

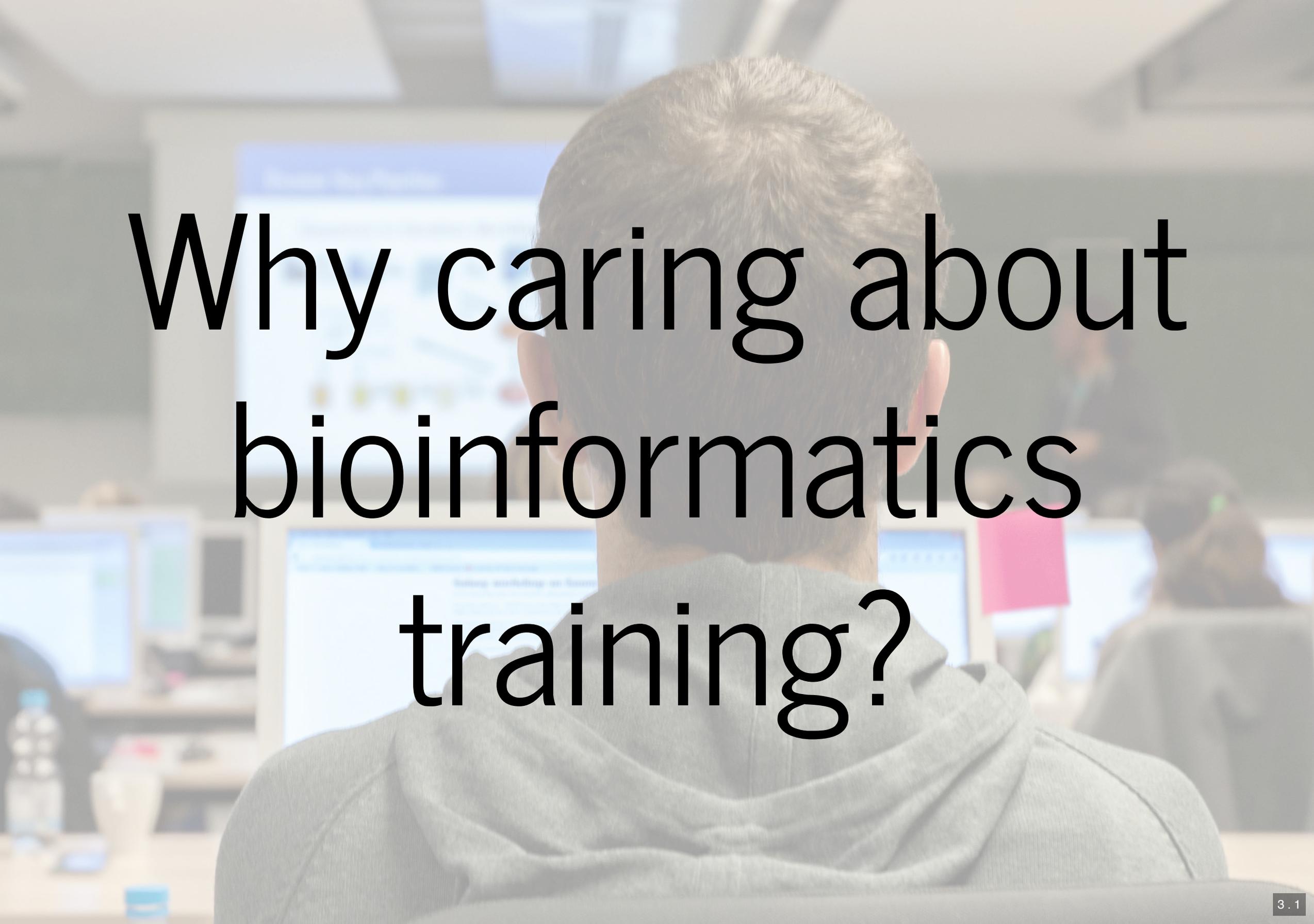
Community-driven training for biological data analysis with the Galaxy Training Network



Picture from Bérénice Batut - Icons from the Noun Project and Flaticon

Bérénice Batut

Galaxy Africa - April 2018



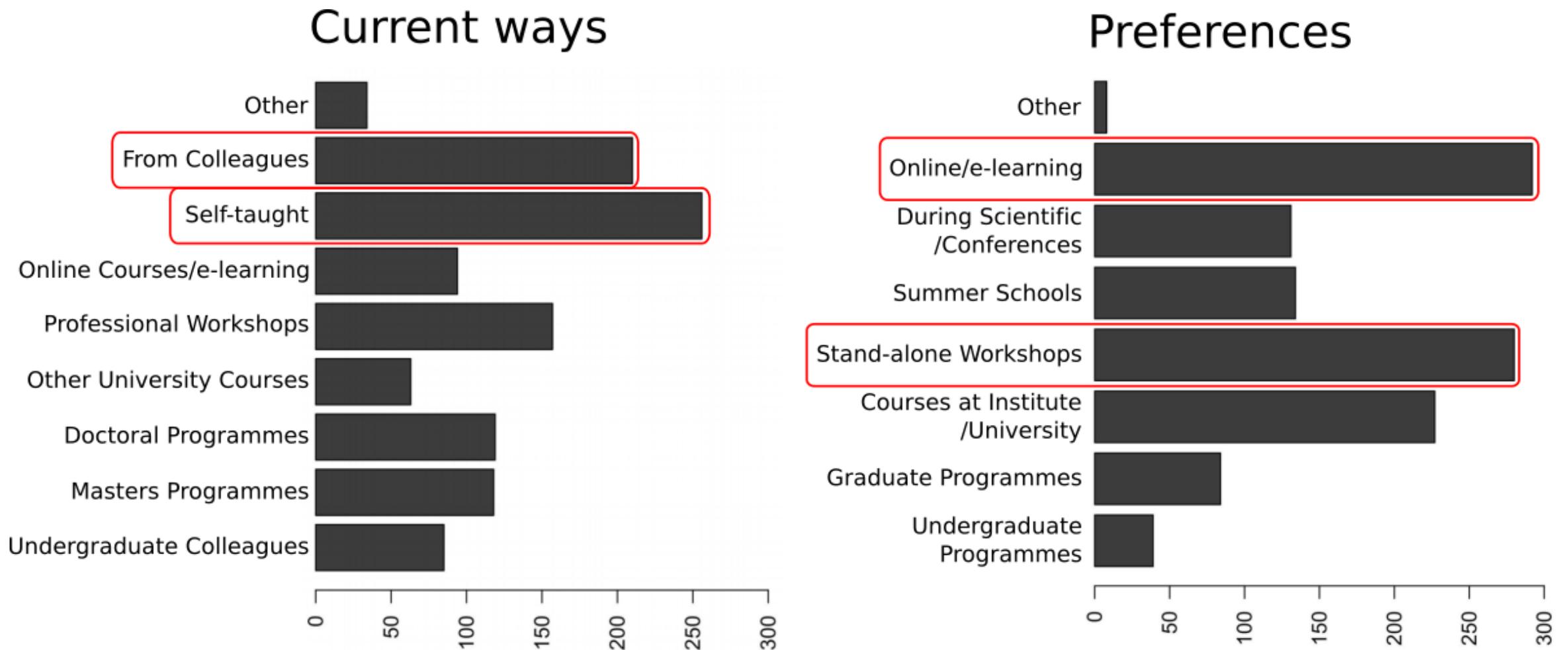
Why caring about
bioinformatics
training?

