

Galaxy training material



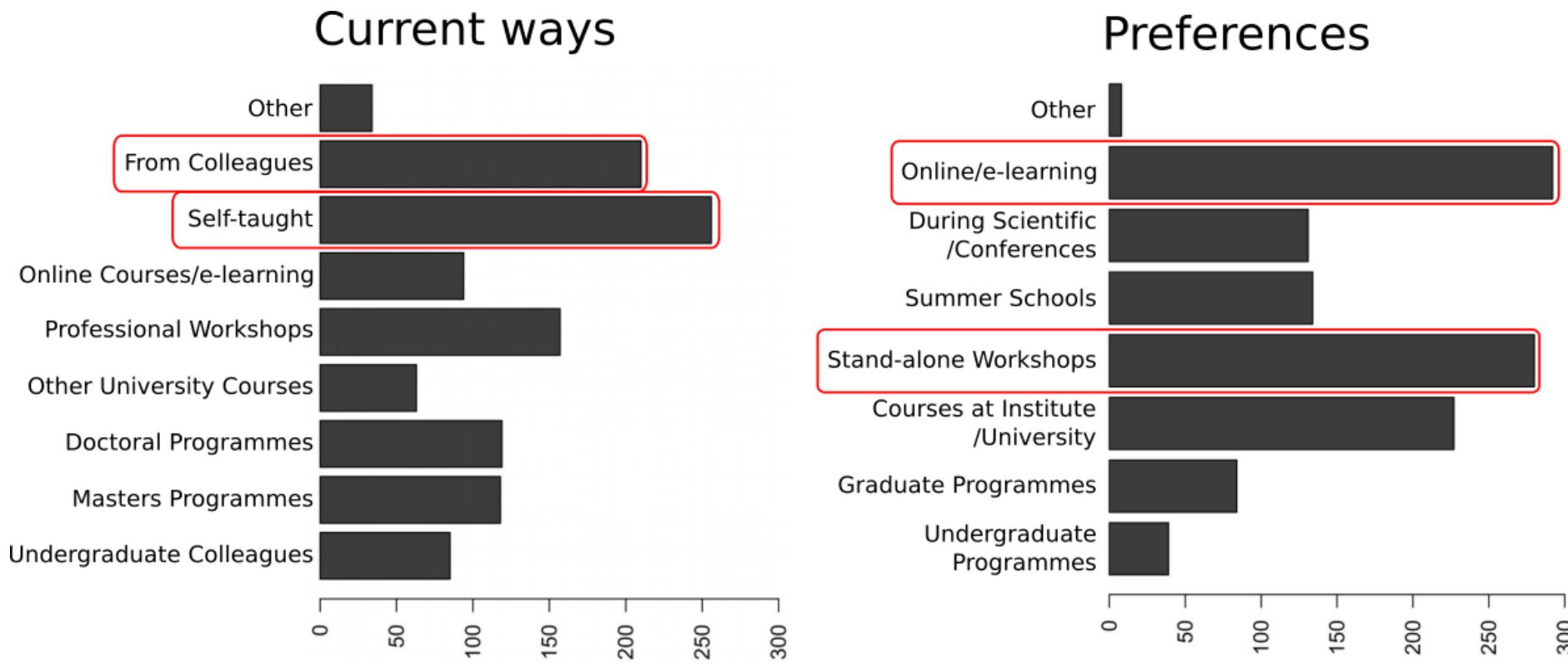
Picture from Bérénice Batut - Icons from the Noun Project and Flaticon

