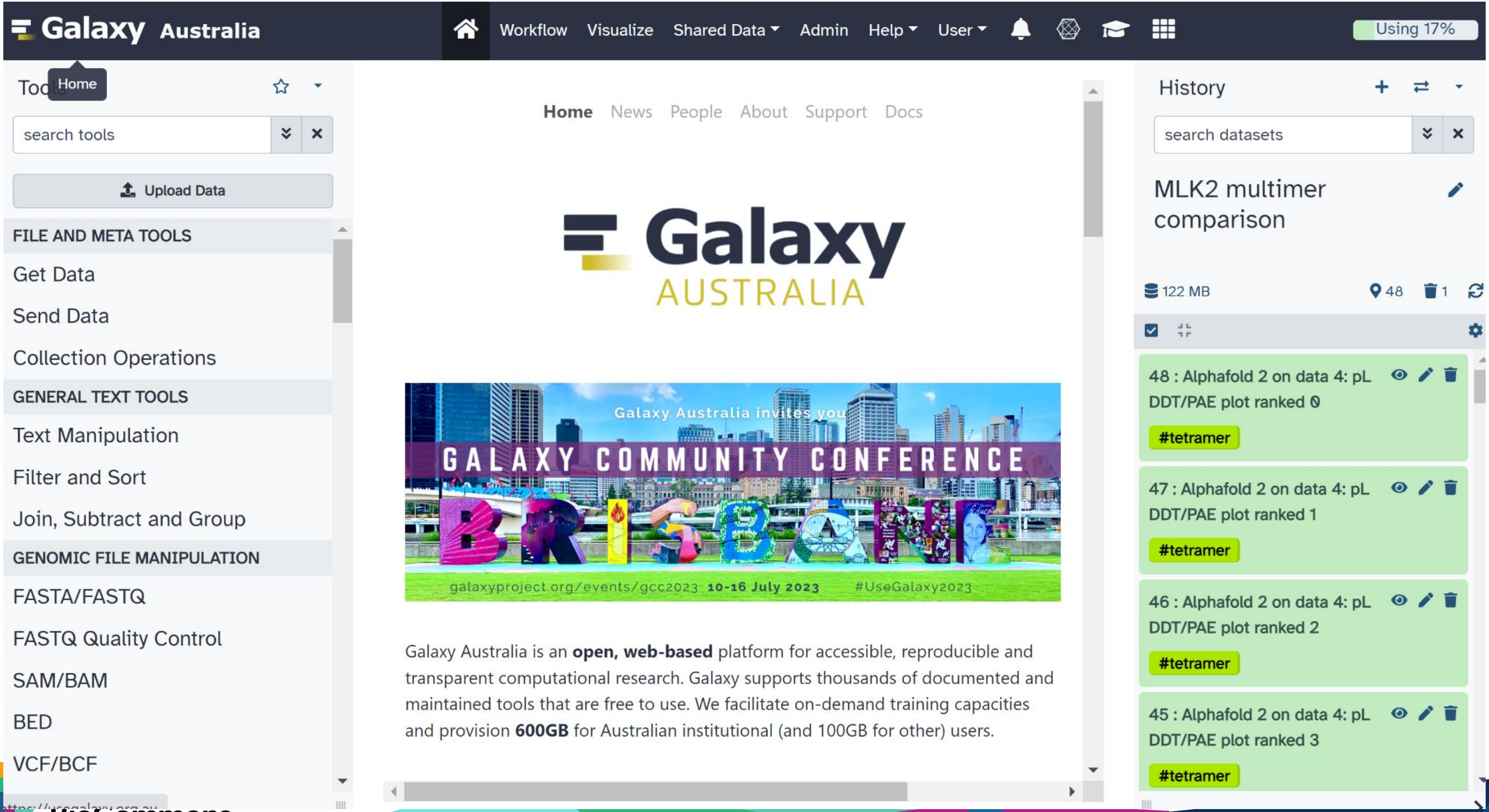
Managing hands-on data analysis training with Galaxy

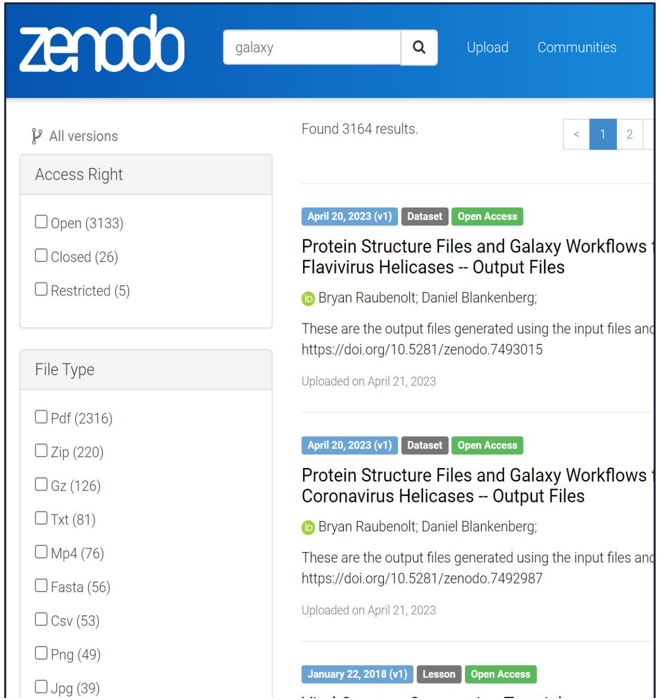
Dr Gareth Price, Project Lead Galaxy Australia
Saskia Hiltemann & Helena Rasche, Galaxy Training Network

Australian
BioCommons



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

Learning Support through Galaxy



All versions

Access Right

☐ Open (3133)
☐ Closed (26)
☐ Restricted (5)

File Type

☐ Pdf (2316)
☐ Zip (220)
☐ Gz (126)
☐ Txt (81)
☐ Mp4 (76)
☐ Fasta (56)
☐ Csv (53)
☐ Png (49)
☐ Jpg (39)

Found 3164 results.

April 20, 2023 (v1) Dataset Open Access

Protein Structure Files and Galaxy Workflows f
Flavivirus Helicases -- Output Files

Bryan Raubenolt; Daniel Blankenberg;

These are the output files generated using the input files and
https://doi.org/10.5281/zenodo.7493015

Uploaded on April 21, 2023

April 20, 2023 (v1) Dataset Open Access

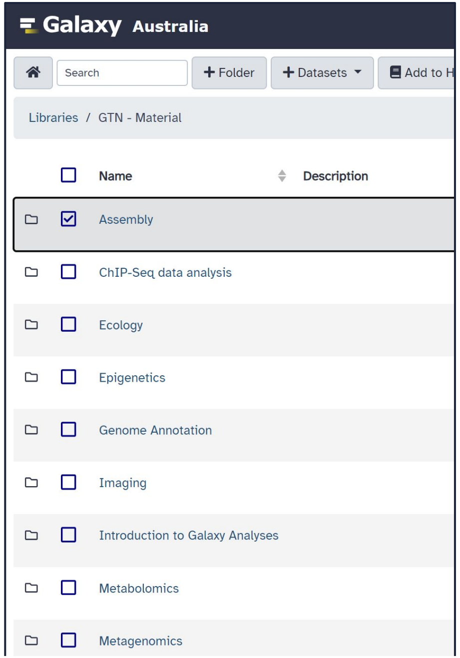
Protein Structure Files and Galaxy Workflows f
Coronavirus Helicases -- Output Files

Bryan Raubenolt; Daniel Blankenberg;

These are the output files generated using the input files and
https://doi.org/10.5281/zenodo.7492987

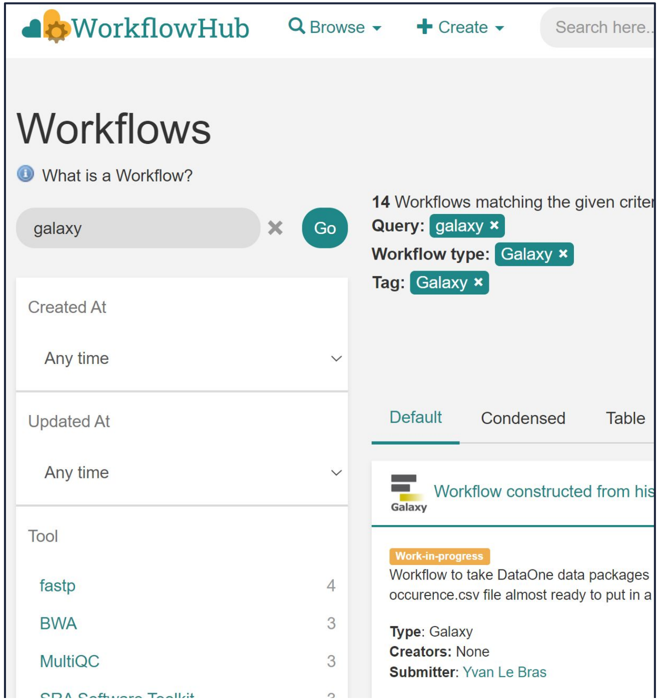
Uploaded on April 21, 2023

January 22, 2018 (v1) Lesson Open Access



Libraries / GTN - Material

	☐	Name	Description
	☑	Assembly	
	☐	ChIP-Seq data analysis	
	☐	Ecology	
	☐	Epigenetics	
	☐	Genome Annotation	
	☐	Imaging	
	☐	Introduction to Galaxy Analyses	
	☐	Metabolomics	
	☐	Metagenomics	



Workflows

What is a Workflow?

galaxy ✖ Go

14 Workflows matching the given criter
Query: galaxy ×
Workflow type: Galaxy ×
Tag: Galaxy ×

Created At
Any time

Updated At
Any time

Tool

fastp	4
BWA	3
MultiQC	3

Workflow constructed from his

Work-in-progress

Workflow to take DataOne data packages
occurence.csv file almost ready to put in a

Type: Galaxy
Creators: None
Submitter: Yvan Le Bras

Australian BioCommons
QCIF
Melbourne Bioinformatics
aarnet Australia's Academic and Research Network
The University of Queensland Australia
ARDC Australian Research Data Commons
The University of Melbourne
Bioplatforms Australia
NCI Australia
Pawsey
NCRIS

An overview of Galaxy Australia training

Australian BioCommons
LaTrobe University
QCIF
The University of Sydney
University of New South Wales
University of Technology Sydney
Griffith University
Royal Melbourne Hospital
Queensland University of Technology
The University of Melbourne
The University of Queensland

2021-22 FY	
Australian Training attendance	619
Australian Teaching attendance	543
Global Training attendance	1998
Total	**3160**

2022-23 FY	
Australian Training attendance	320
Australian Teaching attendance	1041
Global Training attendance	1030
Total	**2391**

Galaxy Australia contributions to GTN

Cameron
Hyde

Simon
Gladman

Anna
Syme

Catherine
Bromhead

Nuwan
Goonasekera

Managing hands-on data analysis training with Galaxy

Saskia Hiltemann & Helena Rasche

2023-07-25

The Galaxy Data Analysis Platform

- User-friendly
- Only a browser needed

- Many tools available
- No installation required
- **Free!**

- Build your own workflows
- Share your work
- ..

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

1111 : Make.shared on data 1108 and data 1109: shared

1110 : Cluster.split on data 1108, data 1106, and data 1107: sensspec

1109 : Cluster.split on data 1108, data 1106, and data 1107: list (OTU List)

1108 : Remove.lineage on data 1072, data 1071, and data 1073: pick.count_table

1107 : Remove.lineage on data 1072, data 1071, and data 1073: pick.fasta

1106 : Remove.lineage on data 1072, data 1071, and data 1073: pick.taxonomy

1105 : Sub.sample on data 1101: logfile

1104 : Sub.sample on data 11

Try Galaxy Now!

- The “Big Three”
    - Galaxy Main (usegalaxy.org)
    - Galaxy Europe (usegalaxy.eu)
    - **Galaxy Australia (usegalaxy.org.au)**

- Many other smaller, often domain-specific Galaxies available
    - List of all public Galaxies (150+): galaxyproject.org/use

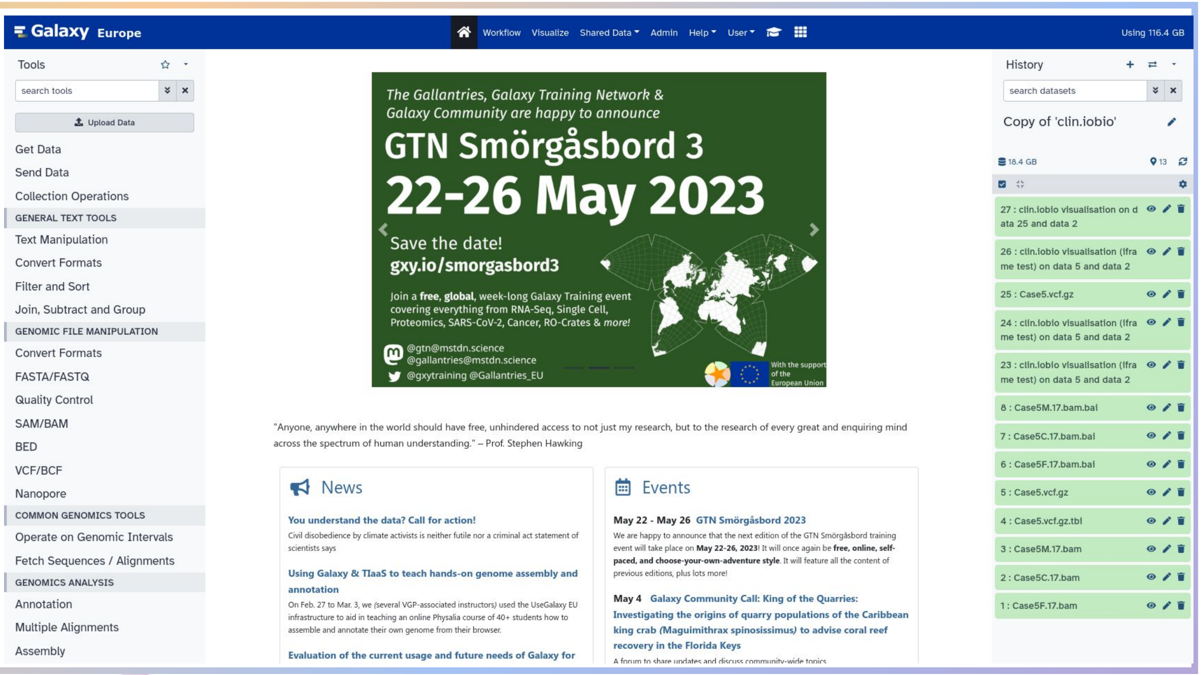
Galaxy for Training

Teach bioinformatics without the hassle

Concepts, not command line!

You’re not teaching Galaxy, you’re teaching the science.

Bring a web browser



The Gallantries, Galaxy Training Network & Galaxy Community are happy to announce

GTN Smörgåsbord 3

22-26 May 2023

Save the date!
gxy.io/smorgasbord3

Join a **free**, **global**, week-long Galaxy Training event covering everything from RNA-Seq, Single Cell, Proteomics, SARS-CoV-2, Cancer, RO-Crates & *more!*

@gtn@mstdn.science
@gallantries@mstdn.science
@gxytraining @Gallantries_EU

With the support of the European Union

"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." – Prof. Stephen Hawking

News

You understand the data? Call for action!
Civil disobedience by climate activists is neither futile nor a criminal act statement of scientists says

Using Galaxy & TIaaS to teach hands-on genome assembly and annotation
On Feb. 27 to Mar. 3, we *(*several VGP-associated instructors*)* used the UseGalaxy EU infrastructure to aid in teaching an online Physalia course of 40+ students how to assemble and annotate their own genome from their browser.

Evaluation of the current usage and future needs of Galaxy for

Events

May 22 - May 26 **GTN Smörgåsbord 2023**
We are happy to announce that the next edition of the GTN Smörgåsbord training event will take place on **May 22-26, 2023**! It will once again be **free, online, self-paced, and choose-your-own-adventure style**. It will feature all the content of previous editions, plus lots more!

May 4 **Galaxy Community Call: King of the Quarries: Investigating the origins of quarry populations of the Caribbean king crab *(Maguimithrax spinosissimus)* to advise coral reef recovery in the Florida Keys**
A forum to share updates and discuss community-wide topics

Why use Galaxy for Teaching?

Galaxy is an **amazing** teaching platform

🎓 **Focus on the science**, not the technical details of tools

💻 **No installation required**, only requirement is a browser

📚 Huge **library** of free, **high quality tutorials**

📊 **Visualizations** of results and workflows

📨 **Share** data with students, and have them share their work back.

👩‍💻 Enable **remote teaching** & **follow their progress** with TIaaS

Wait, there's more..

🛠️ **Lots of tools available**, ~8000 in the Galaxy Tool shed

🖥️ Combine Galaxy with **programming environments** e.g. Jupyter and Rstudio

📚 **Shared data libraries** for sharing input datasets

📊 **Interactive tools** such as genome browsers, visualisation tools

👩‍💻 Large **community of Galaxy instructors**

How to support education with Galaxy?

Materials

350+ GTN Tutorials

Galaxy FAQs

Video Library
- Automated
- Human recorded

Galaxy

Shared data libraries

Tutorial mode:
- Click-to-run Tools
- Click-to-run Workflows

Compute

TIaaS
- Platform
- Stats
- Dashboard

Training Content:

Galaxy Training Network (GTN)

A short history of the GTN

The before-times

A short history of the GTN

2016

The before-times *Tutorials & infra*

A short history of the GTN

2016

The before-times *Tutorials & infra*

A short history of the GTN

The before-times

2016

Tutorials & infra

Cell Systems

Volume 6, Issue 6, 27 June 2018, Pages 752-758.e1

Report

Community-Driven Data Analysis Training for Biology

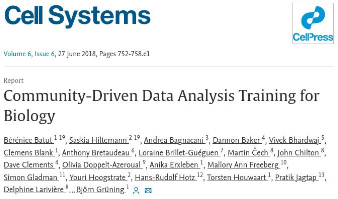
Bérénice Batut[1][19], Saskia Hiltemann[2][19], Andrea Bagnacani[3], Dannon Baker[4], Vivek Bhardwaj[5], Clemens Blank[1], Anthony Bretaudeau[6], Loraine Brillet-Guéguen[7], Martin Čech[8], John Chilton[8], Dave Clements[4], Olivia Doppelt-Azeroual[9], Anika Erxleben[1], Mallory Ann Freeberg[10], Simon Gladman[11], Youri Hoogstrate[2], Hans-Rudolf Hotz[12], Torsten Houwaart[1], Pratik Jagtap[13], Delphine Larivière[8]...Björn Grüning[1]

2018

Paper!