Need for bioinformatic training

*Bioinformatics has become too central to biology
to be left to specialist bioinformaticians*

- Explosion of data to analyze
- Access to computational power
- Thousand of possible tools for specialized analyses

An increasing demand for learning bioinformatics



Graphs of [Brazas et al, 2017](#)

Galaxy - Uni Freiburg

Tools

deepTools

computeMatrix

computeMatrix has two main output options

scale-regions

Distance in bases to which all regions are going to be stretched or shrunk to

500

Set distance up- and downstream of each genomic region

no

Show advanced output settings

no

Show advanced options

no

History

ChIPseq_sept2016_afternoon

31: plotProfile on data 28: Underlying data

30: plotProfile on data 28: Underlying data

29: plotHeatmap on data 28: Underlying data

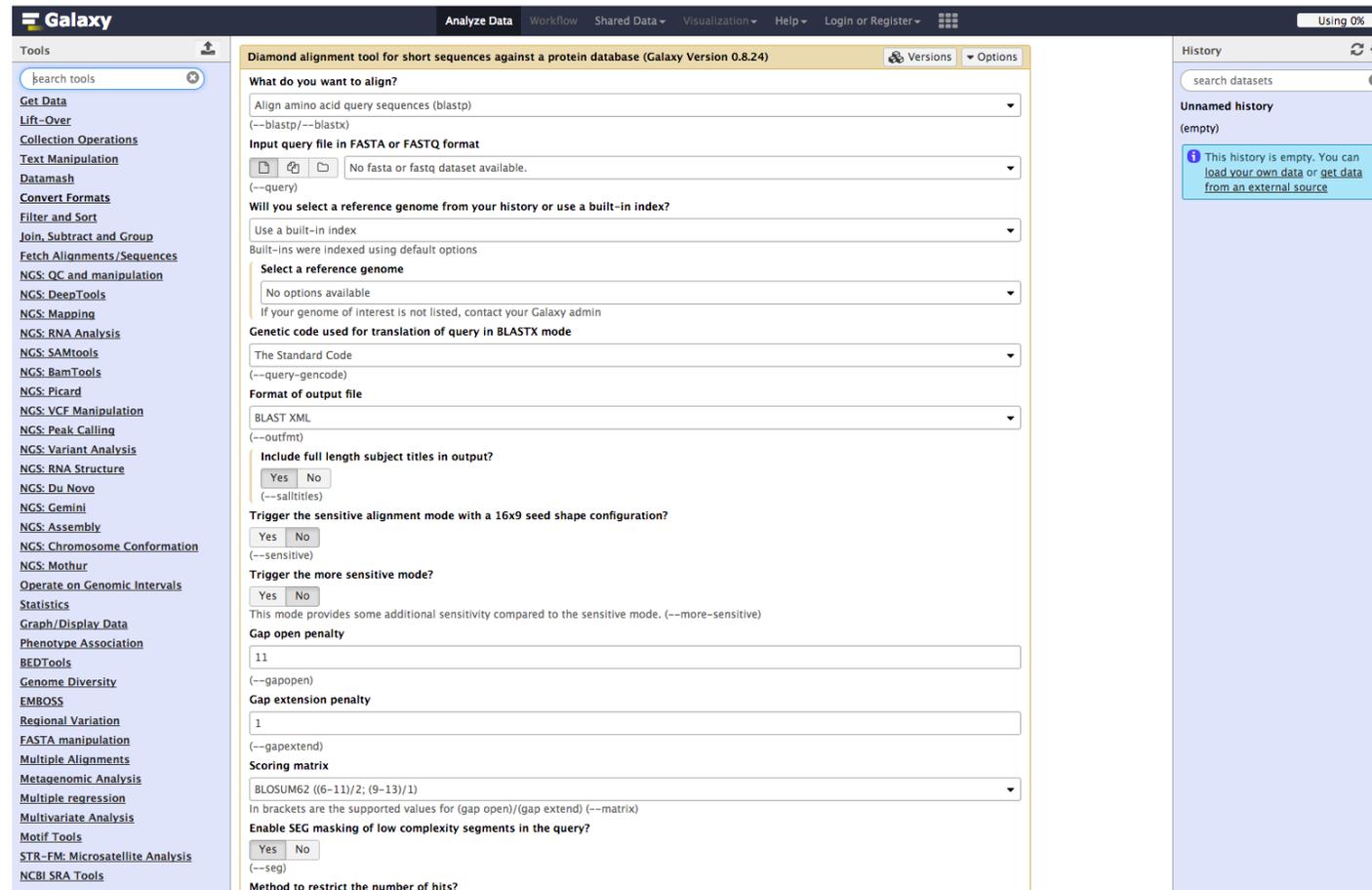
28: computeMatrix on data 14: Underlying data

17: matrix

Galaxy

a great solution!

Computational knowledge: Not required!



- Web interface for numerous bioinformatics tools
- Scalable
- No issue with computer configuration during training

Galaxy

Training
ressources

Best
Practices

Platform
recommendations

Trainer
Directory

Training
Network





Building an infrastructure facilitating data analysis training in life sciences

Requirements for a training infrastructure

- Interactive learning platform
- Support for current research problems
- Usable for effective training for individual users & instructors
- Community driven (content creation and maintenance)
- FAIR: Findable, Accessible, Interoperable, Reusable
- Open

Interactive learning via hands-on tutorials

The image shows a hands-on tutorial for using FastQC in Galaxy. The left pane contains the tutorial text, and the right pane shows the Galaxy interface with the FastQC tool selected.

Hands-on: Quality control

- FastQC** 🛠️: Run FastQC on the FASTQ files to control the quality of the reads
 - “Short read data from your current history”
 - Click on “Multiple datasets”
 - Select all raw datasets

Tip

You can select several files by keeping the CTRL (or COMMAND) key pressed and clicking on the interesting files

- Inspect on the generated webpage for `GSM461177_1` sample

Questions

What is the read length?
▶ Click to view answers

- MultiQC** 🛠️: Aggregate the FastQC reports with
 - “Which tool was used generate logs?” to `FastQC`
 - “Type of FastQC output?” to `Raw data`
 - “FastQC output” to the generated `Raw data` files (multiple datasets)
- Inspect the webpage output from MultiQC

Questions

What is the quality for the sequences for the different files?
▶ Click to view answers

- Trim Galore** 🛠️: Treat for the quality of sequences by running Trim Galore! with
 - “Is this library paired- or single-end?” to `Paired-end`

Galaxy Interface:

- Tools:** fastqc
- NGS: QC and manipulation:** Manipulate FASTQ reads on various attributes, Combine FASTA and QUAL into FASTQ, FastQC Read Quality reports
- Workflows:** All workflows
- FastQC Read Quality reports (Galaxy) Version 0.69:** Short read data from your current history, Contaminant list, Submodule and Limit specifying file, Execute
- Purpose:** FastQC aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a modular set of analyses which you can use to give a quick impression of whether your data has any problems of which you should be aware before doing any further analysis. The main functions of FastQC are:
 - Import of data from BAM, SAM or FastQ/FastQ.gz files (any variant),
 - Providing a quick overview to tell you in which areas there may be problems
 - Summary graphs and tables to quickly assess your data
 - Export of results to an HTML based permanent report
 - Offline operation to allow automated generation of reports without running the interactive application
- FastQC:** This is a Galaxy wrapper. It merely exposes the external package `FastQC` which is documented at `FastQC`. Kindly acknowledge it as well as this
- History:** Unnamed history, 4 shown, (empty), 4: https://zenodo.org/record/1185122/files/GSM461180_2.fastqsanger, 3: https://zenodo.org/record/1185122/files/GSM461180_1.fastqsanger, 2: https://zenodo.org/record/1185122/files/GSM461177_2.fastqsanger, 1: https://zenodo.org/record/1185122/files/GSM461177_1.fastqsanger

Hands-on tutorials built around a "research story"

The screenshot shows the Galaxy Training interface. At the top, there is a dark blue header with the Galaxy Training logo on the left and navigation links: "Transcriptomics", "Introduction slides" (with a dropdown arrow), "Input Dataset", "Literature", "Help" (with a dropdown arrow), and "Edit". Below the header, the main content area has a large heading "Reference-based RNA-Seq data analysis". Underneath this heading is a dark blue bar labeled "Overview". Below the "Overview" bar, there are four sections: "Questions" with two bullet points, "Objectives" with three bullet points, "Requirements" with two blue links, and "Time estimation: 1d" with a clock icon. At the bottom of the screenshot, the word "Introduction" is partially visible.

Galaxy Training! Transcriptomics Introduction slides ▾ Input Dataset Literature Help ▾ Edit

Reference-based RNA-Seq data analysis

Overview

❓ Questions

- What are the effects of Pasilla (PS) gene depletion on splicing events?
- How to analyze RNA sequencing data using a reference genome?