Bérénice Batut, Saskia Hiltemann, The Galaxy Training Network

Biohackathon - Paris - November 2018

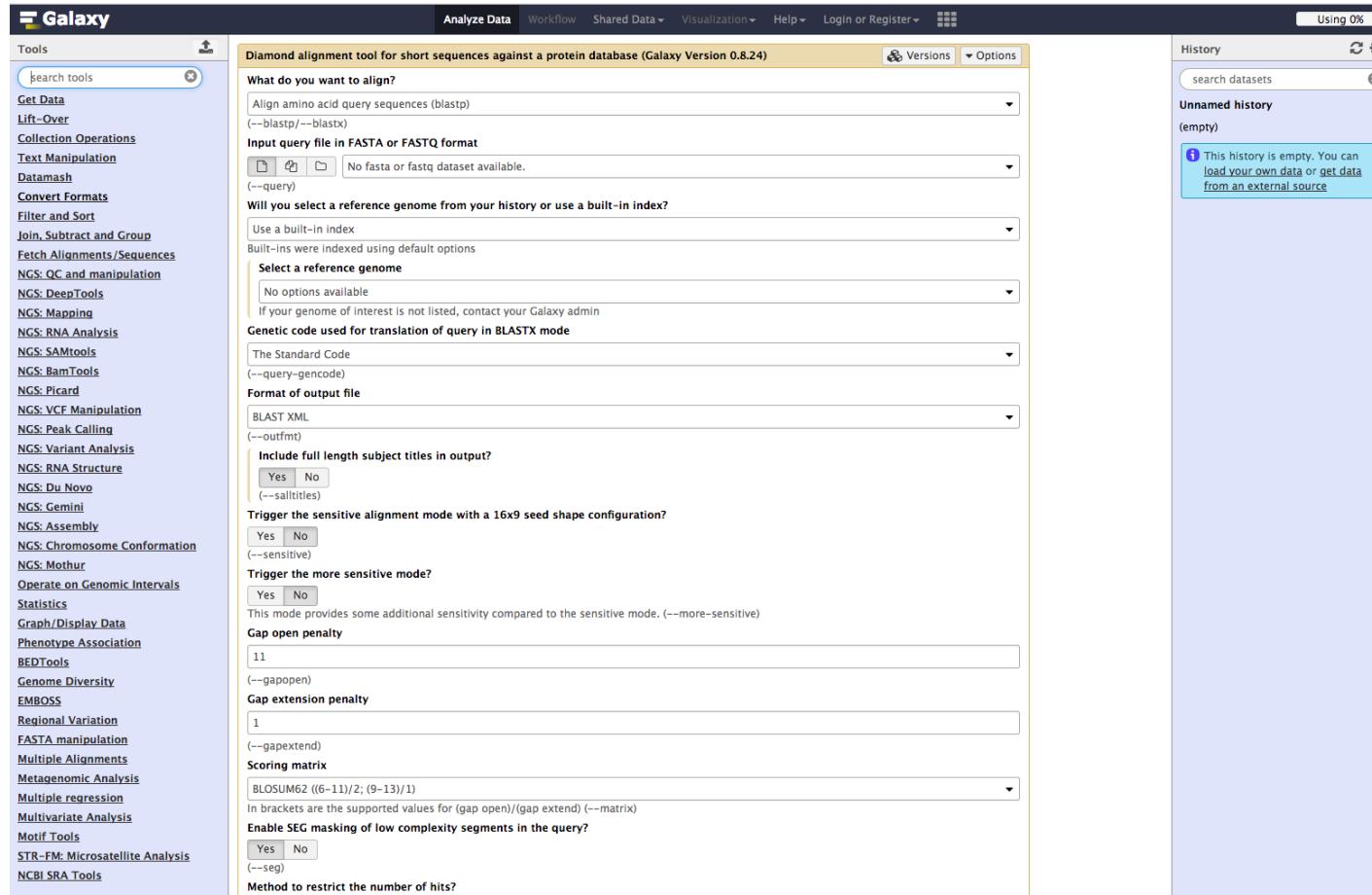
Need and demand for bioinformatic training