A short history of the GTN

2016



Cell Systems

Volume 6, Issue 6, 27 June 2018, Pages 752-758.e1

Report

Community-Driven Data Analysis Training for Biology

Bérénice Batut[1, 19], Saskia Hiltemann[2, 19], Andrea Bagnacani[3], Dannon Baker[4], Vivek Bhardwaj[5], Clemens Blank[1], Anthony Bretaudeau[6], Loraine Brillet-Guéguen[7], Martin Čech[8], John Chilton[8], Dave Clements[4], Olivia Doppelt-Azeroual[9], Anika Erxleben[1], Mallory Ann Freeberg[10], Simon Gladman[11], Youri Hoogstrate[2], Hans-Rudolf Hotz[12], Torsten Houwaart[1], Pratik Jagtap[13], Delphine Larivière[8]…Björn Grüning[1]

2018

The before-times *Tutorials & infra* **Paper!**

A short history of the GTN

2016

Cell Systems

Volume 6, Issue 6, 27 June 2018, Pages 752-758.e1

Report

Community-Driven Data Analysis Training for Biology

Bérénice Batut [1] [19], Saskia Hiltemann [2] [19], Andrea Bagnacani [3], Dannon Baker [4], Vivek Bhardwaj [5], Clemens Blank [1], Anthony Bretaudeau [6], Loraine Brillet-Guéguen [7], Martin Čech [8], John Chilton [8], Dave Clements [4], Olivia Doppelt-Azeroual [9], Anika Erxleben [1], Mallory Ann Freeberg [10], Simon Gladman [11], Youri Hoogstrate [2], Hans-Rudolf Hotz [12], Torsten Houwaart [1], Pratik Jagtap [13], Delphine Larivière [8]...Björn Grüning [1]

2018

The before-times *Tutorials & infra* **Paper!** *Teaching*

A short history of the GTN

Cell Systems

Volume 6, Issue 6, 27 June 2018, Pages 752-758.e1

Report

Community-Driven Data Analysis Training for Biology

Bérénice Batut [1 19], Saskia Hiltemann [2 19], Andrea Bagnacani [3], Dannon Baker [4], Vivek Bhardwaj [5], Clemens Blank [1], Anthony Bretaudeau [6], Loraine Brillet-Guéguen [7], Martin Čech [8], John Chilton [8], Dave Clements [4], Olivia Doppelt-Azeroual [9], Anika Erxleben [1], Mallory Ann Freeberg [10], Simon Gladman [11], Youri Hoogstrate [2], Hans-Rudolf Hotz [12], Torsten Houwaart [1], Pratik Jagtap [13], Delphine Larivière [8]…Björn Grüning [1]

2016

2018

2020

The before-times

Tutorials & infra

Paper!

Teaching

Pandemic

A short history of the GTN

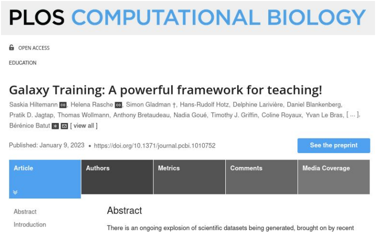
Cell Systems

Volume 6, Issue 6, 27 June 2018, Pages 752-758.e1

Report

Community-Driven Data Analysis Training for Biology

Bérénice Batut[1][19], Saskia Hiltemann[2][19], Andrea Bagnacani[3], Dannon Baker[4], Vivek Bhardwaj[5], Clemens Blank[1], Anthony Bretaudeau[6], Loraine Brillet-Guéguen[7], Martin Čech[8], John Chilton[8], Dave Clements[4], Olivia Doppelt-Azeroual[9], Anika Erxleben[1], Mallory Ann Freeberg[10], Simon Gladman[11], Youri Hoogstrate[2], Hans-Rudolf Hotz[12], Torsten Houwaart[1], Pratik Jagtap[13], Delphine Larivière[8]…Björn Grüning[1]



PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

EDUCATION

Galaxy Training: A powerful framework for teaching!

Saskia Hiltemann, Helena Rasche, Simon Gladman †, Hans-Rudolf Hotz, Delphine Larivière, Daniel Blankenberg, Pratik D. Jagtap, Thomas Wollmann, Anthony Bretaudeau, Nadia Goué, Timothy J. Griffin, Coline Royaux, Yvan Le Bras, […], Bérénice Batut [view all]

Published: January 9, 2023 • https://doi.org/10.1371/journal.pcbi.1010752

Abstract

There is an ongoing explosion of scientific datasets being generated, brought on by recent

2016

2018

2020

2023

The before-times

Tutorials & infra

Paper!

Teaching

Pandemic

Paper!

10.1371/journal.pcbi.1010752

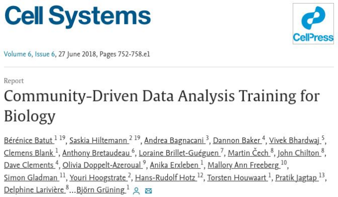
A short history of the GTN

Topic Maintainers

Adopt a Topic

310

Contributors



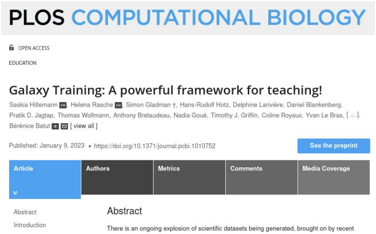
Cell Systems

Volume 6, Issue 6, 27 June 2018, Pages 752-758.e1

Report

Community-Driven Data Analysis Training for Biology

Bérénice Batut[1][19], Saskia Hiltemann[2][19], Andrea Bagnacani[3], Dannon Baker[4], Vivek Bhardwaj[5], Clemens Blank[1], Anthony Bretaudeau[6], Loraine Brillet-Guéguen[7], Martin Čech[8], John Chilton[8], Dave Clements[4], Olivia Doppelt-Azeroual[9], Anika Erxleben[1], Mallory Ann Freeberg[10], Simon Gladman[11], Youri Hoogstrate[2], Hans-Rudolf Hotz[12], Torsten Houwaart[1], Pratik Jagtap[13], Delphine Larivière[8]...Björn Grüning[1]



PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

EDUCATION

Galaxy Training: A powerful framework for teaching!

Saskia Hiltemann, Helena Rasche, Simon Gladman †, Hans-Rudolf Hotz, Delphine Larivière, Daniel Blankenberg, Pratik D. Jagtap, Thomas Wollmann, Anthony Bretaudeau, Nadia Goué, Timothy J. Griffin, Coline Royaux, Yvan Le Bras, [...], Bérénice Batut [view all]

Published: January 9, 2023 • https://doi.org/10.1371/journal.pcbi.1010752

Abstract

There is an ongoing explosion of scientific datasets being generated, brought on by recent

2016

2018

2020

2023

The before-times

Tutorials & infra

Paper!

Teaching

Pandemic

Paper!

GTN Catalog of Tutorials

Large number of tutorials

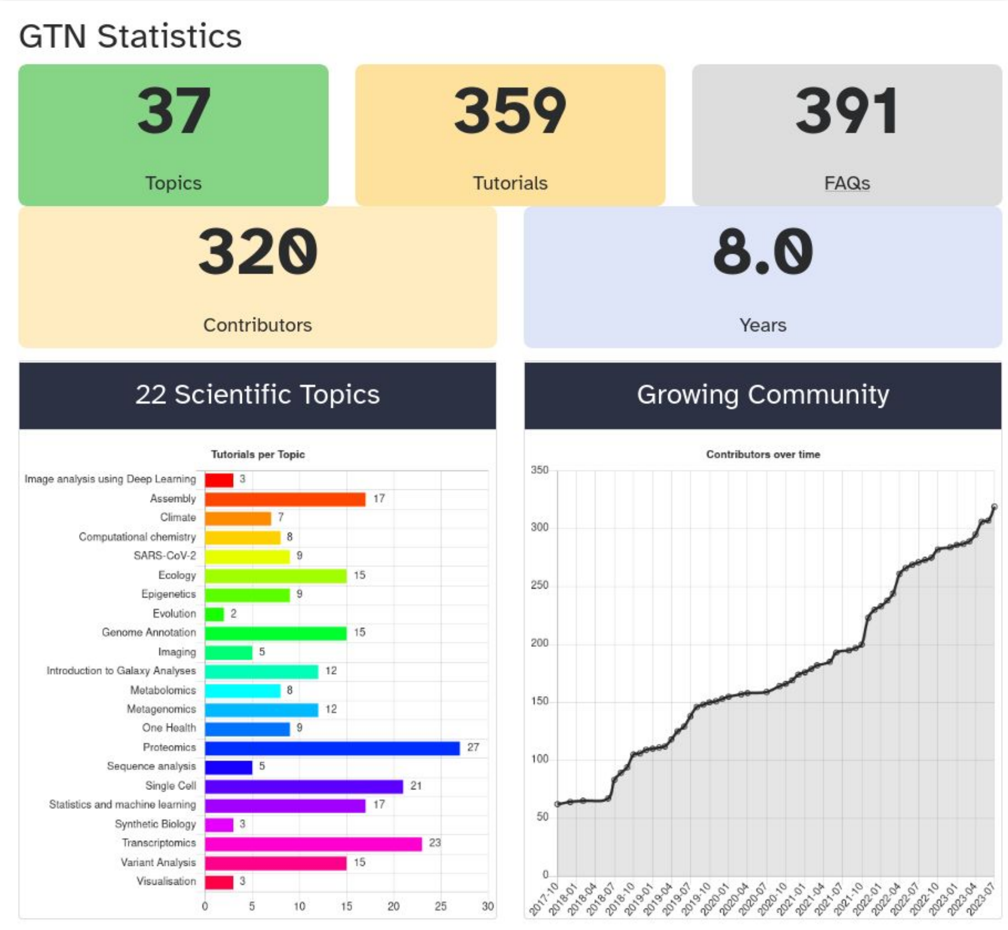
Many different **scientific topics**

Free to use for anybody

Also suitable for **self-study**

Anybody can **contribute** updates and new tutorials

Community-driven



GTN Statistics

37 Topics

359 Tutorials

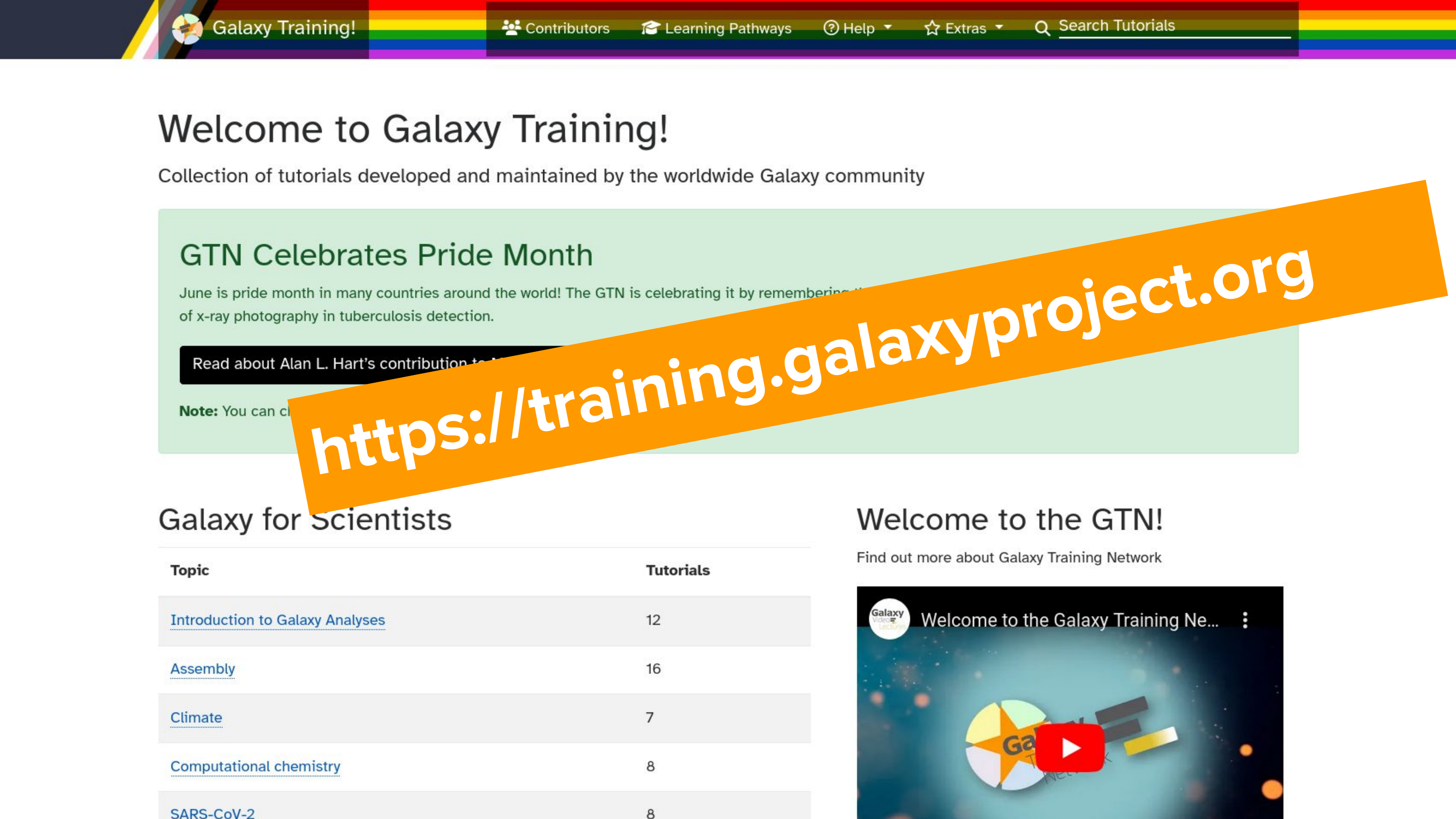
391 FAQs

320 Contributors

8.0 Years

22 Scientific Topics

Growing Community



https://training.galaxyproject.org

Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

GTN Celebrates Pride Month

June is pride month in many countries around the world! The GTN is celebrating it by rememberin
of x-ray photography in tuberculosis detection.

Read about Alan L. Hart's contribution t

Note: You can c

Galaxy for Scientists

Topic	Tutorials
Introduction to Galaxy Analyses	12
Assembly	16
Climate	7
Computational chemistry	8
SARS-CoV-2	8

Welcome to the GTN!

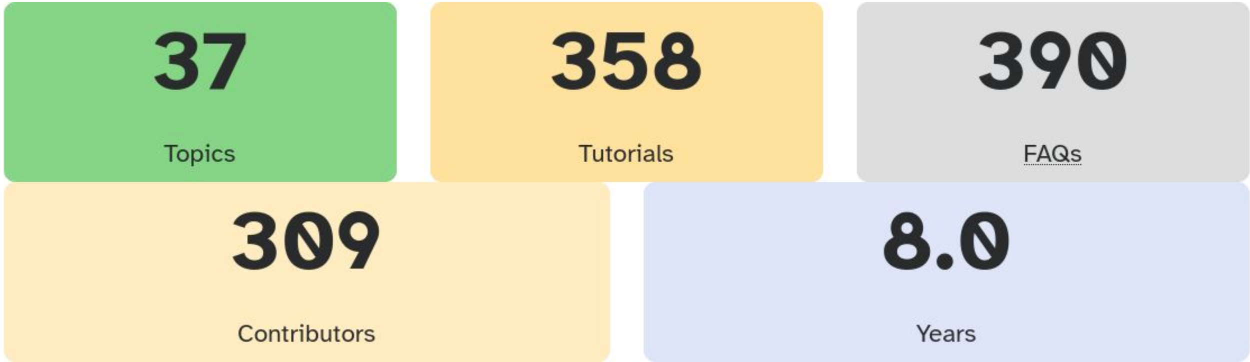
Find out more about Galaxy Training Network

The Galaxy Training Network

By the community,
for the community

Wide range of Scientific & Technical Topics

New Publication!
10.1371/journal.pcbi.1010752



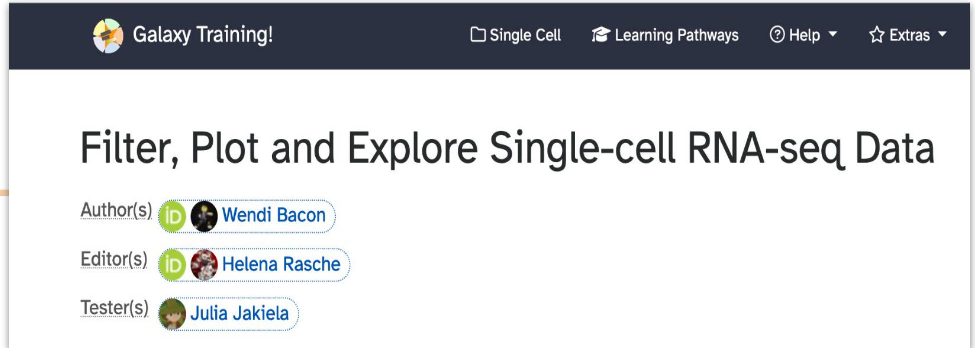
37	Topics
358	Tutorials
390	FAQs
309	Contributors
8.0	Years

A GTN Tutorial

A hands-on journal article

Metadata

- Questions
- Learning objectives
- Prerequisites
- Time estimate
- Supporting Materials