🎯 Objectives

- Analysis of RNA sequencing data using a reference genome
- Analysis of differentially expressed genes
- Identification of functional enrichment among differentially expressed genes

📋 Requirements

- [Galaxy introduction](#)
- [Quality control](#)

🕒 Time estimation: 1d

Introduction

Transcriptomics - Reference-based RNA-Seq data analysis

Hands-on also supported by Interactive Tours

The screenshot displays the Galaxy web interface for Uni Freiburg. The top navigation bar includes 'Galaxy / Uni Freiburg', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The left sidebar lists various tool categories such as 'Get Data', 'Send Data', 'Text Manipulation', and 'Statistics'. The main content area is titled 'Galaxy Tours' and contains a list of interactive tours, including 'Exome Sequencing', 'History Introduction', 'Galaxy UI', and 'Scratchbook - Introduction'. A central dialog box titled 'Welcome to Galaxy' is overlaid on the page, providing instructions on how to navigate the tour and an 'End tour' button. The right sidebar shows a 'History' section with a search bar and a message indicating that the history is empty.

Galaxy / Uni Freiburg Analyze Data Workflow Shared Data Visualization Help User Using 3.8 MB

Tools

search tools

Galaxy Tours

This page presents a list of interactive tours available on this Galaxy server. Select any tour to get started (and remember, you can click 'End Tour' at any time).

- [Exome Sequencing](#) – Sequencing all the protein-coding genes in a genome, the EXOME
- [History Introduction](#) – A detailed introduction to the Galaxy History
- [Galaxy UI](#) – A gentle introduction to the Galaxy User Interface
- [Scratchbook – Introduction](#) – An introduction on how to display multiple datasets and visualizations next to each other.
- [Exome Sequencing](#) – Sequencing all the protein-coding genes in a genome, the EXOME

Welcome to Galaxy

This short tour will guide you through Galaxy's user interface. You can navigate with your arrow keys and leave the tour at any time point with 'Escape' or the 'End tour' button.

