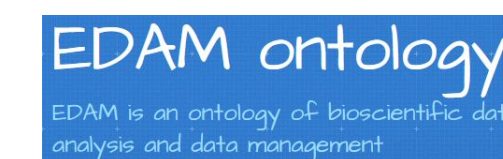
Samaha, G., Chew, T., Willet, C., & Deshpande, N. (2022). *GermlineShortV_biovalidation*. WorkflowHub. <https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.339.1>

Acknowledgements

The community of workflow developers



Snakemake



Thanks!

Any questions?