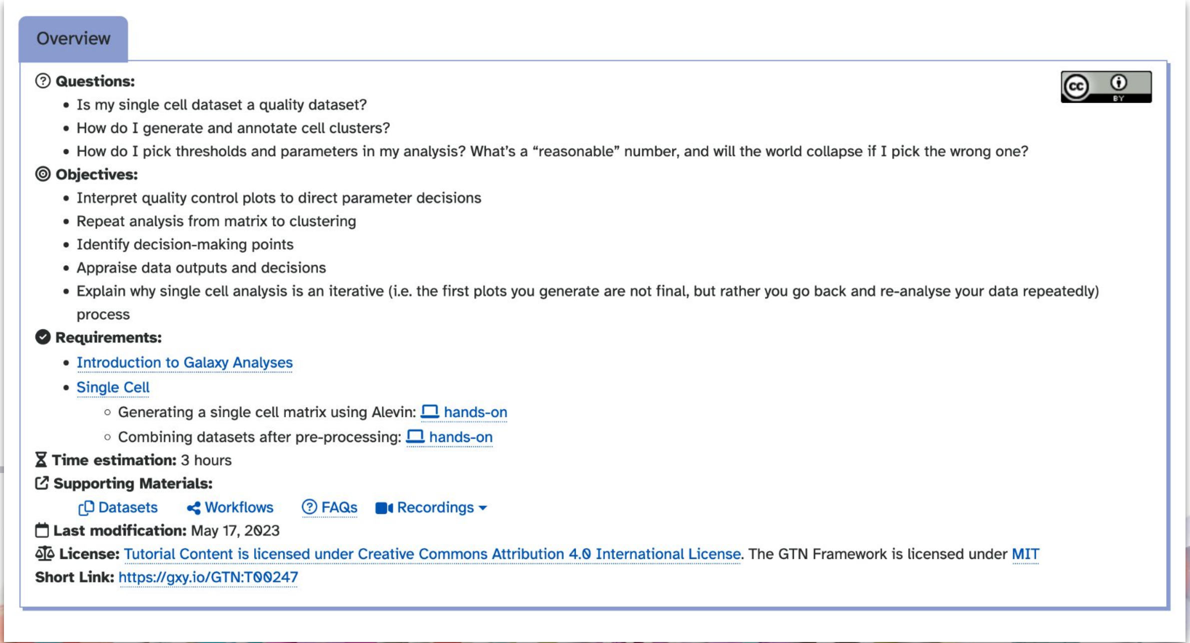


Filter, Plot and Explore Single-cell RNA-seq Data

Author(s) Wendi Bacon

Editor(s) Helena Rasche

Tester(s) Julia Jakiela



Questions:
- Is my single cell dataset a quality dataset?
- How do I generate and annotate cell clusters?
- How do I pick thresholds and parameters in my analysis? What's a “reasonable” number, and will the world collapse if I pick the wrong one?

Objectives:
- Interpret quality control plots to direct parameter decisions
- Repeat analysis from matrix to clustering
- Identify decision-making points
- Appraise data outputs and decisions
- Explain why single cell analysis is an iterative (i.e. the first plots you generate are not final, but rather you go back and re-analyse your data repeatedly) process

Requirements:
- Introduction to Galaxy Analyses
- Single Cell
  - Generating a single cell matrix using Alevin: hands-on
  - Combining datasets after pre-processing: hands-on

Time estimation: 3 hours

Supporting Materials:

Datasets Workflows FAQs Recordings

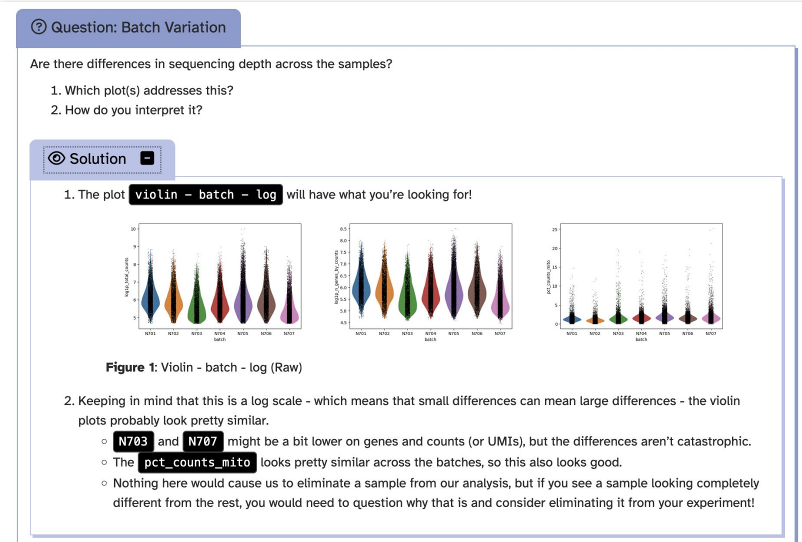
Last modification: May 17, 2023

License: Tutorial Content is licensed under Creative Commons Attribution 4.0 International License. The GTN Framework is licensed under MIT

Short Link: https://gxy.io/GTN:T00247

A GTN Tutorial

- Scientific Background
- Hands-on boxes
- Formative Assessments
  - Question & Answers



Question: Batch Variation

Are there differences in sequencing depth across the samples?

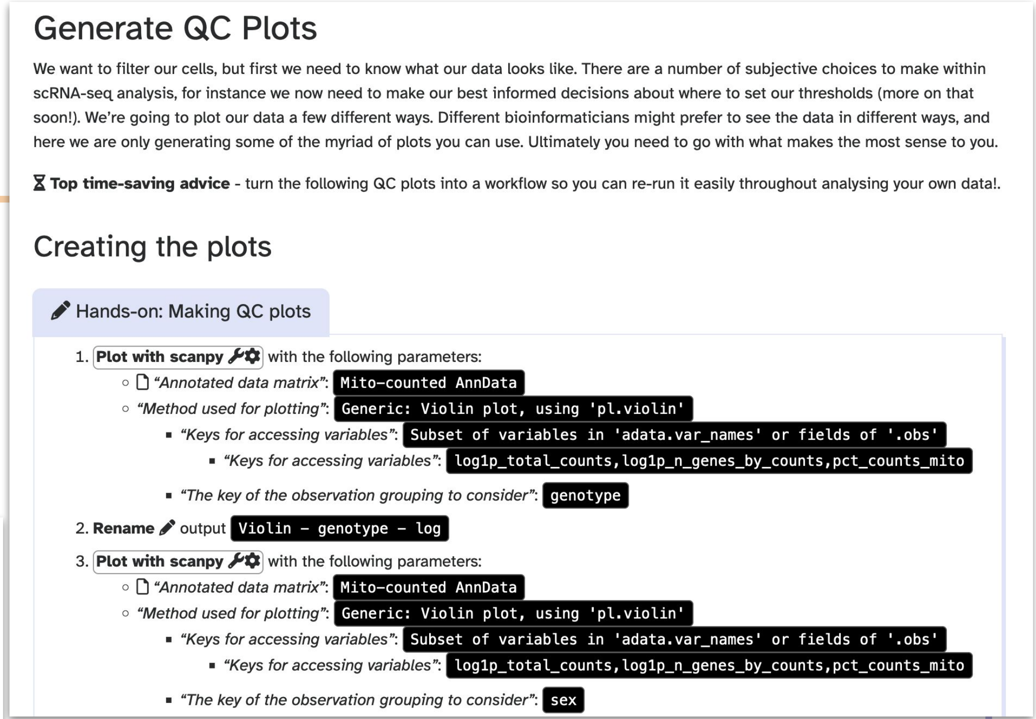
1. Which plot(s) addresses this?
2. How do you interpret it?

Solution

1. The plot `violin - batch - log` will have what you're looking for!

Figure 1: Violin - batch - log (Raw)

2. Keeping in mind that this is a log scale - which means that small differences can mean large differences - the violin plots probably look pretty similar.
   - `N703` and `N707` might be a bit lower on genes and counts (or UMIs), but the differences aren't catastrophic.
   - The `pct_counts_mito` looks pretty similar across the batches, so this also looks good.
   - Nothing here would cause us to eliminate a sample from our analysis, but if you see a sample looking completely different from the rest, you would need to question why that is and consider eliminating it from your experiment!



Generate QC Plots

We want to filter our cells, but first we need to know what our data looks like. There are a number of subjective choices to make within scRNA-seq analysis, for instance we now need to make our best informed decisions about where to set our thresholds (more on that soon!). We're going to plot our data a few different ways. Different bioinformaticians might prefer to see the data in different ways, and here we are only generating some of the myriad of plots you can use. Ultimately you need to go with what makes the most sense to you.

⌛ **Top time-saving advice** - turn the following QC plots into a workflow so you can re-run it easily throughout analysing your own data!.

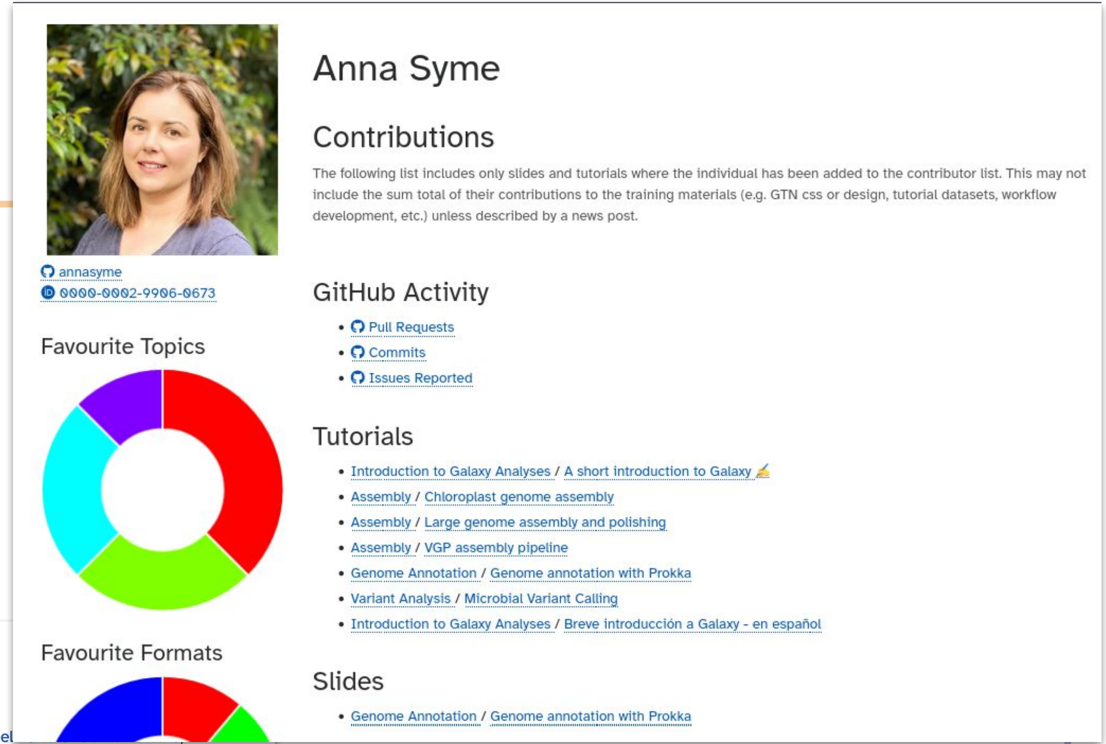
Creating the plots

Hands-on: Making QC plots

1. **Plot with scanpy** with the following parameters:
   - *"Annotated data matrix"*: `Mito-counted AnnData`
   - *"Method used for plotting"*: `Generic: Violin plot, using 'pl.violin'`
     - *"Keys for accessing variables"*: `Subset of variables in 'adata.var_names' or fields of '.obs'`
       - *"Keys for accessing variables"*: `log1p_total_counts,log1p_n_genes_by_counts,pct_counts_mito`
     - *"The key of the observation grouping to consider"*: `genotype`
2. **Rename** output `Violin - genotype - log`
3. **Plot with scanpy** with the following parameters:
   - *"Annotated data matrix"*: `Mito-counted AnnData`
   - *"Method used for plotting"*: `Generic: Violin plot, using 'pl.violin'`
     - *"Keys for accessing variables"*: `Subset of variables in 'adata.var_names' or fields of '.obs'`
       - *"Keys for accessing variables"*: `log1p_total_counts,log1p_n_genes_by_counts,pct_counts_mito`
     - *"The key of the observation grouping to consider"*: `sex`

A GTN Tutorial

- Getting Credit
  - Citable
  - Hall of Fame



Anna Syme

annasyme
0000-0002-9906-0673

Contributions

The following list includes only slides and tutorials where the individual has been added to the contributor list. This may not include the sum total of their contributions to the training materials (e.g. GTN css or design, tutorial datasets, workflow development, etc.) unless described by a news post.

GitHub Activity

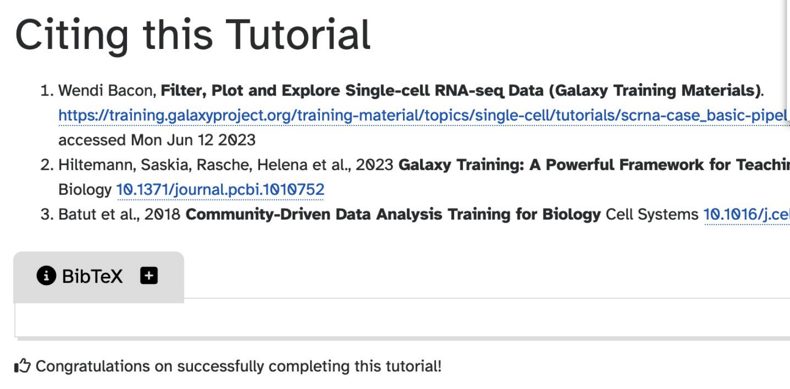
- Pull Requests
- Commits
- Issues Reported

Tutorials

- Introduction to Galaxy Analyses / A short introduction to Galaxy ✍
- Assembly / Chloroplast genome assembly
- Assembly / Large genome assembly and polishing
- Assembly / VGP assembly pipeline
- Genome Annotation / Genome annotation with Prokka
- Variant Analysis / Microbial Variant Calling
- Introduction to Galaxy Analyses / Breve introducción a Galaxy - en español

Slides

- Genome Annotation / Genome annotation with Prokka



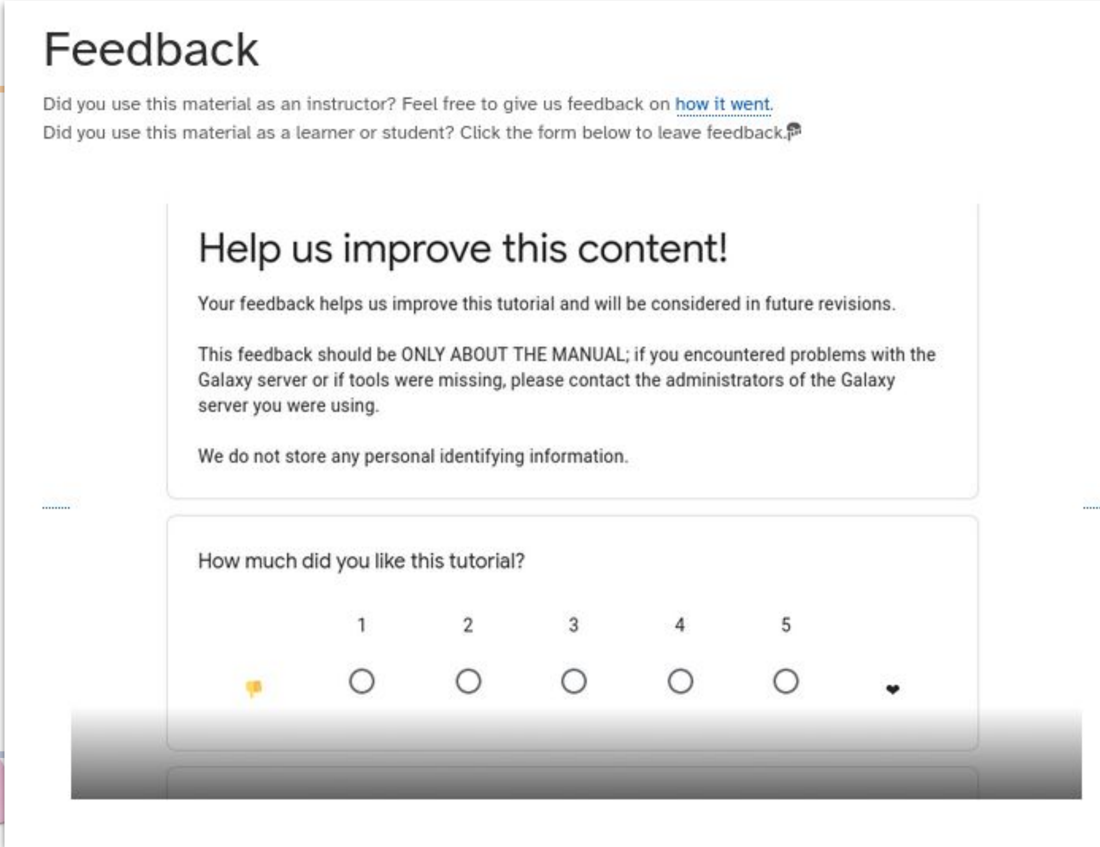
Citing this Tutorial

1. Wendi Bacon, **Filter, Plot and Explore Single-cell RNA-seq Data (Galaxy Training Materials).** https://training.galaxyproject.org/training-material/topics/single-cell/tutorials/scrna-case_basic-pipel accessed Mon Jun 12 2023
2. Hiltemann, Saskia, Rasche, Helena et al., 2023 **Galaxy Training: A Powerful Framework for Teaching!** PLOS Computational Biology 10.1371/journal.pcbi.1010752
3. Batut et al., 2018 **Community-Driven Data Analysis Training for Biology** Cell Systems 10.1016/j.cels.2018.05.012

Congratulations on successfully completing this tutorial!

Feedback

- Embedded Feedback Form at end of every tutorial
- Helps us improve materials



Feedback

Did you use this material as an instructor? Feel free to give us feedback on how it went.
Did you use this material as a learner or student? Click the form below to leave feedback.