« Prev Next » End tour

History

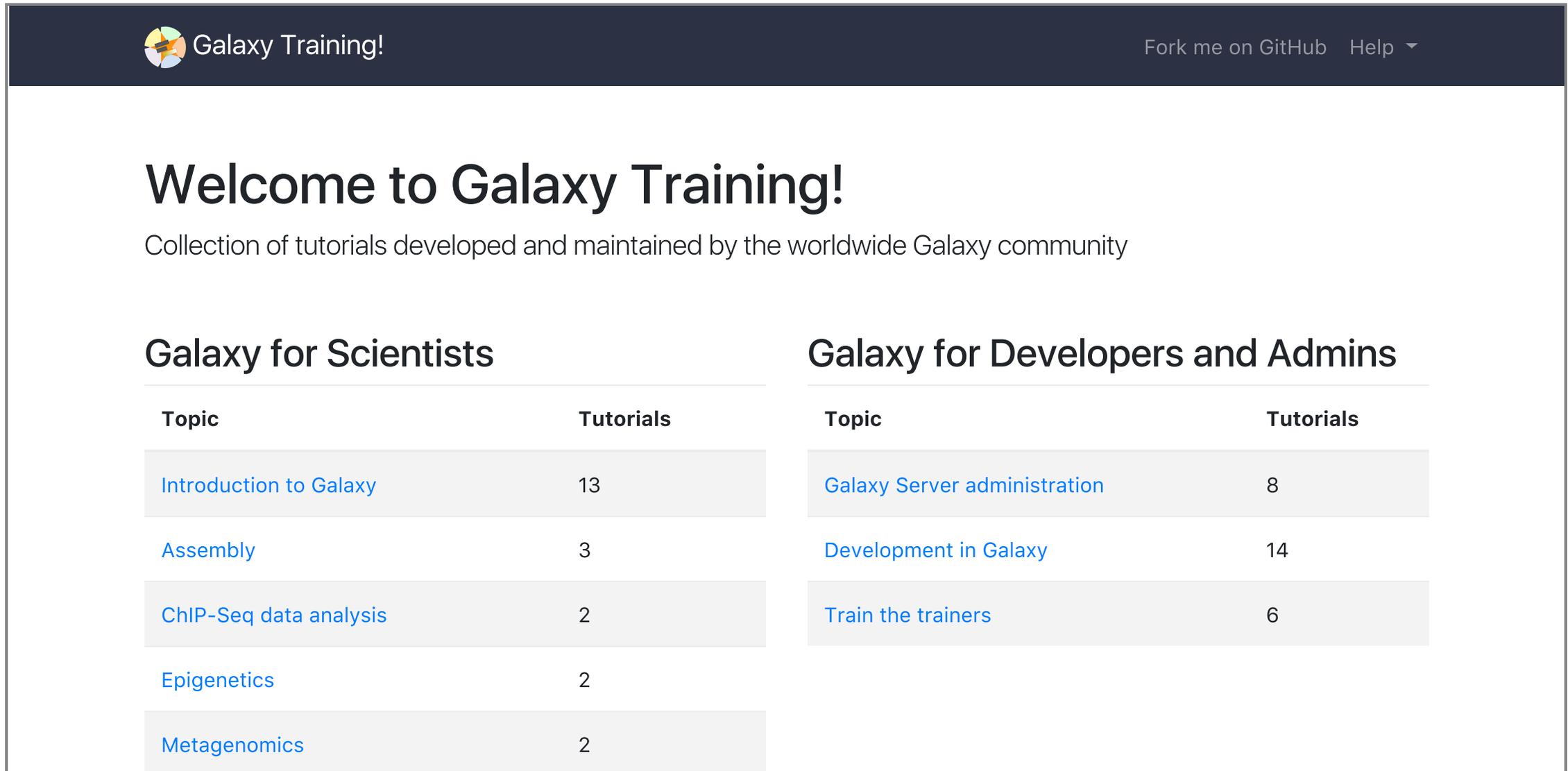
search datasets

Unnamed history

(empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

A collection of materials covering many topics

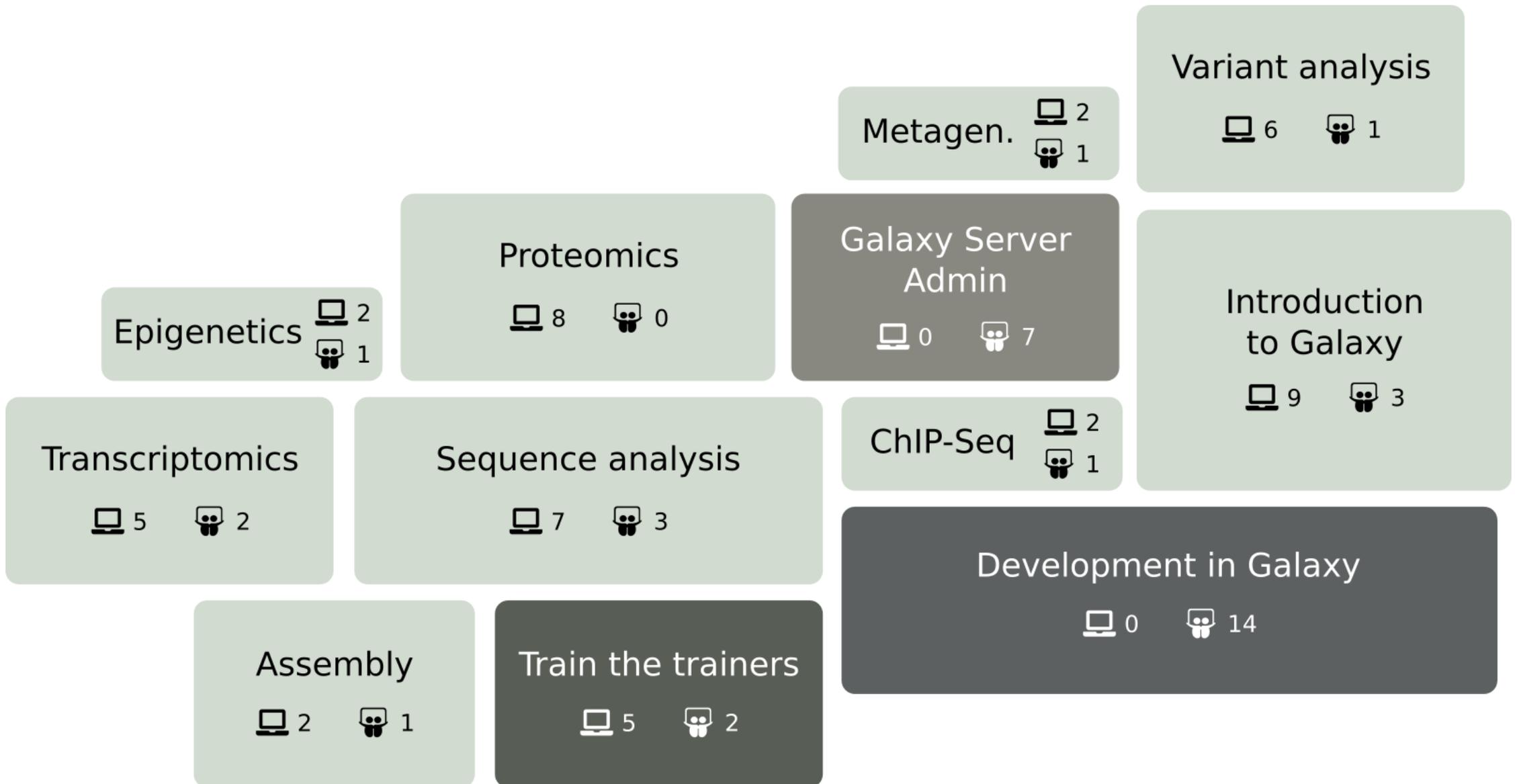


The screenshot shows the Galaxy Training website. At the top left is the logo and text "Galaxy Training!". At the top right are links for "Fork me on GitHub" and "Help". The main heading is "Welcome to Galaxy Training!" followed by the subtitle "Collection of tutorials developed and maintained by the worldwide Galaxy community". Below this are two columns of tables. The left column is titled "Galaxy for Scientists" and the right column is titled "Galaxy for Developers and Admins". Each column contains a table with two columns: "Topic" and "Tutorials".

Galaxy for Scientists		Galaxy for Developers and Admins	
Topic	Tutorials	Topic	Tutorials
Introduction to Galaxy	13	Galaxy Server administration	8
Assembly	3	Development in Galaxy	14
ChIP-Seq data analysis	2	Train the trainers	6
Epigenetics	2		
Metagenomics	2		

<https://training.galaxyproject.org>

A collection of materials covering many topics



More than 80 tutorials!

Used both by individual users

question about tutorial "Analyses of metagenomics data - The global picture"
To: berenice.batut@gmail.com

11. March 2018 at 19:45

Hi

I am doing the tutorial "Analyses of metagenomics data - The global picture" and I have a question.

How I execute the following instruction in galaxy:

Hands-on: Krona

1. **Visualize with Krona** with the following parameters
 - "Input file" to taxonomy output from **Classify.otu** (collection)
 - Set **Is this output from mother?** to **Yes**

I can't find it in menu.

Thanks.

--
██████████

Query regarding sequence analysis tutorial on Galaxy
To: berenice.batut@gmail.com

29. November 2017 at 12:55

Dear sir/mam,
I am following the RNA sequence analysis tutorial on galaxy (<https://galaxyproject.github.io/training-material/topics/sequence-analysis/>). I am unable to find **Augustus** in the Gene predication tutorial. Please help

-- Thanks and Regards

██████████
Near Sola Bridge, Thaltej
Ahmedabad-380054
Gujrat, INDIA

Having trouble with the 16S Microbial Analysis with Mothur tutorial
To: berenice.batut@gmail.com

Training 7. June 2017 at 06:54

NM

Dear Bérénice,

I recently found this tutorial on the galaxy site <https://galaxyproject.github.io/training-material//Metagenomics/tutorials/mothur-miseq-sop> and decided to contact you. I went through it till the "Hands-on: Combine forward and reverse reads into contigs." Unfortunately I'm not sure what dataset to specify on the **rfastq - Reverse Fastq Sequence file**. I made the previous dataset pair but i haven't successfully made any progress going through the workflow. I tried selecting the files manually for the forward and reverse using the multiple datasets icon but it didn't yield the required 6 new collections, trim.contig.fasta and scrap.contig.fasta. Instead i got a log file, contigs.matched file, contigs.qual and contigs.fasta file. In addition, i tried specifying the dataset collection with the paired ataset as input for both **ffastq** and **rfastq** and i got a lot of errors. What could i be doing wrong?

Please advice.

Regards,

██████████

Make.contigs Aligns paired forward and reverse fastq files to contigs as fasta and quality (Galaxy Options)
Version 1.27.0

ffastq - Forward Fastq Sequence file
40: Mock_R2.fastq

rfastq - Reverse Fastq Sequence file
40: Mock_R2.fastq

galaxy reference based rnaseq analysis
To: berenice.batut@gmail.com

28. February 2018 at 19:43

SC

Hi,

Last week, I analyze my data following your reference based rnaseq analysis using HISAT2 with HTseq. But now it seems like it changed to STAR alignment based analysis. Is there a way that I can find other tutorial. If so please provide me the link.

Thanks
Sri

Used both by individual users & instructors

Used both by individual users & instructors

MPI-IE Freiburg @mpi_le

Our Guardians of the galaxy a.k.a. bioinformaticians are at the university #Freiburg giving an #usegalaxy workshop. @fidel_ramirez @vivekbhr

Vivek Bhardwaj @vivekbhr @fidel_ramirez and I are at the #usegalaxy Freiburg workshop helping our new batch of learners doing #bioinformatics with galaxy. From raw reads to heatmaps.