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



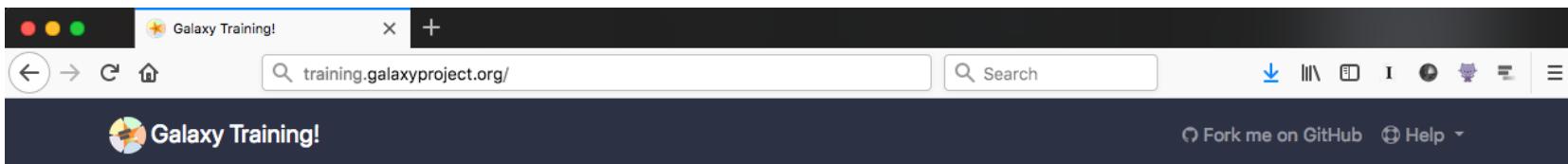
Graphs of Brazas et al, 2017

Galaxy: a great solution !



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

Online training material covering many current research topics



Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

Galaxy for Scientists

Topic	Tutorials
Introduction to Galaxy	14
Assembly	3
ChIP-Seq data analysis	2
Epigenetics	2
Metagenomics	2
Proteomics	8
Sequence analysis	7
Transcriptomics	5

Galaxy for Developers and Admins

Topic	Tutorials
Galaxy Server administration	9
Development in Galaxy	13

How to contribute?

First off, thanks for taking the time to contribute!

You can report mistakes or errors, create more contents, etc. Whatever is your background, there is probably a way to do it: via the GitHub website, via command-line. If you feel it is too much, you can even write it with any text editor and contact us: we will work together to integrate it.

To get you started, check our [dedicated tutorials](#)

<https://training.galaxyproject.org>

Interactive learning via hands-on tutorials built around a "research story"

Sequence quality control is therefore an essential first step in your analysis. We will use similar tools as described in the "Quality control" training: [FastQC](#) and [Cutadapt](#).

Hands-on: Quality control

1. **FastQC** with the following parameters:
 - ["Short read data from your current history"](#): input datasets selected with **Multiple datasets**

Tip: Select multiple datasets

1. Click on [Multiple datasets](#)
2. Select several files by keeping the CTRL (or COMMAND) key pressed and clicking on the files of interest

2. Inspect the webpage output of **FastQC** for the [GSM461177](#) sample

Questions

What is the read length?

[Solution](#) +

3. **MultiQC** with the following parameters to aggregate the FastQC reports:
 - In "Results"
 - "Which tool was used generate logs?": [FastQC](#)
 - In "FastQC output"
 - "Type of FastQC output?": [Raw data](#)

FastQC Read Quality reports (Galaxy)
Version 0.69

Short read data from your current history
No fastq, fastq.gz, fastq.bz2, bam or sam...
Contaminant list
Nothing selected

Manipulate FASTQ reads on various attributes
Combine FASTA and QUAL into FASTQ
FastQC Read Quality reports

Workflows
All workflows

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.