Help us improve this content!

Your feedback helps us improve this tutorial and will be considered in future revisions.

This feedback should be ONLY ABOUT THE MANUAL; if you encountered problems with the Galaxy server or if tools were missing, please contact the administrators of the Galaxy server you were using.

We do not store any personal identifying information.

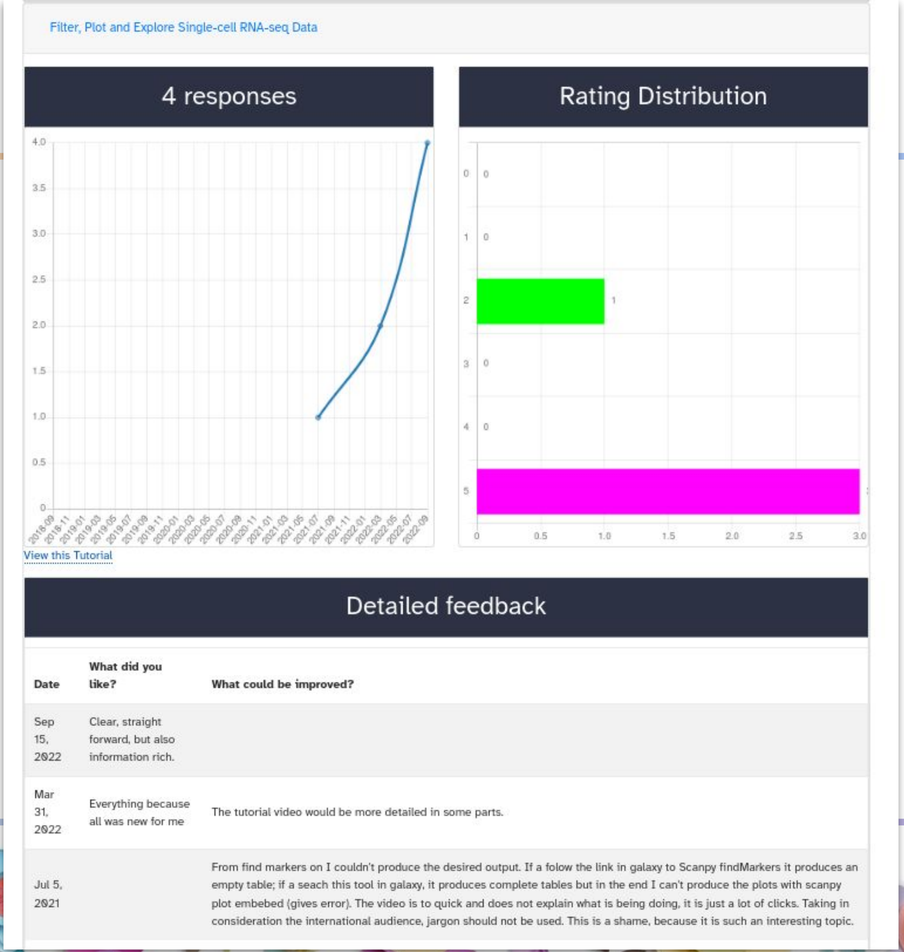
How much did you like this tutorial?

	1	2	3	4	5	
👎	○	○	○	○	○	❤

Feedback

- Embedded Feedback Form at end of every tutorial
- Helps us improve materials
- See [feedback results on the GTN](training.galaxyproject.org/feedback) (public but anonymous)

training.galaxyproject.org/feedback



Detailed feedback

Date	What did you like?	What could be improved?
Sep 15, 2022	Clear, straight forward, but also information rich.	
Mar 31, 2022	Everything because all was new for me	The tutorial video would be more detailed in some parts.
Jul 5, 2021		From find markers on I couldn't produce the desired output. If a folow the link in galaxy to Scanpy findMarkers it produces an empty table; if a seach this tool in galaxy, it produces complete tables but in the end I can't produce the plots with scanpy plot embebed (gives error). The video is to quick and does not explain what is being doing, it is just a lot of clicks. Taking in consideration the international audience, jargon should not be used. This is a shame, because it is such an interesting topic.

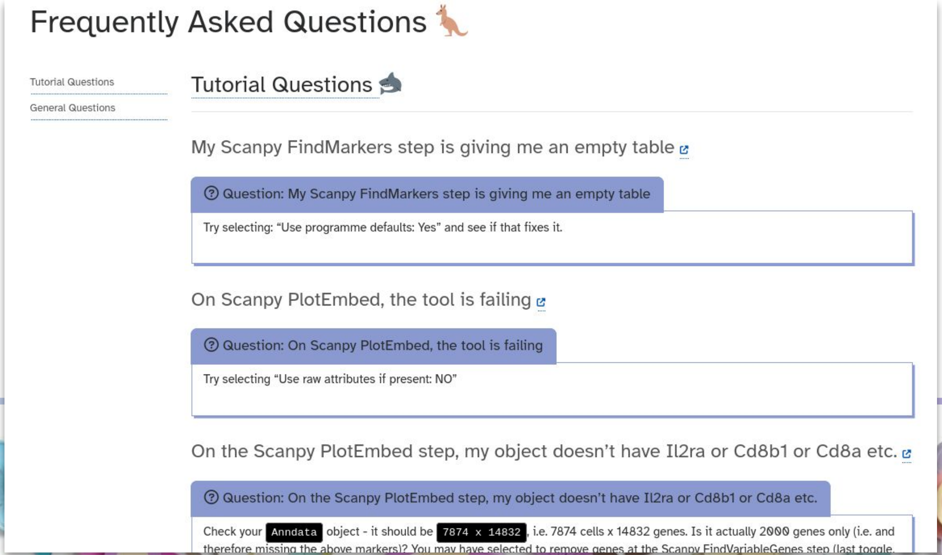
Supporting Materials & FAQs

- Slides
- Datasets
- Workflows
- Videos
- FAQs
    - (also great to prepare for teaching!)

Time estimation: 3 hours

Supporting Materials:

Last modification: Jul 14, 2023



Frequently Asked Questions

Tutorial Questions

My Scanpy FindMarkers step is giving me an empty table

Question: My Scanpy FindMarkers step is giving me an empty table

Try selecting: “Use programme defaults: Yes” and see if that fixes it.

On Scanpy PlotEmbed, the tool is failing

Question: On Scanpy PlotEmbed, the tool is failing

Try selecting “Use raw attributes if present: NO”

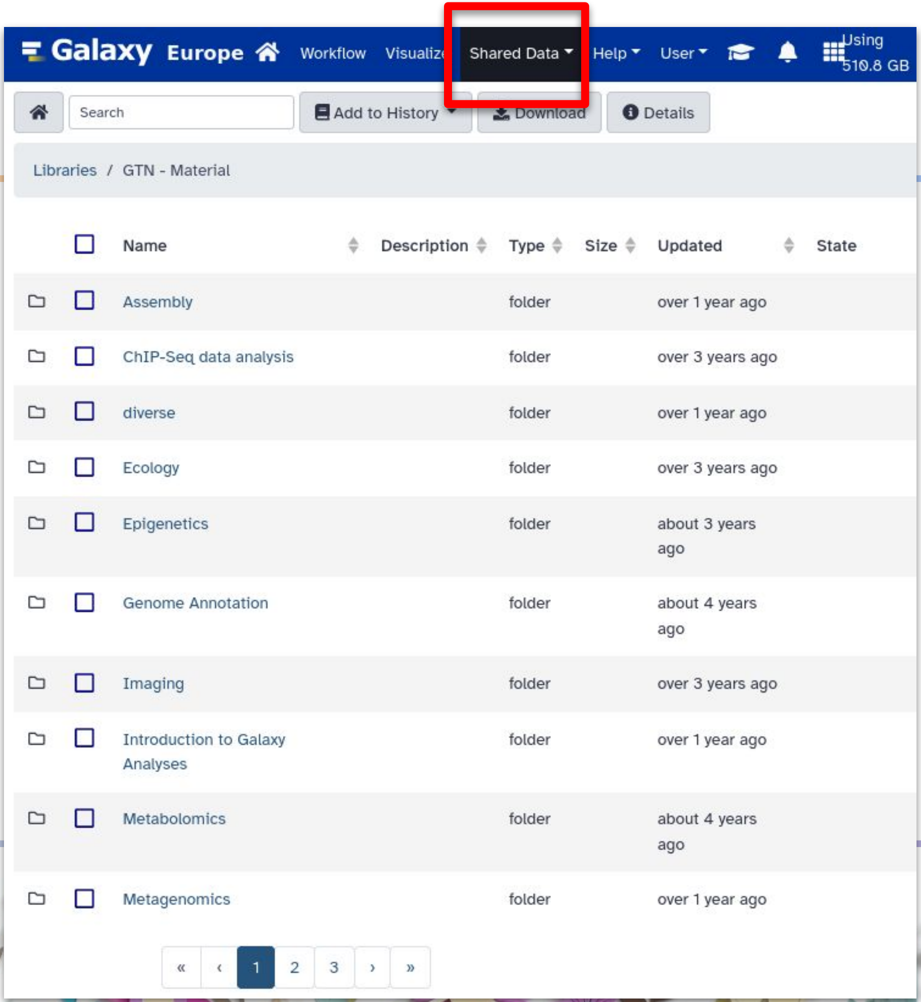
On the Scanpy PlotEmbed step, my object doesn’t have Il2ra or Cd8b1 or Cd8a etc.

Question: On the Scanpy PlotEmbed step, my object doesn’t have Il2ra or Cd8b1 or Cd8a etc.

Check your `Anndata` object - it should be `7874 x 14832`, i.e. 7874 cells x 14832 genes. Is it actually 2000 genes only (i.e. and therefore missing the above markers)? You may have selected to remove genes at the Scanpy FindVariableGenes step (last toggle.

Input datasets

- Deposited into Zenodo
- Automatically uploaded to a **Shared Data Library**
- Students import directly into their history
- No upload required!



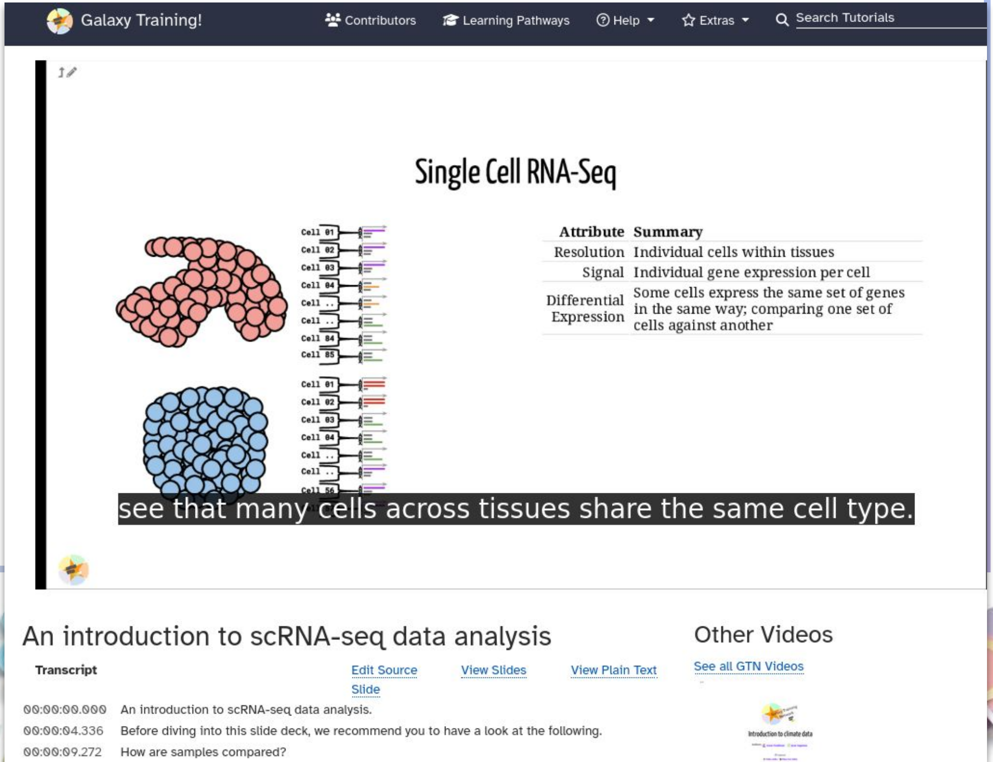
Libraries / GTN - Material

☐	Name	Description	Type	Size	Updated	State
☐	Assembly		folder		over 1 year ago	
☐	ChIP-Seq data analysis		folder		over 3 years ago	
☐	diverse		folder		over 1 year ago	
☐	Ecology		folder		over 3 years ago	
☐	Epigenetics		folder		about 3 years ago	
☐	Genome Annotation		folder		about 4 years ago	
☐	Imaging		folder		over 3 years ago	
☐	Introduction to Galaxy Analyses		folder		over 1 year ago	
☐	Metabolomics		folder		about 4 years ago	
☐	Metagenomics		folder		over 1 year ago	

Automated Video Slides

- Based on speaker notes from other instructors
- Text-to-Speech

- Easy to prepare for teaching!
- Easy to keep up-to-date!



Attribute	Summary
Resolution	Individual cells within tissues
Signal	Individual gene expression per cell
Differential Expression	Some cells express the same set of genes in the same way; comparing one set of cells against another

see that many cells across tissues share the same cell type.

An introduction to scRNA-seq data analysis

Transcript

Edit Source Slide
View Slides
View Plain Text

00:00:00.000 An introduction to scRNA-seq data analysis.
00:00:04.336 Before diving into this slide deck, we recommend you to have a look at the following.
00:00:09.272 How are samples compared?

Other Videos

See all GTN Videos

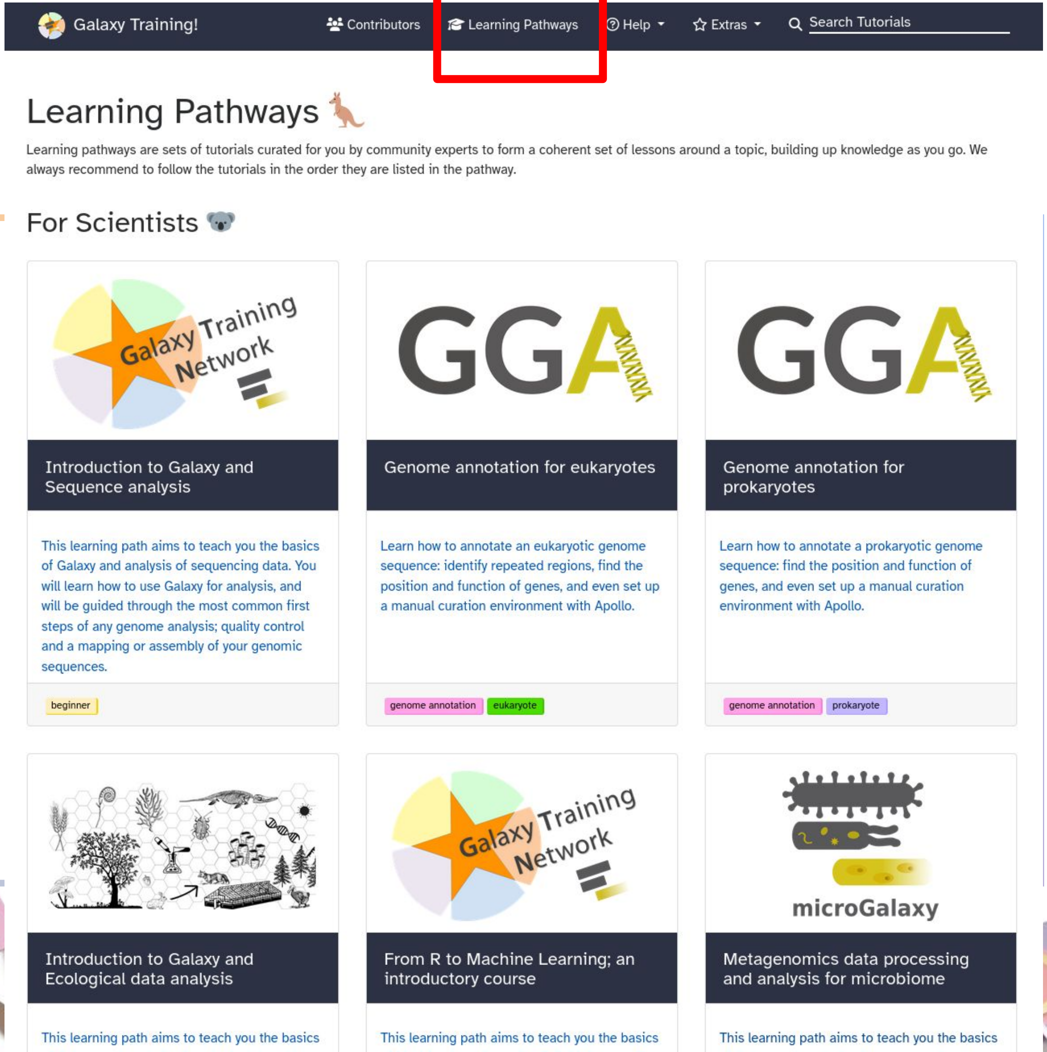
Learning Pathways

"From Zero to Hero"

A journey around a topic, taking you from the introductory tutorials on to the more advanced lessons.

E.g. a curriculum for a week-long training course

New Learning Pathways Welcome!
(ask us how to add yours!)



Learning Pathways 🦘

Learning pathways are sets of tutorials curated for you by community experts to form a coherent set of lessons around a topic, building up knowledge as you go. We always recommend to follow the tutorials in the order they are listed in the pathway.