À l'origine en anglais
08:22 - 27 févr. 2018

MiModD @MiModDNews

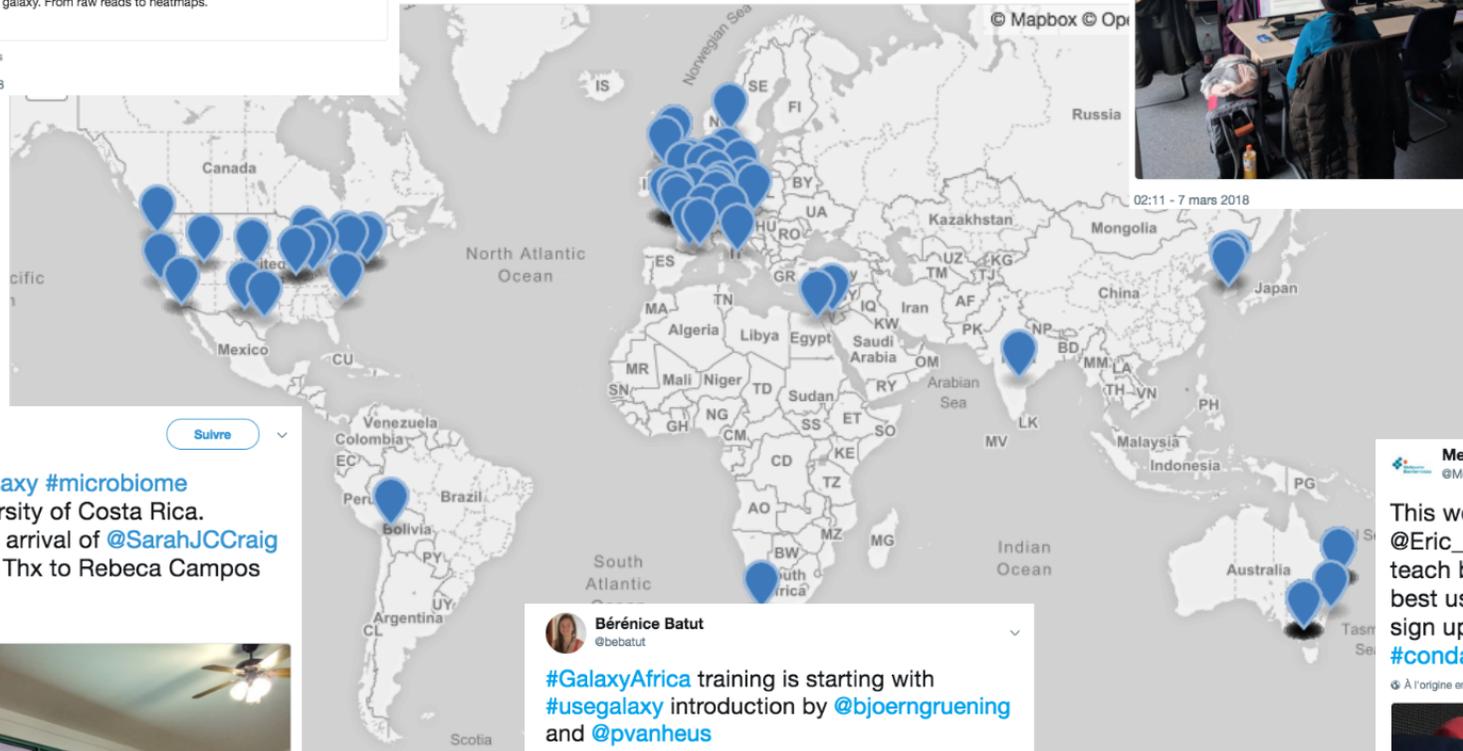
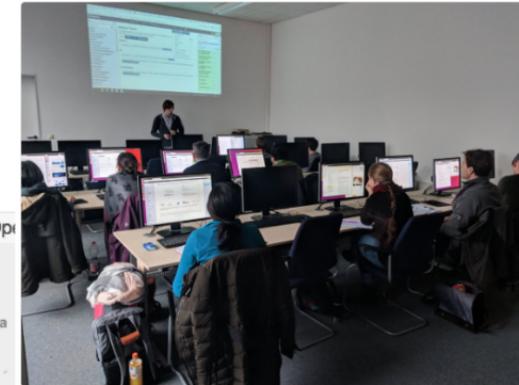
Use the #usegalaxy training material for MiModD at galaxyproject.github.io/training-material ... and read sourceforge.net/p/mimodd/wiki/ ... to get started with MiModD on public servers

À l'origine en anglais
06:53 - 27 mars 2018

de.STAIR @denbi_deSTAIR

#denbi de.STAIR training on RNA-Seq data analysis at @kieluni @denbiOffice #usegalaxy denbi.de/22-training-ca ...

À l'origine en anglais



Kateryna @MakovaLab

Kicking off #usegalaxy #microbiome workshop at University of Costa Rica. Looking forward to arrival of @SarahJCCraig & @DBlankenber. Thx to Rebeca Campos Sanchez for invite!

À l'origine en anglais



Bérénice Batut @bebatut

#GalaxyAfrica training is starting with #usegalaxy introduction by @bjoerngruening and @pvanheus

À l'origine en anglais



MelBioInf @MelBioInf

This week we're lucky enough to have @Eric_Rasche and @shiltemann with us to teach bioinformatics tool developers how to best use Bioconda and Galaxy. Still time to sign up for their workshop next week! #conda #usegalaxy bit.ly/2mJE65r

À l'origine en anglais



Requirements for a training infrastructure

- Interactive learning platform
- Support for current research problems
- Usable for effective training for individual users & instructors
- Community driven (content creation and maintenance)
- FAIR: Findable, Accessible, Interoperable, Reusable
- Open

Building an infrastructure to facilitate community-led content development

- Makes tutorial creation a convenient, hassle-free process
- Enables transparent peer-review and curation to guarantee high-quality and current content

Separation between content and format

Here treatment is the primary factor which we are interested in. The sequencing type is some further information that we know about the data that might affect the analysis. This particular multi-factor analysis allows us to assess the effect of the treatment, while taking the sequencing type into account, too.

```
> ### {% icon comment %} Comment
>
> We recommend you to add as many factors as you think may affect gene expression
in your experiment. It can be the sequencing type like here, but it can also be the
manipulation (if different persons are involved in the library preparation), ...
{: .comment}

> ### {% icon hands_on %} Hands-on: Analysis of the differential gene expression
(1)
>
> 1. Create a new history
> 2. Import the seven count files from [Zenodo]
(https://dx.doi.org/10.5281/zenodo.290221)
> - `GSM461176_untreat_single.deseq.counts`
> - `GSM461177_untreat_paired.deseq.counts`
> - `GSM461178_untreat_paired.deseq.counts`
> - `GSM461179_treat_single.deseq.counts`
> - `GSM461180_treat_paired.deseq.counts`
> - `GSM461181_treat_paired.deseq.counts`
> - `GSM461182_untreat_single.deseq.counts`
>
> 3. **DESeq2** {% icon tool %}: Run **DESeq2** with:
> - "Treatment" as first factor with "treated" and "untreated" as levels and
selection of count files corresponding to both levels
>
>     > ### {% icon tip %} Tip
>     >
>     > You can select several files by keeping the CTRL (or COMMAND) key pressed
and clicking on the interesting files
>     {: .tip}
>
```

Markdown



Here treatment is the primary factor which we are interested in. The sequencing type is some further information that we know about the data that might affect the analysis. This particular multi-factor analysis allows us to assess the effect of the treatment, while taking the sequencing type into account, too.

Comment

We recommend you to add as many factors as you think may affect gene expression in your experiment. It can be the sequencing type like here, but it can also be the manipulation (if different persons are involved in the library preparation), ...

Hands-on: Analysis of the differential gene expression (1)

1. Create a new history
2. Import the seven count files from [Zenodo](https://dx.doi.org/10.5281/zenodo.290221)
 - [GSM461176_untreat_single.deseq.counts](#)
 - [GSM461177_untreat_paired.deseq.counts](#)
 - [GSM461178_untreat_paired.deseq.counts](#)
 - [GSM461179_treat_single.deseq.counts](#)
 - [GSM461180_treat_paired.deseq.counts](#)
 - [GSM461181_treat_paired.deseq.counts](#)
 - [GSM461182_untreat_single.deseq.counts](#)
3. **DESeq2** : Run **DESeq2** with:
 - "Treatment" as first factor with "treated" and "untreated" as levels and selection of count files corresponding to both levels

Tip

You can select several files by keeping the CTRL (or COMMAND) key pressed and clicking on the interesting files

User-friendly HTML

<https://training.galaxyproject.org>

Creating a tutorial

Galaxy Training! Fork me on GitHub Help ▾

Train the trainers

Train the trainers

Material

Lesson	Slides	Hands-on
Introduction to training with Galaxy		
Creating a new tutorial - Writing content in Markdown		
Creating a new tutorial - Defining metadata		
Creating a new tutorial - Setting up the infrastructure		
Creating a new tutorial - Creating Interactive Galaxy Tours		

<http://galaxyproject.github.io/training-material/topics/training/>

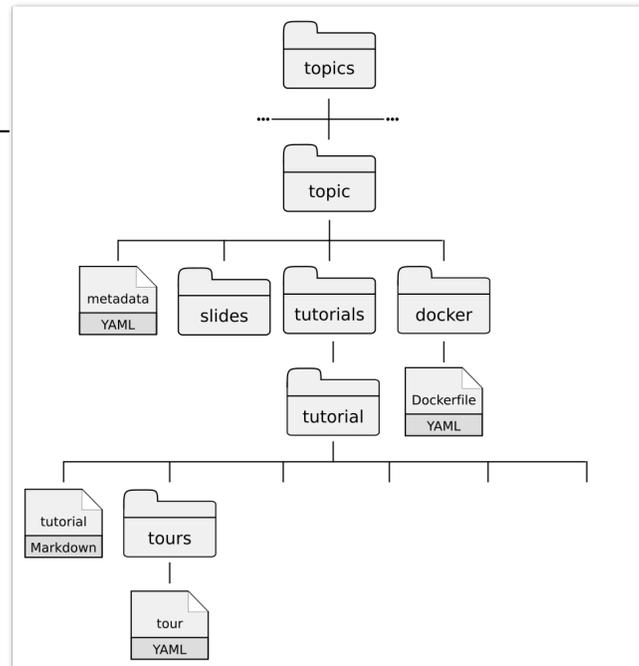
One GitHub repository to collect everything