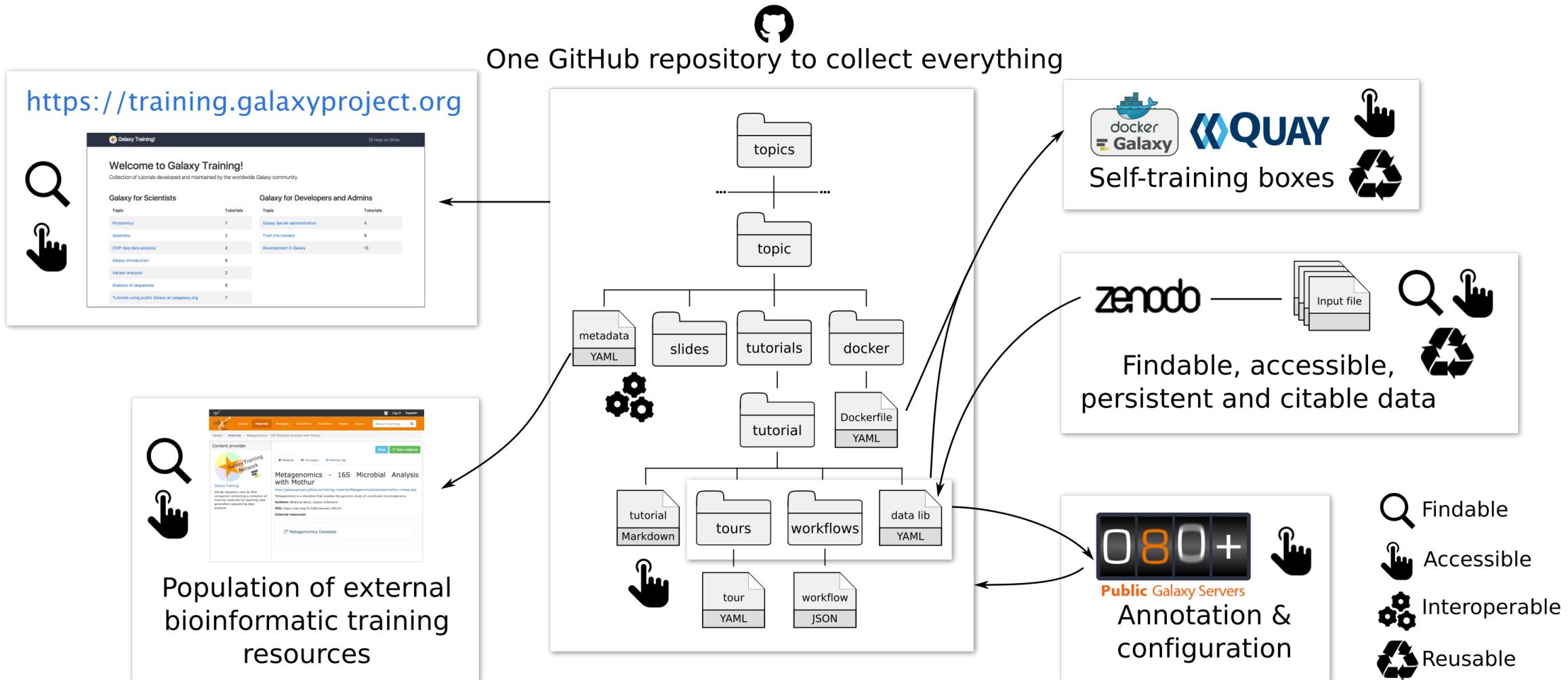
The main functions of FastQC are:

- 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

Usable for effective training for individual users & instructors

An effective (FAIR) training infrastructure



Community-driven



Support

Contributing to the Galaxy Training Material

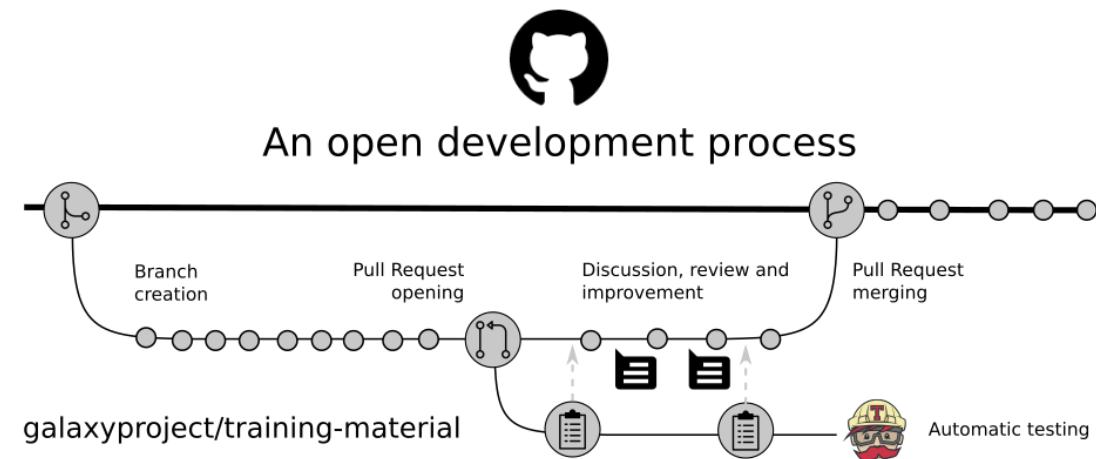
Galaxy is a great solution to train the bioinformatics concepts. Numerous bioinformatics tools are available (almost 5,000 in the Toolshed), it can be used by people without any computer science skills, it trains to use technology, providing available resources and efforts that have made them accessible to researchers. It is however not always easy to find the right material.