For Scientists 🐨

Introduction to Galaxy and Sequence analysis

This learning path aims to teach you the basics of Galaxy and analysis of sequencing data. You will learn how to use Galaxy for analysis, and will be guided through the most common first steps of any genome analysis; quality control and a mapping or assembly of your genomic sequences.

beginner

Genome annotation for eukaryotes

Learn how to annotate an eukaryotic genome sequence: identify repeated regions, find the position and function of genes, and even set up a manual curation environment with Apollo.

genome annotation | eukaryote

Genome annotation for prokaryotes

Learn how to annotate a prokaryotic genome sequence: find the position and function of genes, and even set up a manual curation environment with Apollo.

genome annotation | prokaryote

Introduction to Galaxy and Ecological data analysis

This learning path aims to teach you the basics

From R to Machine Learning; an introductory course

This learning path aims to teach you the basics

Metagenomics data processing and analysis for microbiome

This learning path aims to teach you the basics

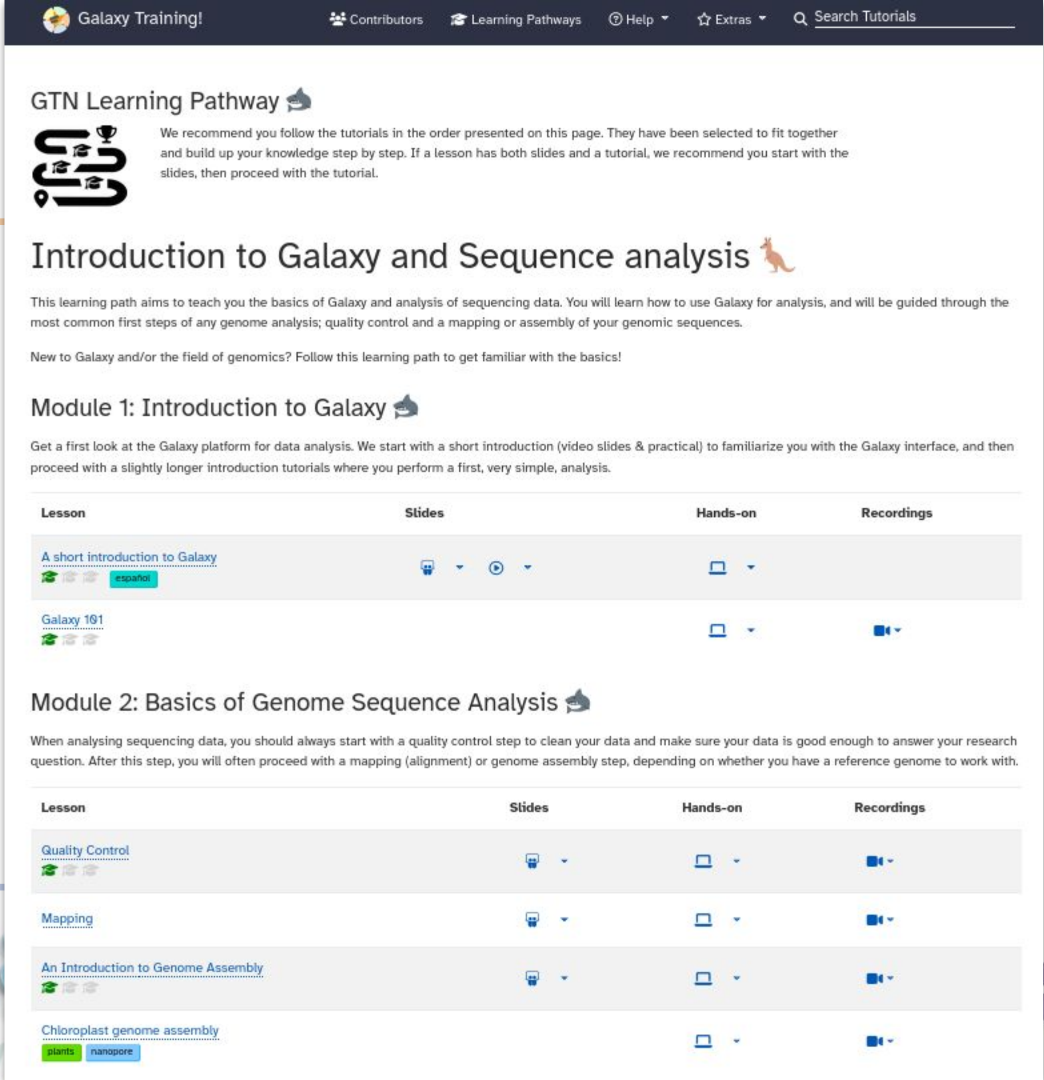
Learning Pathways

“From Zero to Hero”

A journey around a topic, taking you from the introductory tutorials on to the more advanced lessons.

E.g. a curriculum for a week-long training course

New Learning Pathways Welcome!
(ask us how to add yours!)



GTN Learning Pathway

We recommend you follow the tutorials in the order presented on this page. They have been selected to fit together and build up your knowledge step by step. If a lesson has both slides and a tutorial, we recommend you start with the slides, then proceed with the tutorial.

Introduction to Galaxy and Sequence analysis

This learning path aims to teach you the basics of Galaxy and analysis of sequencing data. You will learn how to use Galaxy for analysis, and will be guided through the most common first steps of any genome analysis; quality control and a mapping or assembly of your genomic sequences.

New to Galaxy and/or the field of genomics? Follow this learning path to get familiar with the basics!

Module 1: Introduction to Galaxy

Get a first look at the Galaxy platform for data analysis. We start with a short introduction (video slides & practical) to familiarize you with the Galaxy interface, and then proceed with a slightly longer introduction tutorials where you perform a first, very simple, analysis.

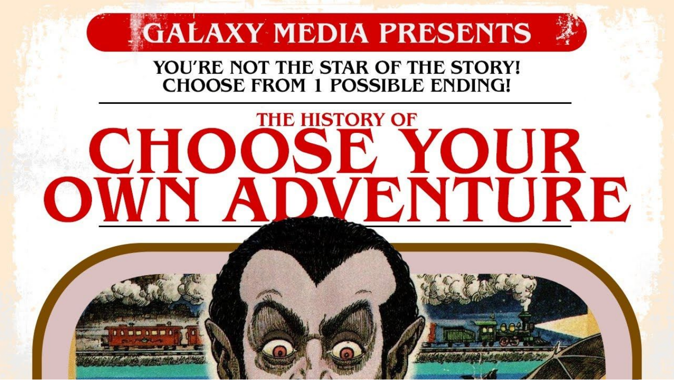
Lesson	Slides	Hands-on	Recordings
A short introduction to Galaxy español	● ●	●	
Galaxy 101		●	●

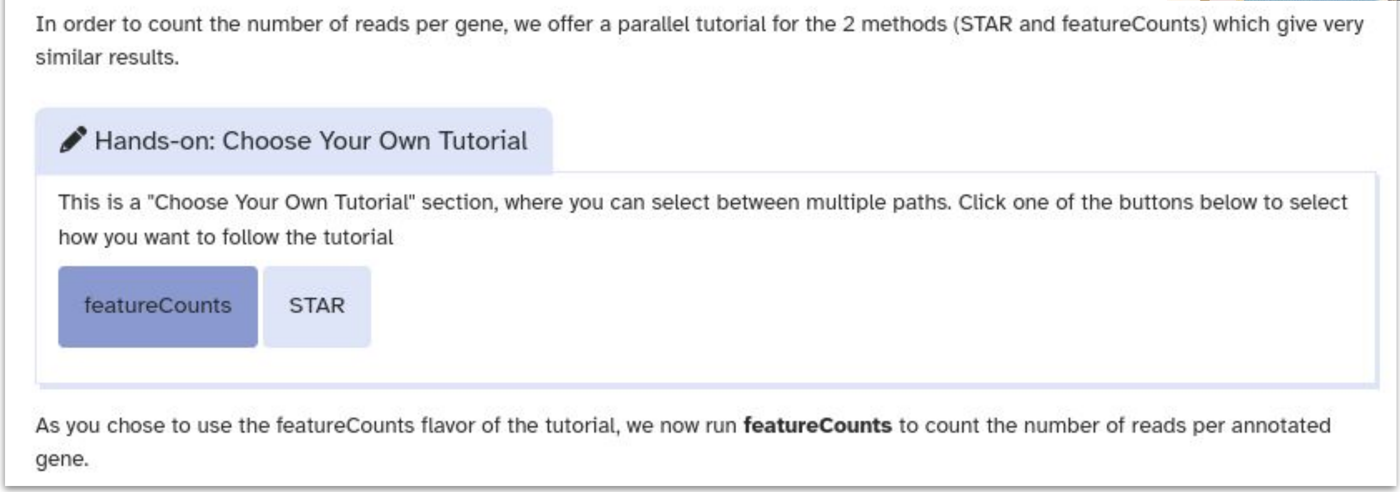
Module 2: Basics of Genome Sequence Analysis

When analysing sequencing data, you should always start with a quality control step to clean your data and make sure your data is good enough to answer your research question. After this step, you will often proceed with a mapping (alignment) or genome assembly step, depending on whether you have a reference genome to work with.

Lesson	Slides	Hands-on	Recordings
Quality Control	●	●	●
Mapping	●	●	●
An Introduction to Genome Assembly	●	●	●
Chloroplast genome assembly plants nanopore		●	●

CYOA: Choose Your Own Adventure





Counting reads per genes

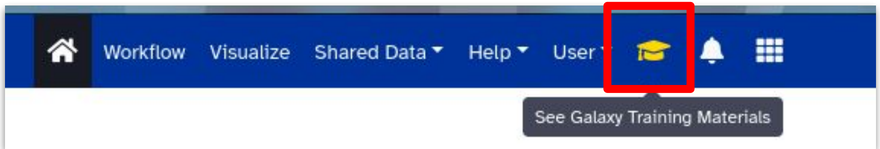
In order to count the number of reads per gene, we offer a parallel tutorial for the 2 methods (STAR and featureCounts) which give very similar results.

Hands-on: Choose Your Own Tutorial

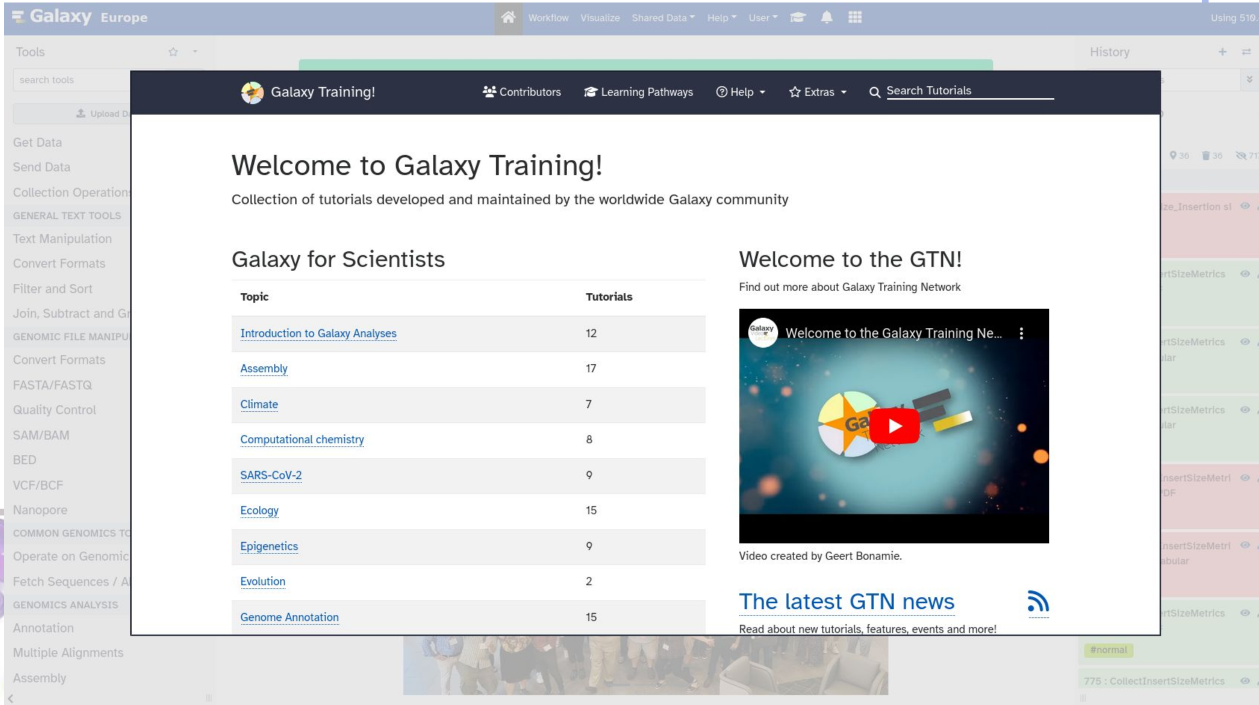
This is a "Choose Your Own Tutorial" section, where you can select between multiple paths. Click one of the buttons below to select how you want to follow the tutorial

As you chose to use the featureCounts flavor of the tutorial, we now run **featureCounts** to count the number of reads per annotated gene.

Tutorial Mode



- Open Galaxy Tutorial Mode



Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

Galaxy for Scientists

Topic	Tutorials
Introduction to Galaxy Analyses	12
Assembly	17
Climate	7
Computational chemistry	8
SARS-CoV-2	9
Ecology	15
Epigenetics	9
Evolution	2
Genome Annotation	15

Welcome to the GTN!

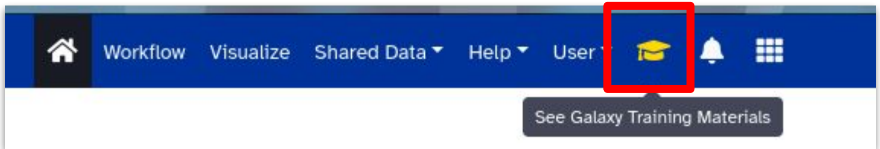
Find out more about Galaxy Training Network

Video created by Geert Bonamie.

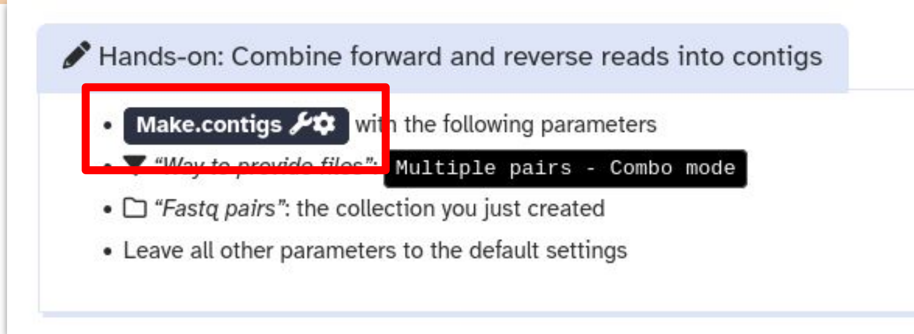
The latest GTN news

Read about new tutorials, features, events and more!

Tutorial Mode

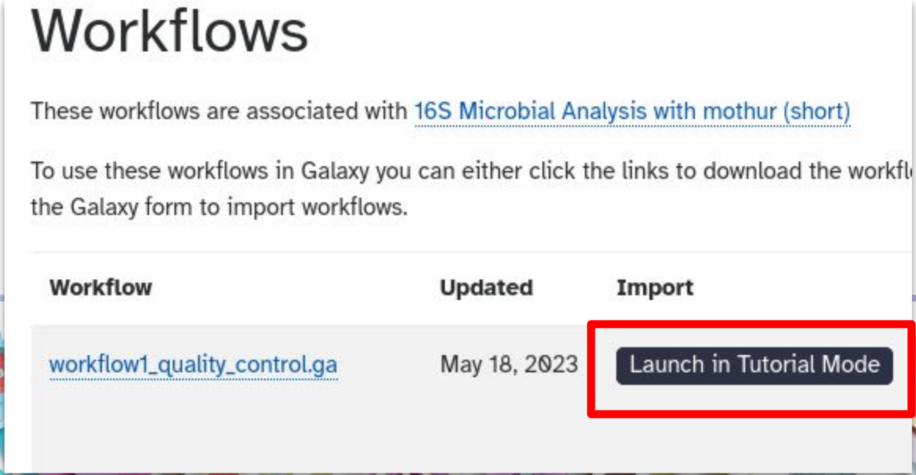


- Open Galaxy Tutorial Mode
- Click on **tool names** in tutorials to open them directly in your Galaxy
- Click on **Workflows** in tutorials to import them directly in your Galaxy



Hands-on: Combine forward and reverse reads into contigs

- **Make.contigs** with the following parameters
  - *“Way to provide files”*: `Multiple pairs - Combo mode`
  - *“Fastq pairs”*: the collection you just created
- Leave all other parameters to the default settings



Workflows

These workflows are associated with 16S Microbial Analysis with mothur (short)

To use these workflows in Galaxy you can either click the links to download the workfl the Galaxy form to import workflows.