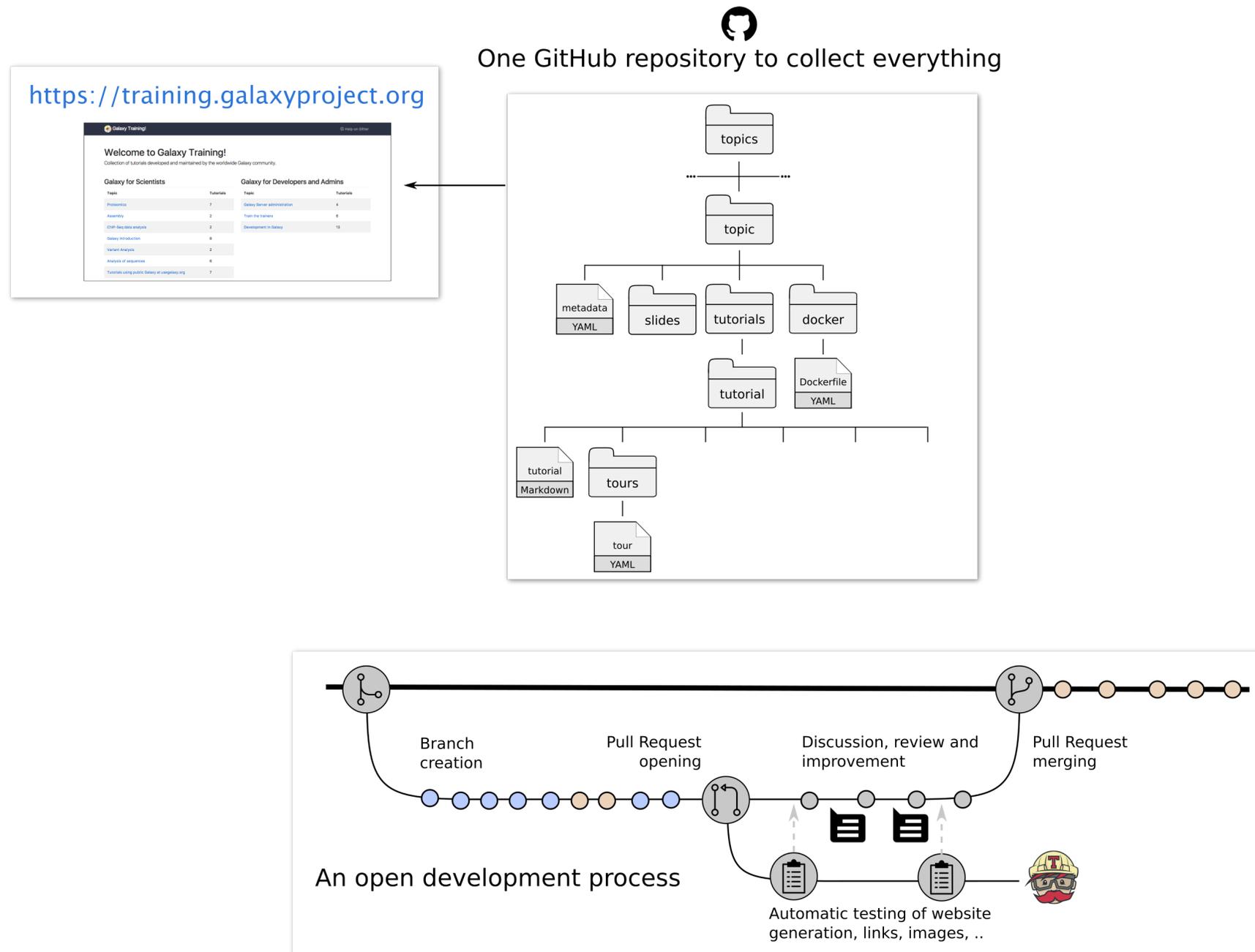
One GitHub repository to collect everything

<https://training.galaxyproject.org>

Galaxy for Scientists		Galaxy for Developers and Admins	
Tutorials	Topics	Topics	Tutorials
Phenomics	7	Galaxy Server administration	4
Assembly	2	Train the trainers	6
ChIP-Seq data analysis	2	Development in Galaxy	13
Galaxy introduction	8		
Next-Genomics	2		
Analysis of sequences	6		
Tutorials using public Galaxy at usegalaxy.org	7		



An open and accessible development process



Community-driven

Community-driven

A constant support

Galaxy Training Network/Lobby <https://new.galaxyproject.org/MailingLists/>

Messages:

- Gildas Le Corguillé @lecorguille (May 31 16:01): 👍
- John Chilton @jmchilton (Jun 01 21:11): I tried checking out build_pipeline and now only the usegalaxy tutorials show up in my training site. Not admin or dev tutorials at all - any ideas?
- Björn Grüning @bgruening (Jun 01 21:12): This branch is not yet working ... @dannon needs to finish up his methal smith build magic
- John Chilton @jmchilton (Jun 01 21:18): Are there slide content changes or just restructuring and metadata reorganization in that branch?
- Dannon Baker @dannon (Jun 01 21:18): Yep, the build `Blocked Plug-in` contains a full reorganization of the content, which doesn't build quite yet. (but that's where we'll want new stuff for now) I wouldn't guess there are likely conflicting slide content changes
- Björn Grüning @bgruening (Jun 01 21:22): Afaik there are no content changes, just the build procedure and the organisation.
- Victoria Dominguez del Angel @vdda (Jun 02 16:15): I'm with Berenice, please take care 🙏
- Slugg70 @Slugg70 (Jun 05 12:29): Hi all, Torsten Seemann came up with an idea after looking at the GTN website. He would like to see tags on the various tutorials for things like Eukaryotic vs prokaryotic specific, or virus etc etc... I reckon it's a good idea.
- Björn Grüning @bgruening (Jun 05 12:29): 👍 We need more tags, also for supported Galaxy instance etc ...
- Slugg70 @Slugg70 (Jun 05 12:32): I agree. Minimum Galaxy version at least.
- Yvan Le Bras @yvanlebras (Jun 05 12:50): +1
- Mallory Freeberg @malloryfreeberg (Jun 05 16:04): +1 Also willing to help with this 😊
- Slugg70 @Slugg70 (Jun 05 16:07): Could be something to add at the hackathon?
- Mallory Freeberg @malloryfreeberg (Jun 05 16:08): Absolutely. I'll add it to suggested data hack topics
- Mallory Freeberg @malloryfreeberg (Jun 05 16:16):

Click here to type a chat message. Supports GitHub flavoured markdown.

PEOPLE: [Grid of 20 user avatars]

ACTIVITY:

- shiltemann on general_metagenomics_tutorial update tutorial (compare) 01:04
- nsoranzo commented #358 Jun 15
- shiltemann on master Add authors of Introduction sli... Merge pull request #359 from ns... (compare) Jun 15
- shiltemann closed #359 Jun 15
- nsoranzo opened #359 Jun 15
- shiltemann on fix-slides change slide deck type (compare) Jun 15
- shiltemann synchronize #358 Jun 15
- shiltemann opened #358 Jun 15
- shiltemann on fix-slides change slide deck type (compare) Jun 15
- nsoranzo commented #354 Jun 15
- shiltemann on general_metagenomics_tutorial start updating amplicon part (compare) Jun 15