In 2016, the Galaxy Training Network decided to set up a new infrastructure for delivering easily Galaxy related training material. The idea was to develop something open and online based on a community effort, as always in Galaxy.

Lessons

	Slides	Hands-on
Overview of the Galaxy Training Material	+	
Contributing with GitHub via command-line	+	+
Contributing with GitHub via its interface	+	
Creating a new tutorial		
Creating a new tutorial - Creating interactive Galaxy Tours		
Creating a new tutorial - Defining the technical infrastructure		
Creating a new tutorial - Slides	+	
Creating a new tutorial - Writing content in Markdown		

Tutorials to contribute



Discussions

Real time chat
Galaxy-Training-Network/Lobby

A screenshot of a real-time chat interface titled "Galaxy-Training-Network/Lobby". The interface shows a list of messages and participants, with a pink header bar at the top.



Many ideas for this week!

CoFest

-  Content
 - New topics, new tutorials
 - Training handbook
-  Global infrastructure
 - FAIRness evaluation of the training
 - Plan internationalisation and localisation support
-  Technical support
 - Workflow testing
 - Docker images
-  Community
 - More visibility for training events
 - Instructors and training philosophy

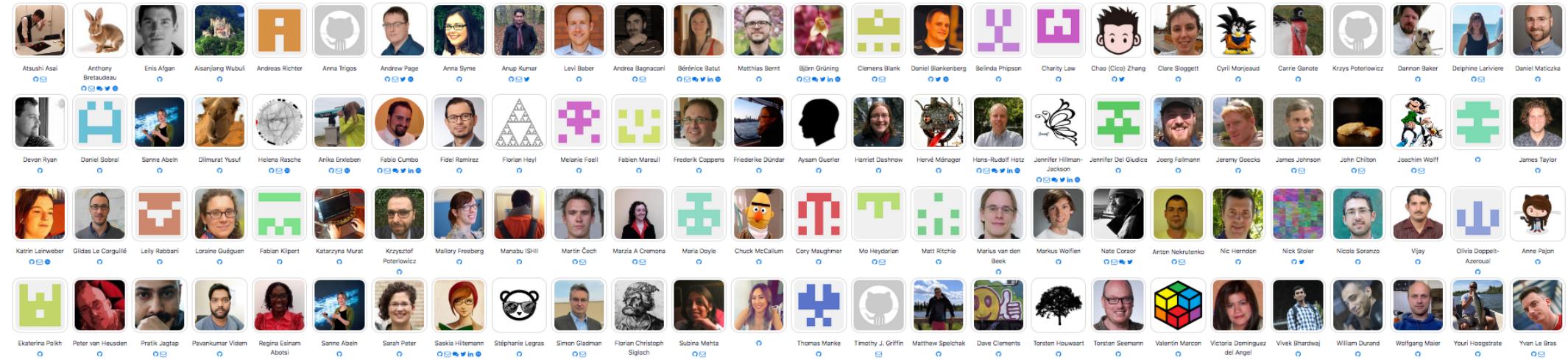
Why should you join?

No specific knowledge or skills required, just motivation!

- Become part of an awesome community!
- Learn about training, Galaxy, and bioinformatics
- Share your insights and knowledge with the community
- Earn a place in our Hall of Fame

Hall of Fame

A big *Thank You!* to the 104 awesome contributors to the Galaxy training materials!



This project would not be possible without its contributors

Thank you!

Join us for this hackathon!



 github.com/galaxyproject/training-material