Workflow	Updated	Import
workflow1_quality_control.ga	May 18, 2023	Launch in Tutorial Mode

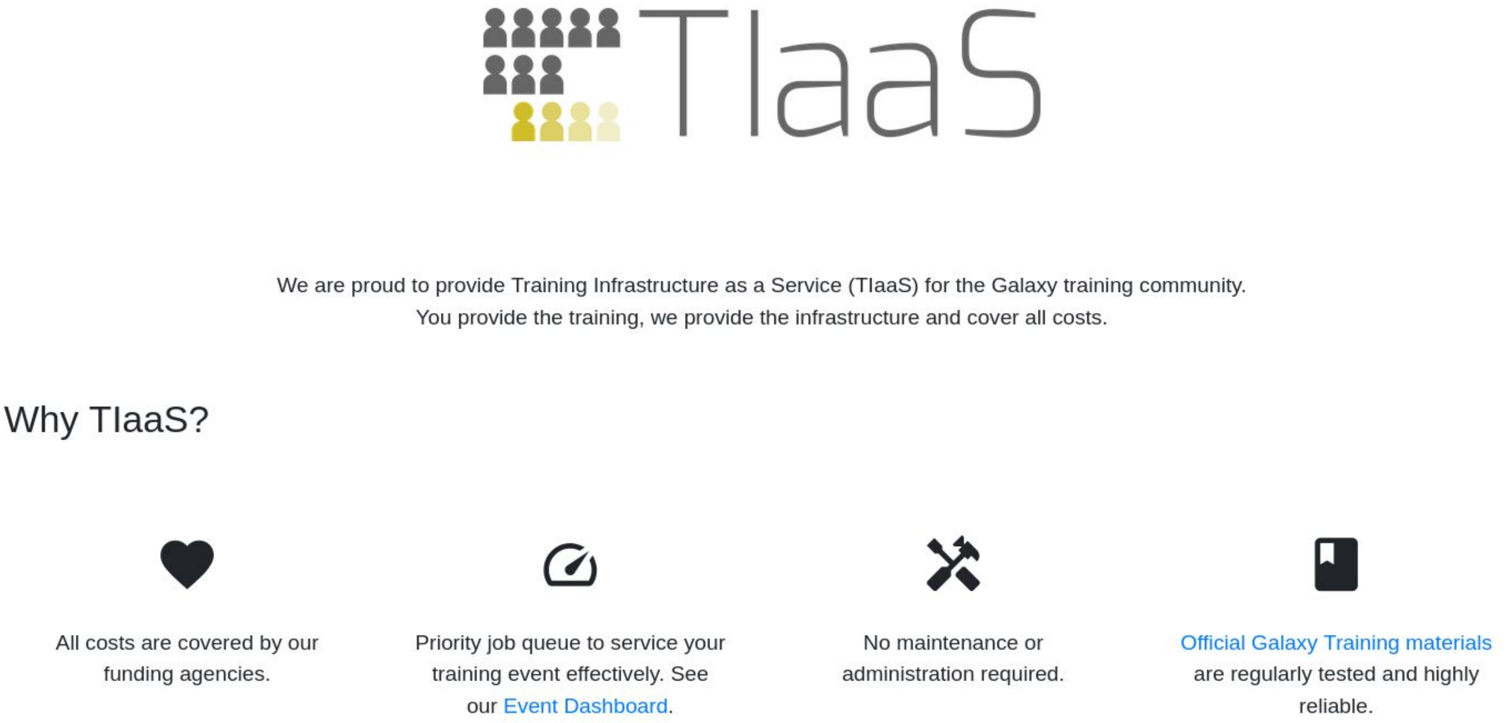
Training Infrastructure

TIaaS: Training Infrastructure as a Service

Challenge: Training Infrastructure is not trivial to set up

- Finding compute resources
- Setting up the environment for running your training
- Testing the tutorials on this environment
- Finding somebody who can do this (usually not the course organizers/teachers)
- **You just want to worry about the course content, not the infrastructure!**

Solution: Training Infrastructure as a Service



We are proud to provide Training Infrastructure as a Service (TIaaS) for the Galaxy training community.
You provide the training, we provide the infrastructure and cover all costs.

Why TIaaS?

All costs are covered by our funding agencies.

Priority job queue to service your training event effectively. See our Event Dashboard.

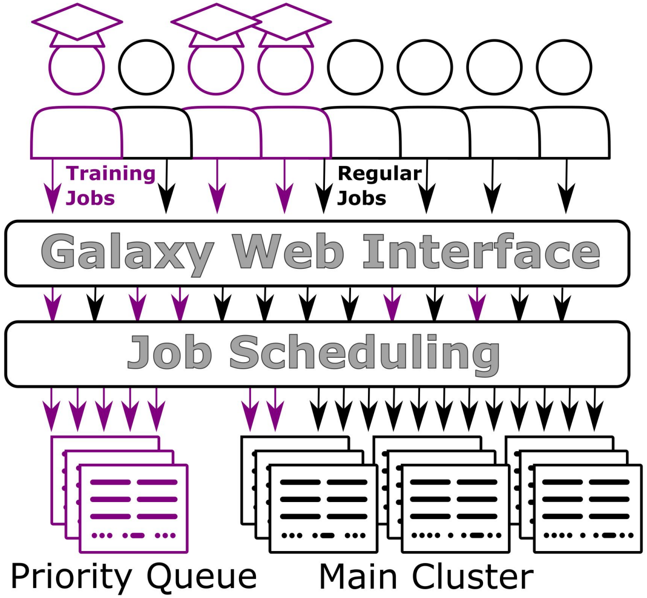
No maintenance or administration required.

Official Galaxy Training materials are regularly tested and highly reliable.

TIaaS: How it works

- Dedicated resources for your course
- Dedicated job queue
- If Galaxy is busy, your workshop doesn't notice

TIaaS

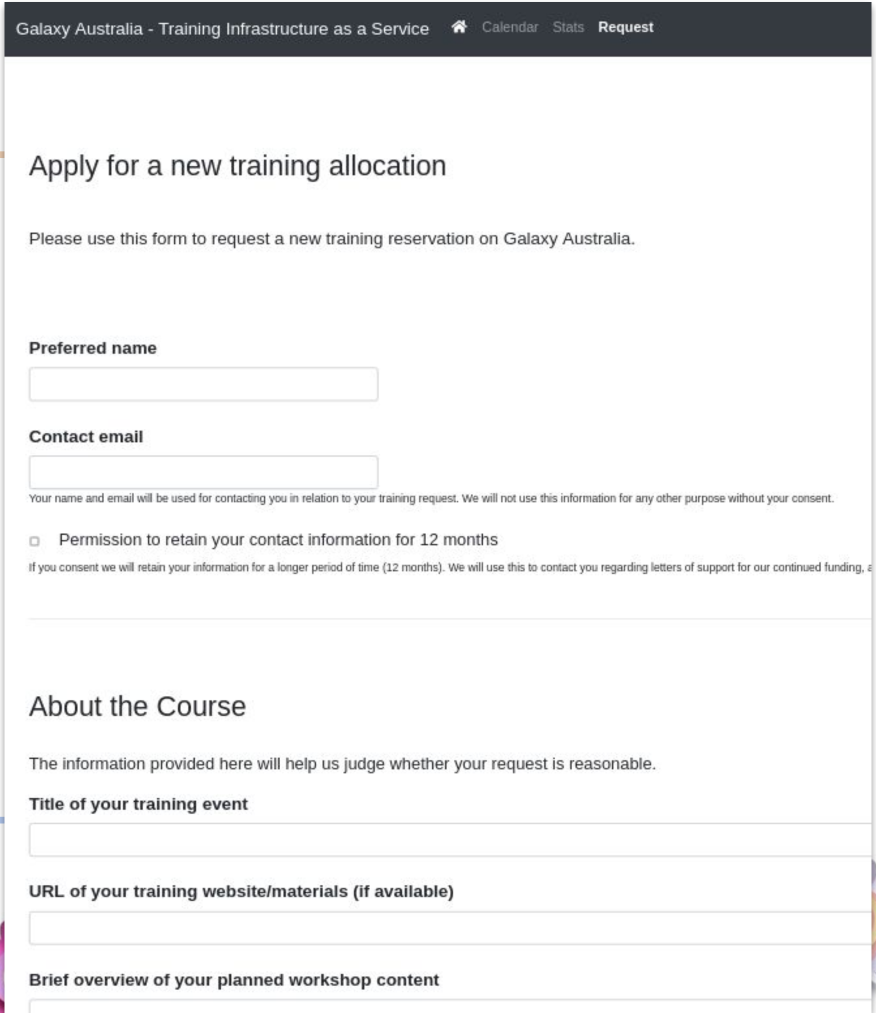


Educators: Requesting TIaaS

- **It's FREE!**
- Available on Galaxy AU, EU, US

- Request Form
  - Dates
  - Number of participants
  - Which (GTN) tutorials you will use?

[https://usegalaxy.org.au/**tiaas**](https://usegalaxy.org.au/tiaas)



Apply for a new training allocation

Please use this form to request a new training reservation on Galaxy Australia.

Preferred name

Contact email

Your name and email will be used for contacting you in relation to your training request. We will not use this information for any other purpose without your consent.

☐ Permission to retain your contact information for 12 months

If you consent we will retain your information for a longer period of time (12 months). We will use this to contact you regarding letters of support for our continued funding, a

About the Course

The information provided here will help us judge whether your request is reasonable.

Title of your training event

URL of your training website/materials (if available)

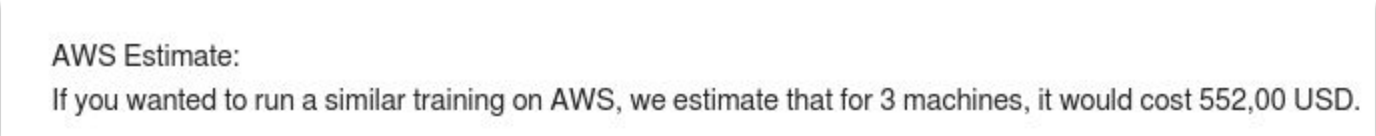
Brief overview of your planned workshop content

Galaxy Admins: Approval Process

- Will evaluate your request, possibly ask for more info
- Assess how much compute resources are needed for your training

- You will get a special join link for your training to give to learners

usegalaxy.org.au/join-training/bioc2023/

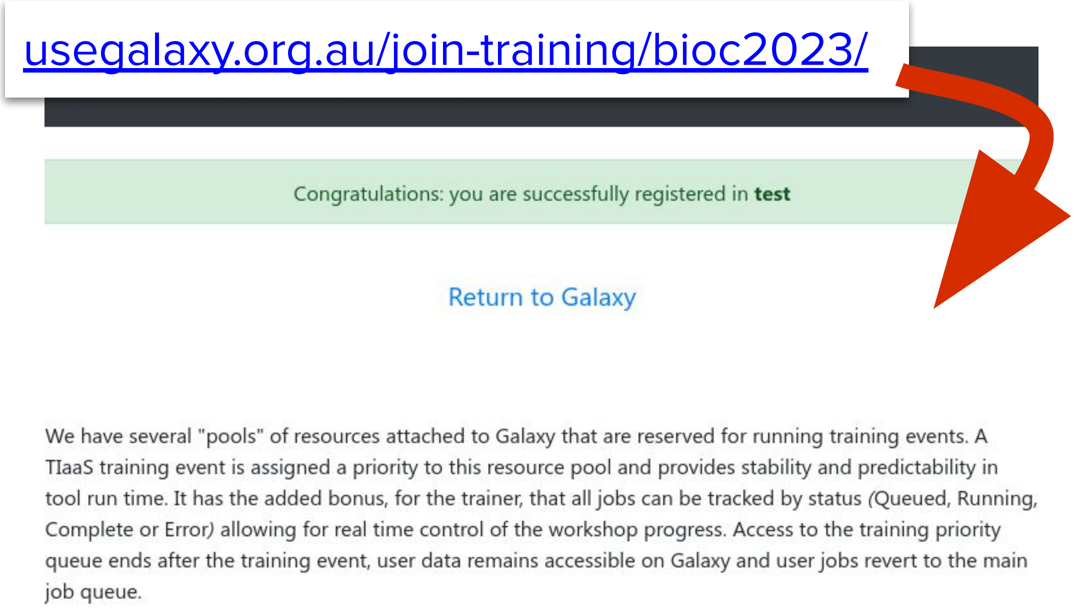
- 

AWS Estimate:

If you wanted to run a similar training on AWS, we estimate that for 3 machines, it would cost 552,00 USD.

Learners: Using TIaaS

usegalaxy.org.au/join-training/bioc2023/



Congratulations: you are successfully registered in **test**

We have several "pools" of resources attached to Galaxy that are reserved for running training events. A TIaaS training event is assigned a priority to this resource pool and provides stability and predictability in tool run time. It has the added bonus, for the trainer, that all jobs can be tracked by status (Queued, Running, Complete or Error) allowing for real time control of the workshop progress. Access to the training priority queue ends after the training event, user data remains accessible on Galaxy and user jobs revert to the main job queue.

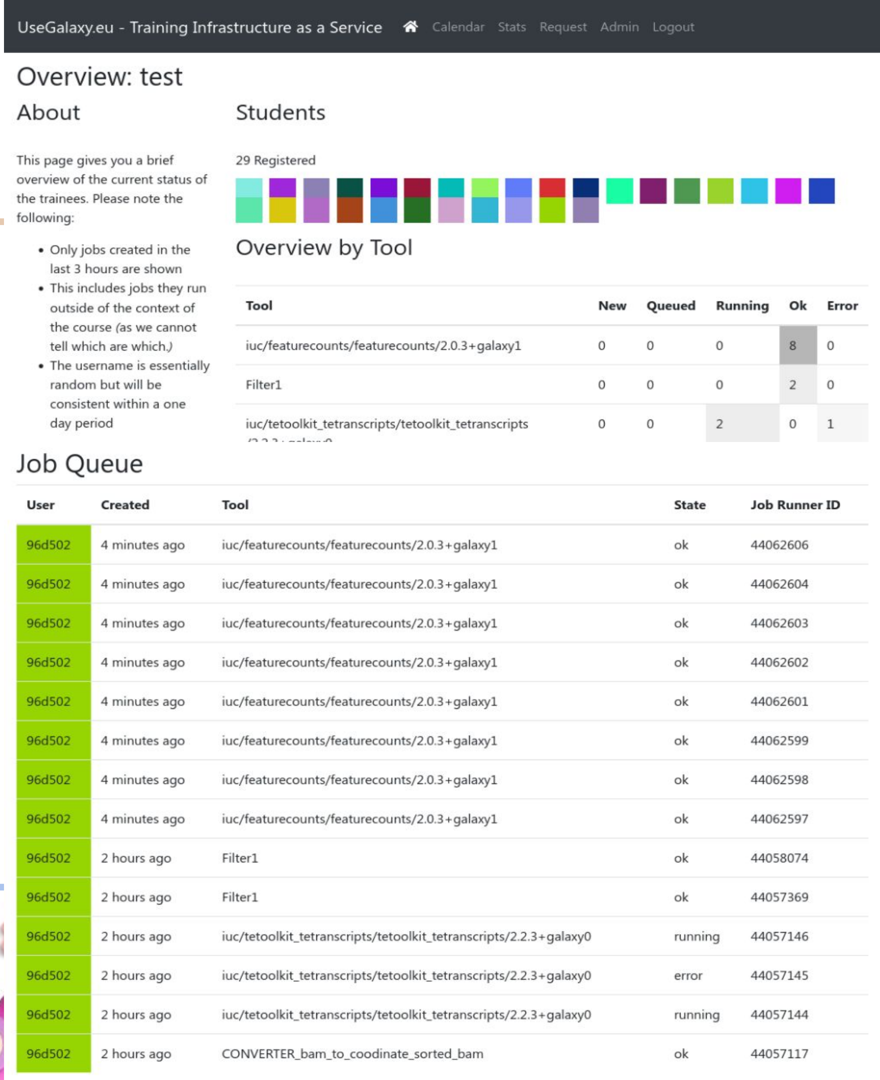
1. Teacher shares link to students
2. Participants click the link
3. …
4. Done! Everything else is auto*magic*

Educators: TIaaS Dashboard

Great for courses:

- Track student progress
- Easily spot problems
- Aggregated & Anonymized

Makes remote training not only possible but easy



Overview: test

About

This page gives you a brief overview of the current status of the trainees. Please note the following:

- Only jobs created in the last 3 hours are shown
- This includes jobs they run outside of the context of the course (as we cannot tell which are which.)
- The username is essentially random but will be consistent within a one day period

Students

29 Registered

Overview by Tool

Tool	New	Queued	Running	Ok	Error
iuc/featurecounts/featurecounts/2.0.3+galaxy1	0	0	0	8	0
Filter1	0	0	0	2	0
iuc/tetoolkit_tetranscripts/tetoolkit_tetranscripts	0	0	2	0	1

Job Queue

User	Created	Tool	State	Job Runner ID
96d502	4 minutes ago	iuc/featurecounts/featurecounts/2.0.3+galaxy1	ok	44062606
96d502	4 minutes ago	iuc/featurecounts/featurecounts/2.0.3+galaxy1	ok	44062604
96d502	4 minutes ago	iuc/featurecounts/featurecounts/2.0.3+galaxy1	ok	44062603
96d502	4 minutes ago	iuc/featurecounts/featurecounts/2.0.3+galaxy1	ok	44062602
96d502	4 minutes ago	iuc/featurecounts/featurecounts/2.0.3+galaxy1	ok	44062601
96d502	4 minutes ago	iuc/featurecounts/featurecounts/2.0.3+galaxy1	ok	44062599
96d502	4 minutes ago	iuc/featurecounts/featurecounts/2.0.3+galaxy1	ok	44062598
96d502	4 minutes ago	iuc/featurecounts/featurecounts/2.0.3+galaxy1	ok	44062597
96d502	2 hours ago	Filter1	ok	44058074
96d502	2 hours ago	Filter1	ok	44057369
96d502	2 hours ago	iuc/tetoolkit_tetranscripts/tetoolkit_tetranscripts/2.2.3+galaxy0	running	44057146
96d502	2 hours ago	iuc/tetoolkit_tetranscripts/tetoolkit_tetranscripts/2.2.3+galaxy0	error	44057145
96d502	2 hours ago	iuc/tetoolkit_tetranscripts/tetoolkit_tetranscripts/2.2.3+galaxy0	running	44057144
96d502	2 hours ago	CONVERTER_bam_to_coodinate_sorted_bam	ok	44057117

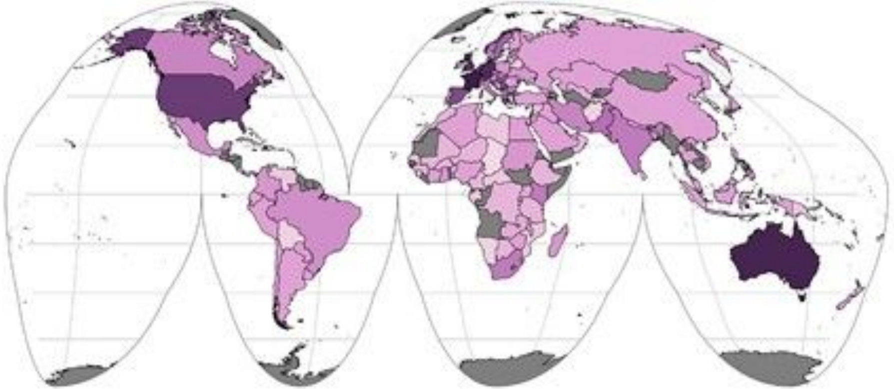
10.1093/gigascience/giad048

Powered by Training Infrastructure as a Service (TIaaS)

508 events

143 countries

24 k Learners



JOURNAL ARTICLE

Training Infrastructure as a Service

Helena Rasche ✉, Cameron Hyde, John Davis, Simon Gladman, Nate Coraor, Anthony Bretaudeau, Gianmauro Cuccuru, Wendi Bacon, Beatriz Serrano-Solano, Jennifer Hillman-Jackson, Saskia Hiltemann, Miaomiao Zhou, Björn Grüning, Andrew Stubbs

Author Notes

GigaScience, Volume 12, 2023, giad048, https://doi.org/10.1093/gigascience/giad048

Published: 03 July 2023 **Article history** ▾

Abstract

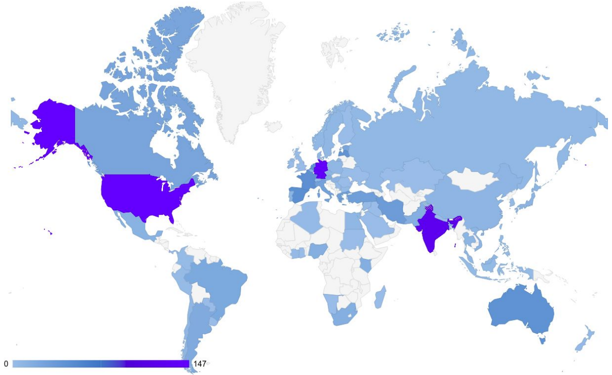
Background

Hands-on training, whether in bioinformatics or other domains, often requires significant technical resources and knowledge to set up and run. Instructors must have access to powerful compute infrastructure that can support

Remote Training

Smörgåsbord Events

- Started in the pandemic (2021, 2022, 2023 ..2024?)
- Completely online training event
- Asynchronous (self-paced & video based)
- Many topics and tutorials, YOU decide your program
- Global
- 24/7 support from the community on Slack
- **Free**!