Gitter: Galaxy-Training-Network/Lobby

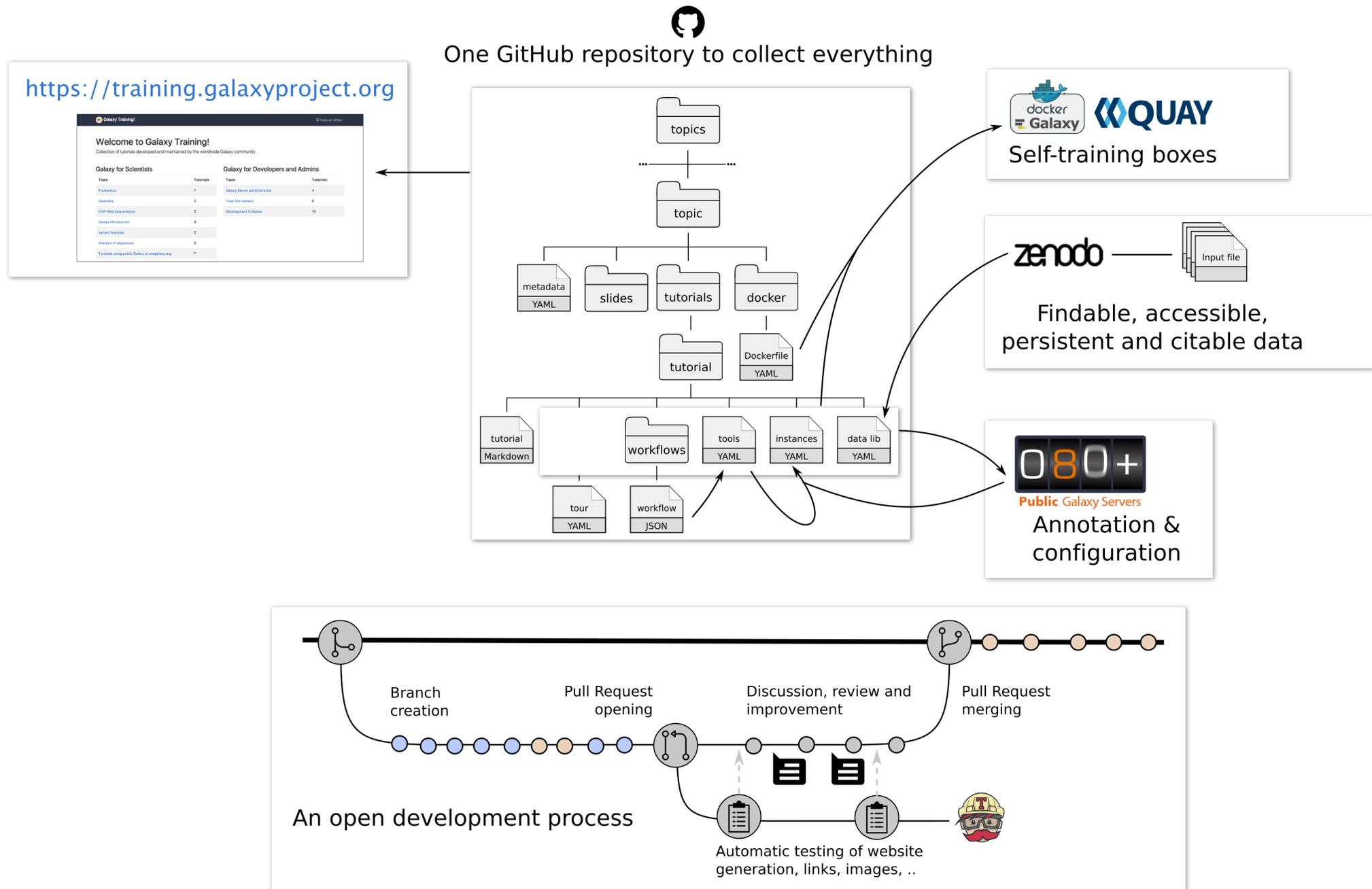
Galaxy Tour Builder

A web extension to develop interactive tours

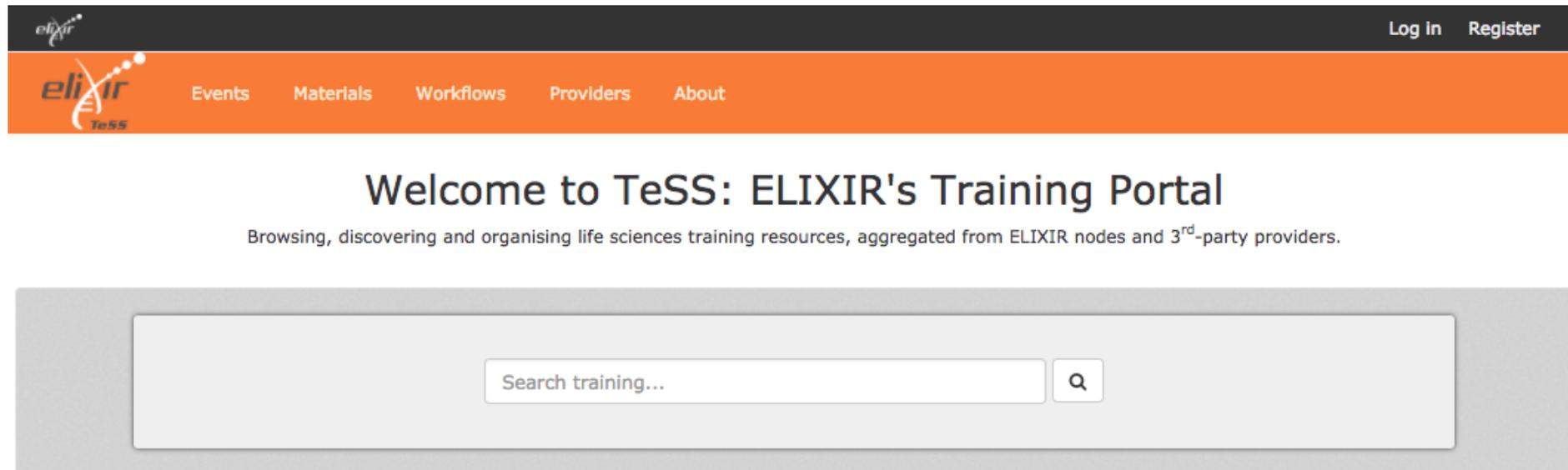
The screenshot shows the Galaxy web interface at <https://usegalaxy.org>. The main content area displays a tour page for the ISMB/ECCB 2017 Tutorial. The page includes a navigation menu on the left with categories like 'Get Data', 'Text Manipulation', and 'NGS: QC and manipulation'. The main content area features a heading 'Galaxy is an open source, web-based platform for data intensive biomedical research...' followed by a promotional banner for the 'ISMB/ECCB 2017 PRAGUE' tutorial with the text 'Making Galaxy work for you' and 'Register now'. Below the banner is a 'Tweets by @galaxyproject' section with two tweets. The right sidebar shows a 'History' panel with a search bar and a list of datasets, including one that has been deleted and another with a warning message.

<https://github.com/TailorDev/galaxy-tourbuilder>

Ensuring accessibility of tutorials



TeSS: the ELIXIR's training portal



The screenshot shows the top navigation bar of the TeSS website. It features the ELIXIR logo on the left and 'Log In' and 'Register' links on the right. Below this is an orange navigation bar with the 'elixir TeSS' logo and menu items: 'Events', 'Materials', 'Workflows', 'Providers', and 'About'. The main content area has a large heading 'Welcome to TeSS: ELIXIR's Training Portal' and a sub-heading 'Browsing, discovering and organising life sciences training resources, aggregated from ELIXIR nodes and 3rd-party providers.' Below this is a search bar with the placeholder text 'Search training...' and a magnifying glass icon.

📅 Events



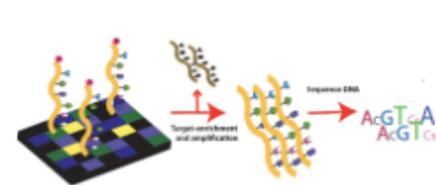
Discover the latest training events and news from ELIXIR nodes and 3rd-party providers.

📖 Materials



Browse the catalogue of training materials offered by ELIXIR nodes and 3rd-party providers.

👤 Workflows



Create training workflows to visualise learning steps and link to resources specific to your training needs.