May 2023: 3rd Annual Smorgasbord

- 3000+ registrations
- 24 modules, 191 video tutorials, 116 hours
- 79 speakers, 92 contributors
- 141 instructors on Slack

NEW collaborations: **BioConductor, RO-crates**

Training Modules

Below you can find some recommended training modules for various topics

Learning Galaxy

Intro to Galaxy
Intro to Galaxy (Non Genomics)
Advanced Galaxy Features

Browse by Analysis Type

I'm interested in...

Genomics
Transcriptomics
Proteomics
Single-Cell Analysis
Meta -omics
Genome Annotation
Genome Assembly
Machine Learning

Browse by Sample Type

I'm interested in...

One Health Data analysis
SARS-CoV-2 Data analysis
Plant Analysis
Microbial Analysis
Mycobacterium Tuberculosis

Non-Genomics modules

Ecology
Climate
Computational Chemistry
Admin Training
Developer Training

New Community, New Researchers

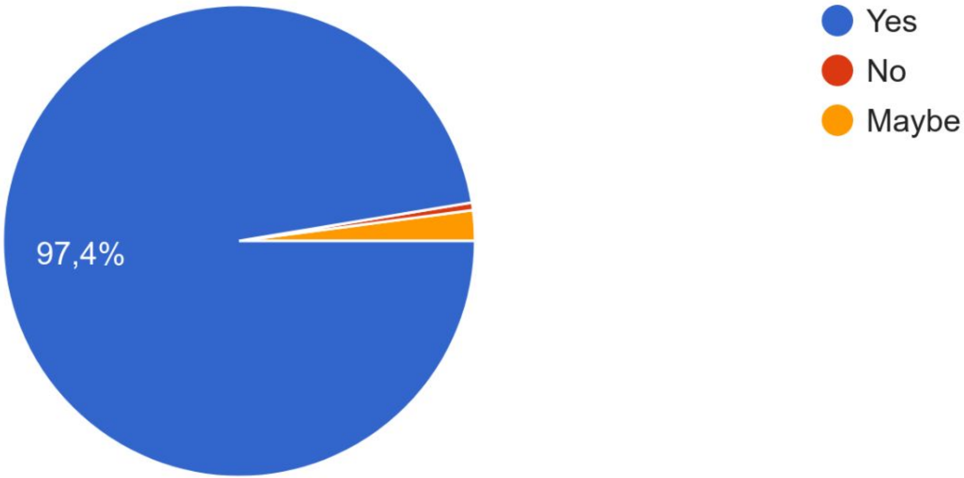
Was this your first Galaxy training?

193 svar

They Love It

Would you recommend this course to others?

192 svar



Virtual Training Reaches the World

This would not have happened with in-person trainings. It simply isn't feasible.

kg/CO_2

0*

Countries

131

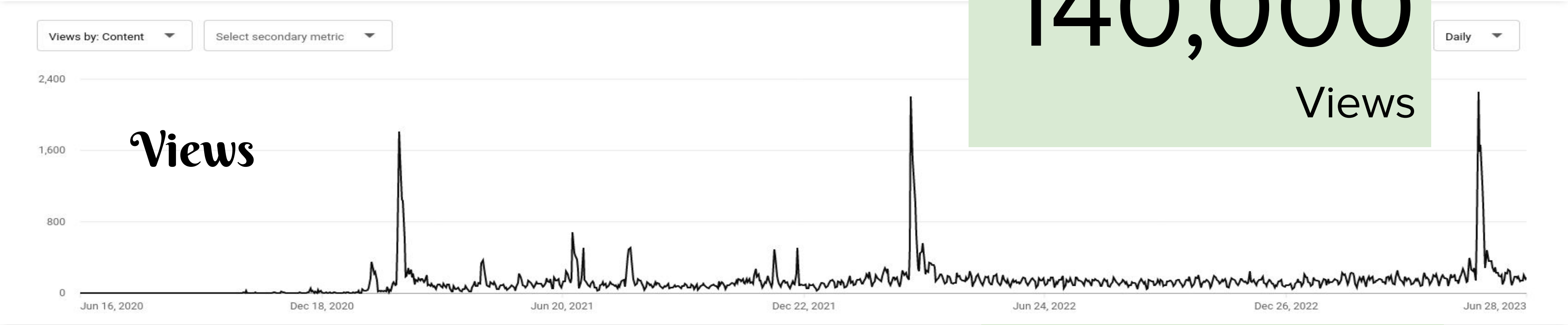
* approximately, computers, internet networking costs not counted.

The Smörgåsbord Effect

140,000

Views

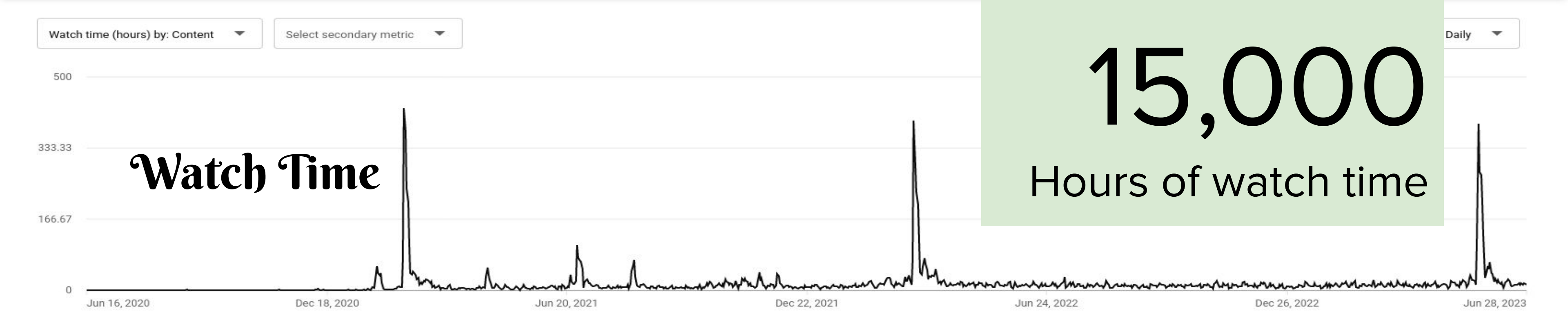
Views



15,000

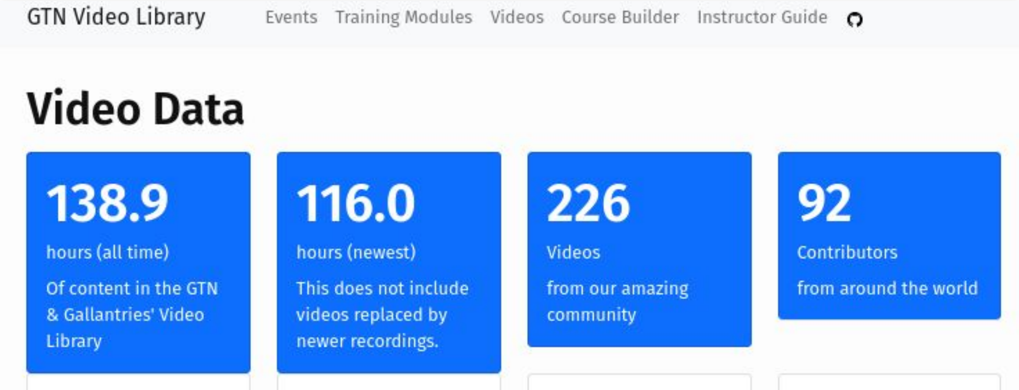
Hours of watch time

Watch Time



GTN Video Library

- All videos recorded for the 3 Smörgåsbords are available in the GTN video library
- Integrated with the GTN
- You can reuse them for your training events!



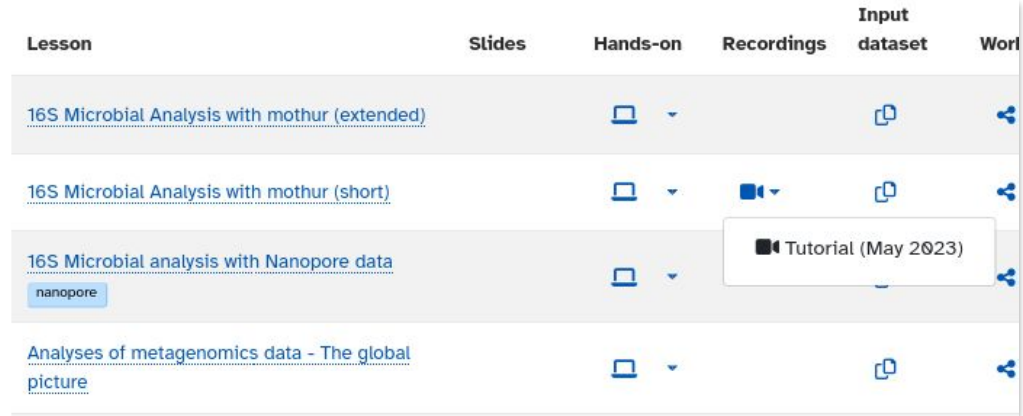
Video Data

138.9
hours (all time)
Of content in the GTN & Gallantries' Video Library

116.0
hours (newest)
This does not include videos replaced by newer recordings.

226
Videos
from our amazing community

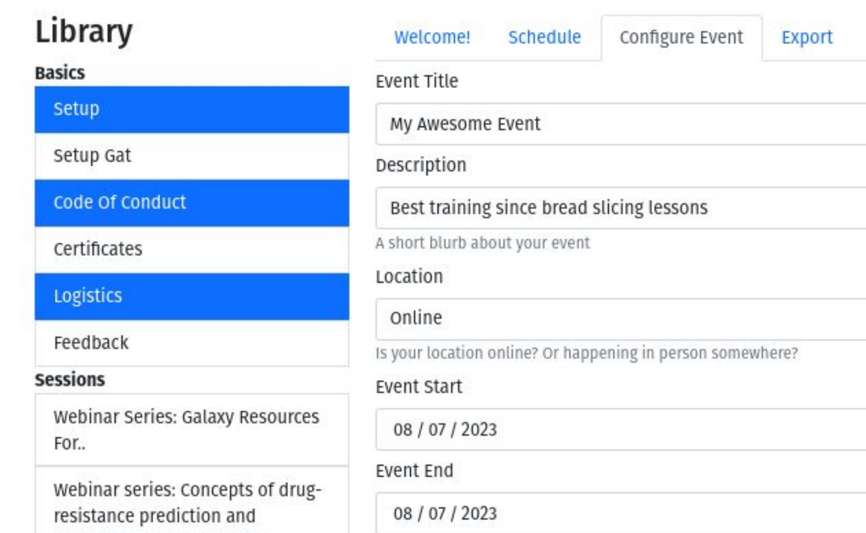
92
Contributors
from around the world



Lesson	Slides	Hands-on	Recordings	Input dataset	Worl
16S Microbial Analysis with mothur (extended)		💻 ▾		⧉	<
16S Microbial Analysis with mothur (short)		💻 ▾	🎥 ▾	⧉	<
16S Microbial analysis with Nanopore data nanopore		💻 ▾			<
Analyses of metagenomics data - The global picture		💻 ▾		⧉	<

🎥 Tutorial (May 2023)

Course Builder: Create your own training events



Basics

- Setup
- Setup Gat
- Code Of Conduct
- Certificates
- Logistics
- Feedback

Sessions

- Webinar Series: Galaxy Resources For..
- Webinar series: Concepts of drug-resistance prediction and mutation-based clustering

Event Title: My Awesome Event

Description: Best training since bread slicing lessons

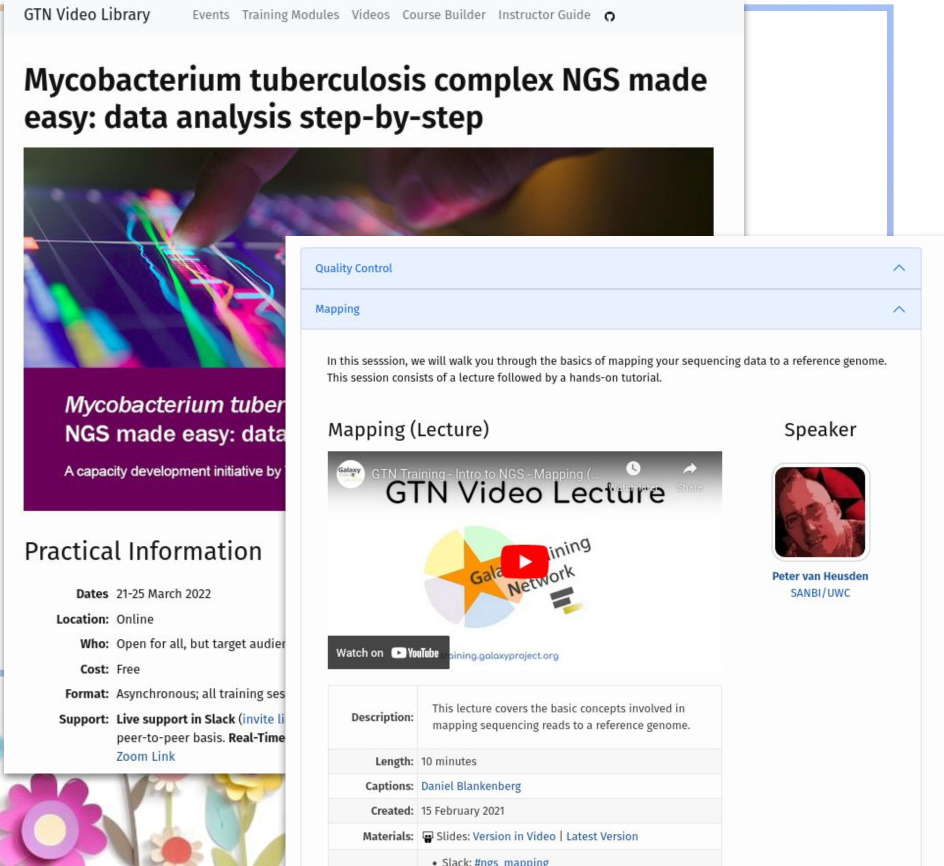
A short blurb about your event

Location: Online

Is your location online? Or happening in person somewhere?

Event Start: 08 / 07 / 2023

Event End: 08 / 07 / 2023



Mycobacterium tuberculosis complex NGS made easy: data analysis step-by-step

Practical Information

Dates	21-25 March 2022
Location:	Online
Who:	Open for all, but target audier
Cost:	Free
Format:	Asynchronous; all training ses
Support:	**Live support in Slack** (invite li peer-to-peer basis. **Real-Time** Zoom Link

Quality Control

Mapping

In this sesssion, we will walk you through the basics of mapping your sequencing data to a reference genome. This session consists of a lecture followed by a hands-on tutorial.

Mapping (Lecture)

Speaker

Peter van Heusden
SANBI/UWC

Description:	This lecture covers the basic concepts involved in mapping sequencing reads to a reference genome.
Length:	10 minutes
Captions:	Daniel Blankenberg
Created:	15 February 2021
Materials:	Slides: Version in Video | Latest Version • Slack: #ngs_mapping

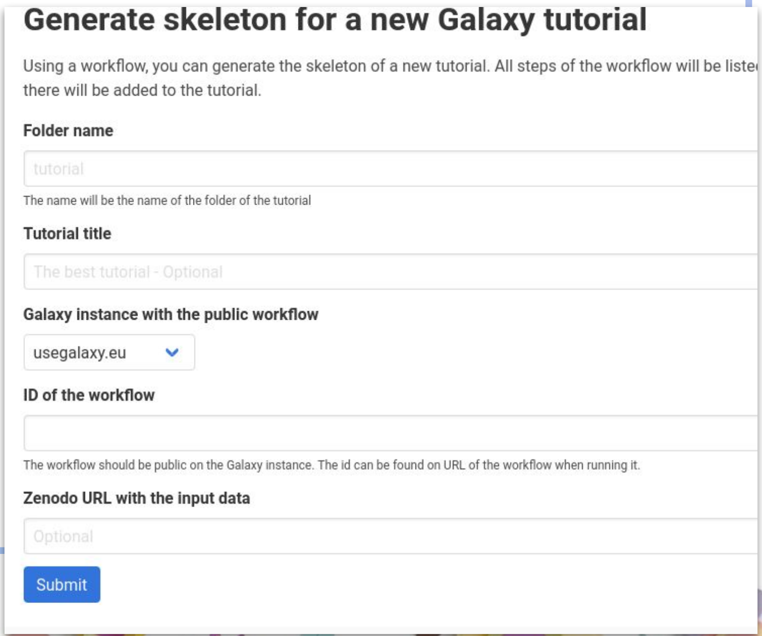
Community

Support & Contributing

Contributing as an Educator

- Everything collaboratively maintained via GitHub
- Markdown based, review & testing, editorial board per topic (can be you!)

- **Want to update tutorials or add your own?**
  - Workflow-to-tutorial service
  - GTN contributing tutorials
  - We will help you!



Generate skeleton for a new Galaxy tutorial

Using a workflow, you can generate the skeleton of a new tutorial. All steps of the workflow will be liste there will be added to the tutorial.

Folder name

tutorial

The name will be the name of the folder of the tutorial

Tutorial title

The best tutorial - Optional

Galaxy instance with the public workflow

usegalaxy.eu

ID of the workflow

The workflow should be public on the Galaxy instance. The id can be found on URL of the workflow when running it.

Zenodo URL with the input data

Optional

Contributing as an Educator

- **Record a video tutorial!**
  - Walk through a tutorial of your choice, performing and explaining all steps
  - Many tutorials do not have recordings yet!
  - Existing videos quickly out of date!
  - Will be automatically included in next Smörgåsbord!
  - Instructions

- **Teach a course with GTN materials?**
  - Add a learning pathway based on your training event!

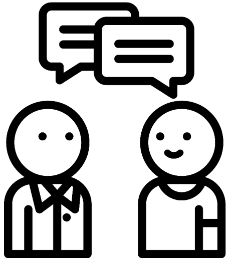
How to engage with the community

Matrix Chat: Galaxy-Training-Network/Lobby

Collaboration Fests: Quarterly (paused temporarily, to resume soon hopefully)

Train-the-trainer: materials, learning pathway & events

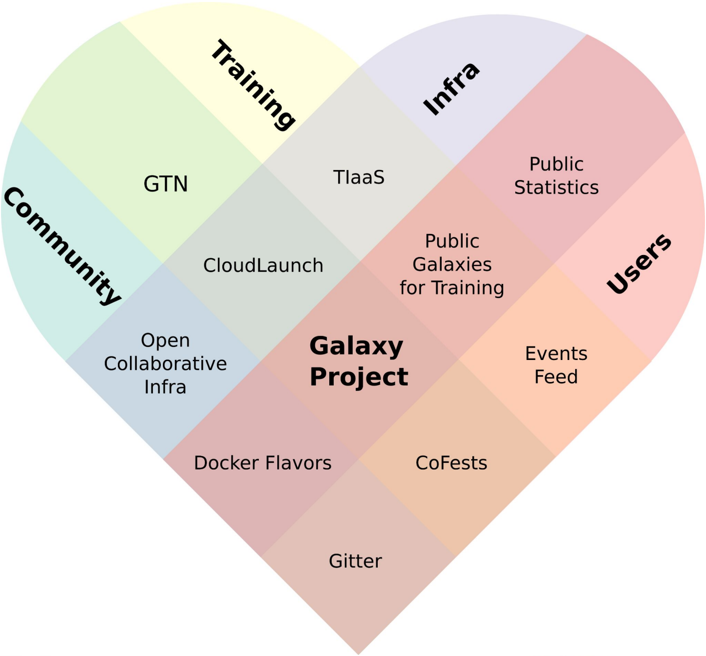
Read our papers: 2018, 2023



GTN CoFests

With the support of the
Erasmus+ Programme
of the European Union

Acknowledgments


Training
Infra
Community
Users
GTN
TlaaS
Public Statistics
CloudLaunch
Public Galaxies for Training
Open Collaborative Infra
Galaxy Project
Events Feed
Docker Flavors
CoFests
Gitter

This would not exist without the

- Galaxy Community
- The GTN Community
- ELIXIR, Gallantries, all funders

Thank you

Geographic Accessibility

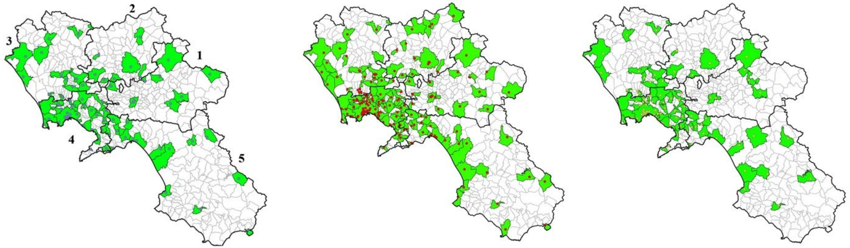
10.1007/s00168-022-01146-6

- **Online materials** can make it accessible world-wide.
- Geographic barriers are a major impediment.
- Not everyone has higher education in reach!
- Are there other barriers (e.g. visas)

(a) Classic Lyceum (CL)
$N_k = 67$ (12%)
$P_k = 188{,}400$ (54%)

(b) Scientific Lyceum (SL)
$N_k = 121$ (22%)
$P_k = 252{,}341$ (72%)

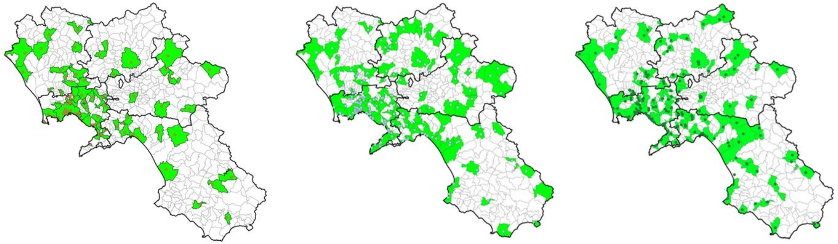
(c) Linguistic Lyceum (LL)
$N_k = 70$ (13%)
$P_k = 201{,}072$ (57%)



(d) Human Sciences Lyceum (HS)
$N_k = 67$ (12%)
$P_k = 182{,}668$ (52%)

(e) Technical Institutions (TI)
$N_k = 110$ (20%)
$P_k = 221{,}167$ (63%)

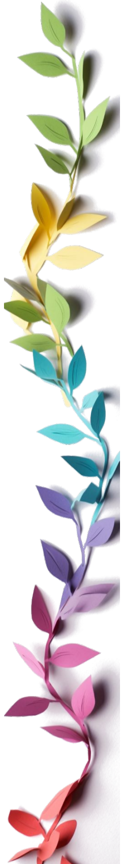
(f) Vocational Institutions (VI)
$N_k = 100$ (18%)
$P_k = 203{,}799$ (58%)



Financial Accessibility

Barriers

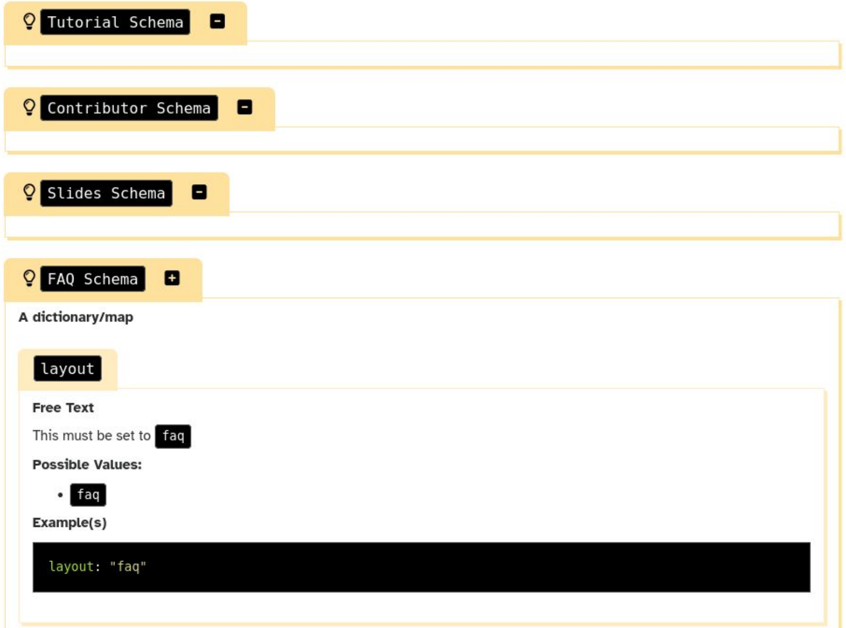
- Training can be expensive
  - Travel
  - Hotel
  - Visas
  - etc.
- Most universities require students to float the cost
- Students don’t always have that money
- Courses can price out LMIC students!



Solutions

- **Make it free**
- You work in an HIC
- You can get a grant if you need
- Differential pricing for LMICs, e.g. OECD list

Cognitive Accessibility: Not just for learners



A dictionary/map

`layout`

Free Text

This must be set to `faq`

Possible Values:

- `faq`

Example(s)

```
layout: "faq"
```

Document your tools!

Document your framework!

Document all of your procedures with simple how to guides

Focus on actionable steps

Multiple Formats

Every page has links to the plain text markdown on GitHub

We have a relatively comprehensive API

“Whatever format you want, you can get”[1] [2]

Compare with Reddit:

shutdown of 3rd party apps means r/Blind will probably shut down.

[1] as long as it’s markdown or html

[2] in the future: PDF, EPUB, kiwix?

Slide Format Spectrum

The GTN converts our slide decks automatically to videos

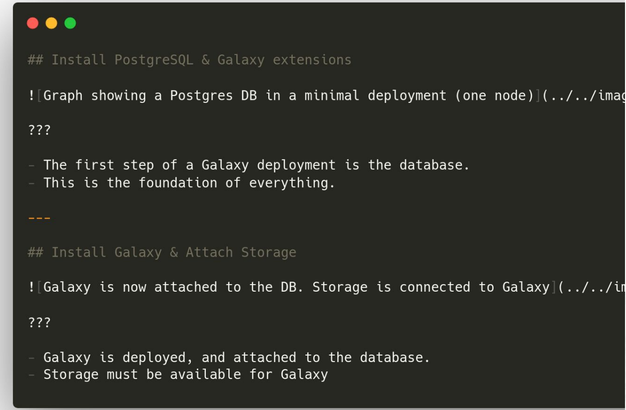
De Bruijn Graph Assembly

Authors: Simon Gladman

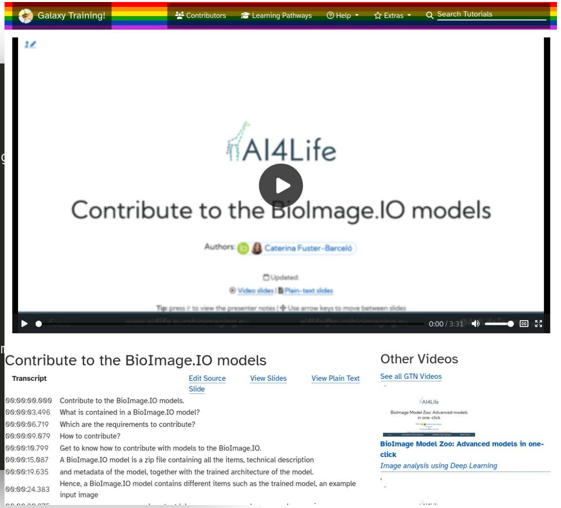
Updated:

Plain-text slides

Tip: press P to view the presenter notes | Use arrow keys to move between slides



```
## Install PostgreSQL & Galaxy extensions

![Graph showing a Postgres DB in a minimal deployment (one node)](../../imag

???

- The first step of a Galaxy deployment is the database.
- This is the foundation of everything.

---

## Install Galaxy & Attach Storage

![Galaxy is now attached to the DB. Storage is connected to Galaxy](../../im

???

- Galaxy is deployed, and attached to the database.
- Storage must be available for Galaxy
```



Contribute to the BioImage.IO models

Transcript

Edit Source Slide | View Slides | View Plain Text

00:00:00.000	Contribute to the BioImage.IO models.
00:00:03.496	What is contained in a BioImage.IO model?
00:00:06.719	Which are the requirements to contribute?
00:00:09.079	How to contribute?
00:00:10.799	Get to know how to contribute with models to the BioImage.IO.
00:00:15.087	A BioImage.IO model is a zip file containing all the items, technical description
00:00:19.635	and metadata of the model, together with the trained architecture of the model.
00:00:24.383	Hence, a BioImage.IO model contains different items such as the trained model, an example input image

Other Videos

See all GTN Videos

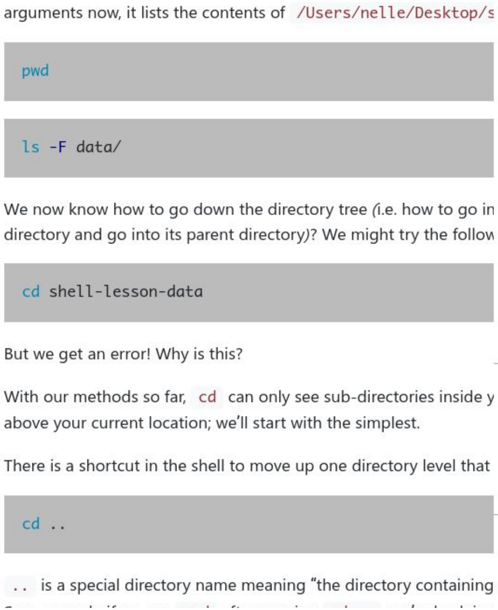
BioImage Model Zoo: Advanced models in one-click

Image analysis using Deep Learning

Tutorial Format Spectrum

- Do you prefer text?
- Or an interactive jupyter notebook?

We make both available (when relevant)



arguments now, it lists the contents of `/Users/nelle/Desktop/s`

```
pwd
```

```
ls -F data/
```

We now know how to go down the directory tree *(*i.e. how to go in

directory and go into its parent directory*)*? We might try the follow

```
cd shell-lesson-data
```

But we get an error! Why is this?

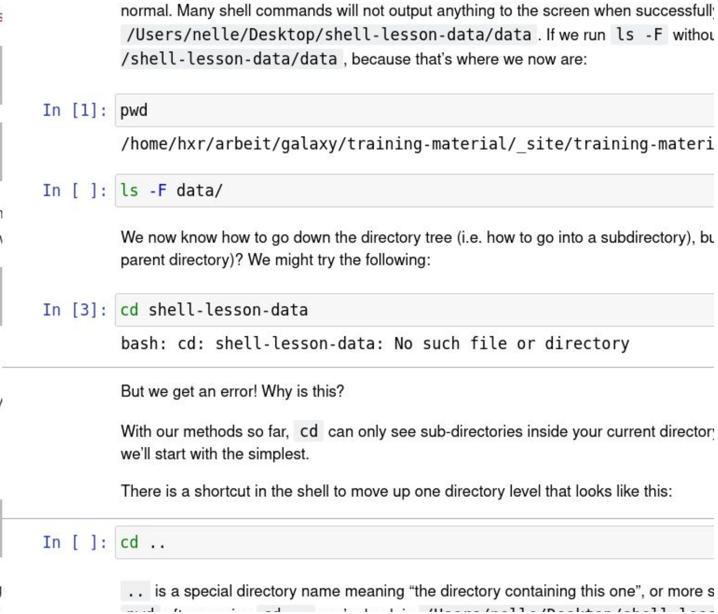
With our methods so far, `cd` can only see sub-directories inside y

above your current location; we'll start with the simplest.

There is a shortcut in the shell to move up one directory level that

```
cd ..
```

`..` is a special directory name meaning "the directory containing



normal. Many shell commands will not output anything to the screen when successfull

`/Users/nelle/Desktop/shell-lesson-data/data` . If we run `ls -F` withou

`/shell-lesson-data/data` , because that's where we now are:

```
In [1]: pwd
        /home/hxr/arbeit/galaxy/training-material/_site/training-materi

In [ ]: ls -F data/
```

We now know how to go down the directory tree (i.e. how to go into a subdirectory), bu

parent directory)? We might try the following:

```
In [3]: cd shell-lesson-data
        bash: cd: shell-lesson-data: No such file or directory
```

But we get an error! Why is this?

With our methods so far, `cd` can only see sub-directories inside your current director

we'll start with the simplest.

There is a shortcut in the shell to move up one directory level that looks like this:

```
In [ ]: cd ..
```

`..` is a special directory name meaning "the directory containing this one", or more s

oschemas

- LD+JSON annotations for your documents
- Makes your resources “Machine Readable”
- The GTN automatically adds this for:
    - Learning Materials
    - Trainers
- Using all of the Bioschemas profiles:
    - Person/0.2-DRAFT-2019_07_19
    - TrainingMaterial/1.0-RELEASE
- Not a task you want to do by hand!
    -

```

<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "LearningResource",
  "http://purl.org/dc/terms/conformsTo": {
    "@id": "https://bioschemas.org/profiles/TrainingMaterial/1.0-RELEASE",
    "@type": "CreativeWork"
  },
  "audience": {
    "@type": "EducationalAudience",
    "educationalRole": "Students"
  },
  "citation": {
    "@type": "CreativeWork",
    "name": "Community-Driven Data Analysis Training for Biology",
    "url": "https://doi.org/10.1016/j.cels.2018.05.012"
  },
  "copyrightHolder": {
    "@type": "Organization",
    "email": "galaxytrainingnetwork@gmail.com",
    "name": "Galaxy Training Network",
    "url": "https://galaxyproject.org/teach/gtn/"
  },
  "dateModified": "2019-05-01 12:49:44 +0000",
  "discussionUrl": "https://gitter.im/Galaxy-Training-Network/Lobby",
  "headline": "16S Microbial Analysis with mothur (short)",
  "interactivityType": "mixed",
  "isAccessibleForFree": true,
  "isFamilyFriendly": true,
  "license": "https://spdx.org/licenses/CC-BY-4.0.html",
```