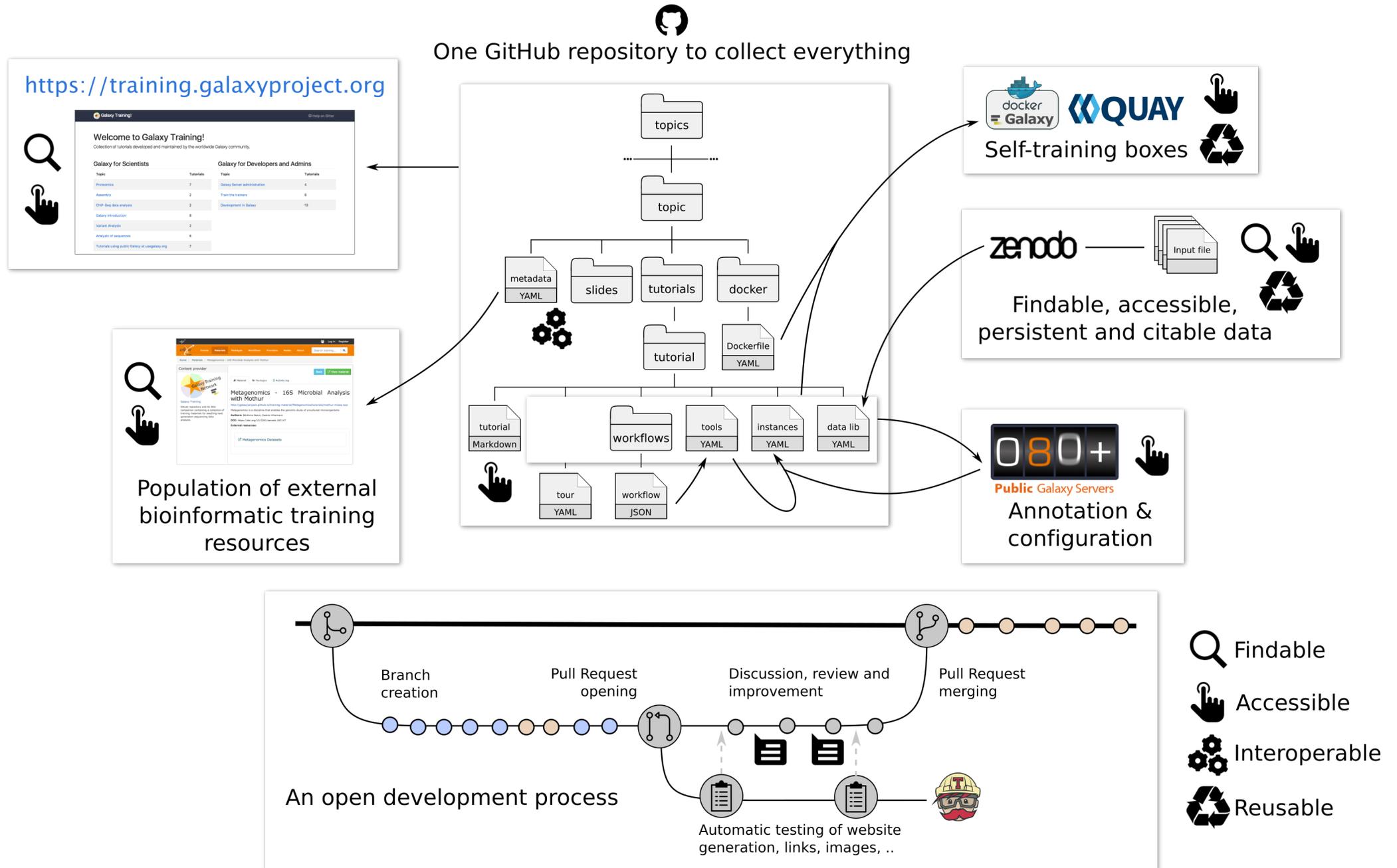
🏠 Providers



Browse training providers to discover training resources they offer and follow links to their materials and courses.

<https://tess.elixir-europe.org/>

Findable, Accessible, Interoperable, Reusable



Requirements for a training infrastructure

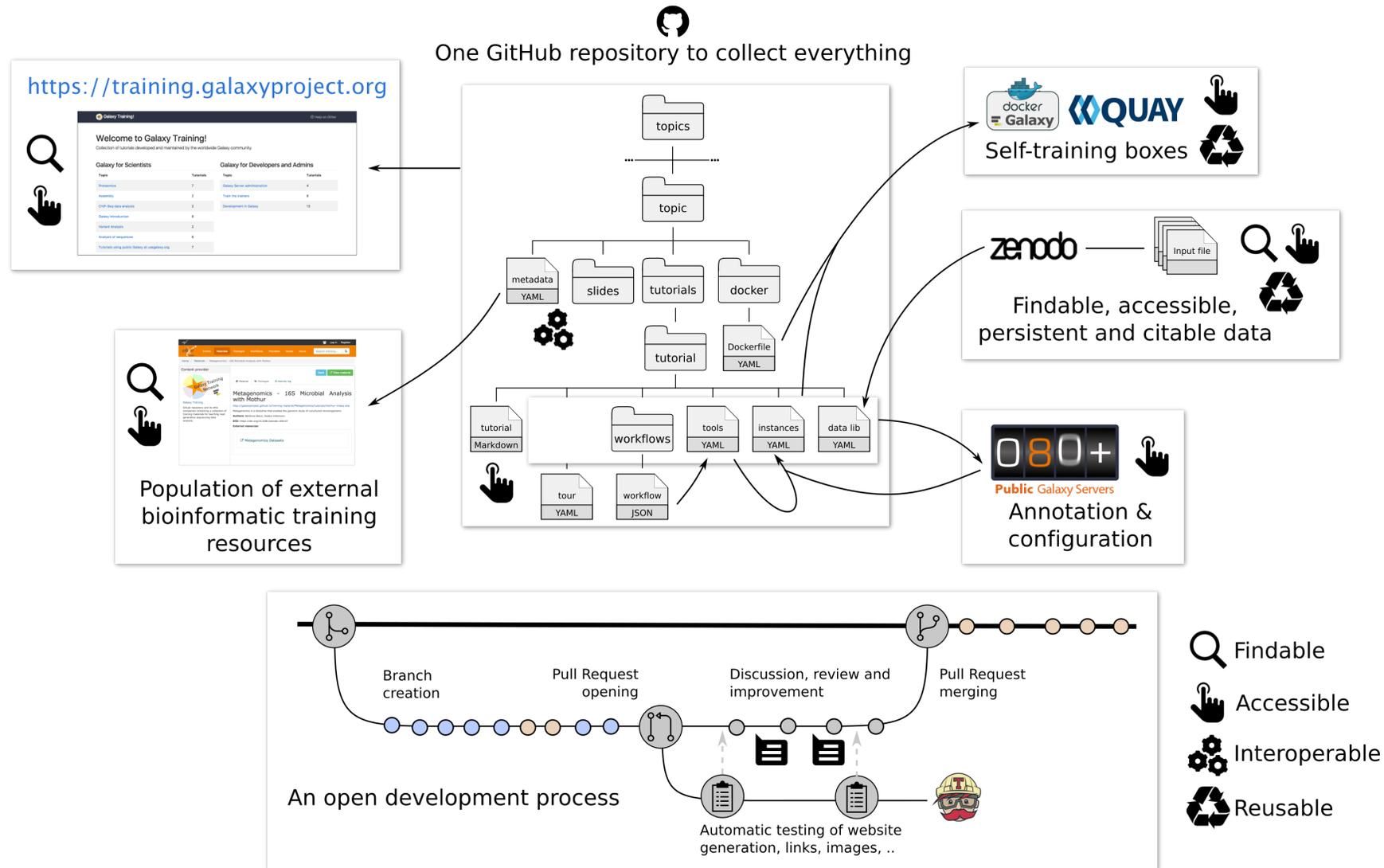
- ☑ Interactive learning platform
- ☑ Support for current research problems
- ☑ Usable for effective training for individual users & instructors
- ☑ Community driven (content creation and maintenance)
- ☑ FAIR: Findable, Accessible, Interoperable, Reusable
- ☑ Open



Thank you!

Sponsors





Rx Community-driven data analysis training for biology

🌐 training.galaxyproject.org

🐙 github.com/galaxyproject/training-material