

Recent Developments in Galaxy

CGRB Spring Conference
Oregon State University
April 19, 2019

Dave Clements
Johns Hopkins University
Galaxy Community



#usegalaxy @galaxyproject

What is Galaxy? **Keith Bradnam's** definition

"A web-based platform that provides a simplified interface to many popular bioinformatics tools."

From

"13 Questions You May Have About Galaxy"

<http://bit.ly/13questions>

Tools



kraken|



Metagenomic Analysis

[Kraken-mpa-report](#) view report of classification for multiple samples

[Kraken-report](#) view sample report of a classification

[Kraken](#) assign taxonomic labels to sequencing reads

[Kraken taxonomic report](#) view report of classification for multiple samples

[Kraken-translate](#) convert taxonomy IDs to names **updated**

[Kraken-filter](#) filter classification by confidence score **updated**

[Convert Kraken](#) data to Galaxy taxonomy representation

Workflows

All workflows

1	2
#ID	Kraken on data 6 and data 5: Cla
Oscillatoria_nigro-viridis_PCC_7112	
Arthrospira_platensis_NIES-39	
Moraxella_catarrhalis_BBH18	
Pseudomonas_aeruginosa_LES431	
Pseudomonas_resinovorans_NBRC_106553	
Erwinia_pyrifoliae_Ep1/96	
Erwinia_billingiae_Eb661	
Escherichia_coli_APEC_O1	
Escherichia_coli_SMS-3-5	
Escherichia_coli_ABU_83972	
Escherichia_coli_P12b	
Escherichia_coli_O104:H4_str_2009EL-2071	
Escherichia_coli_O104:H4_str_2009EL-2050	
Escherichia_coli_O139:H28_str_E24377A	
Klebsiella_pneumoniae_KCTC_2242	
Klebsiella_variicola_At-22	
Providencia_stuartii_MRSN_2154	
Salmonella_enterica_subsp_enterica_serovar_Heidelberg_str_B182	
Salmonella_enterica_subsp_enterica_serovar_Paratyphi_C_str_RKS4594	
Salmonella_enterica_subsp_enterica_serovar_Typhimurium_str_798	
Salmonella_enterica_subsp_enterica_serovar_Newport_str_USMARC-S3124.1	
Shigella_dysenteriae_Sd197	
Yersinia_pseudotuberculosis_IP_32953	
Buchnera_aphidicola_str_Bp_(Baizongia_pistaciae)	
Buchnera_aphidicola_(Cinara_tujafilina)	

History



search datasets



imported: Robyn Santo Thesis

15 shown, 6 hidden

609.1 MB



17: Kraken taxonomic report on data 9 (Abundances)



16: Sort on data 15



15: Kraken taxonomic report on data 8 (Abundances)



13: Sort on data 12



12: Kraken taxonomic report on data 7 (Abundances)



9: Kraken on data 6 and data 5: Classification



8: Kraken on data 4 and data 3: Classification



7: Kraken on data 2 and data 1: Classification



Interface

- *Rule-based loader to create dataset collections at load time*
- *Dataset name tags preserved throughout analysis*
- *Coming soon: non-linear views of your current analysis*

The screenshot shows the Galaxy web interface. At the top, there's a navigation bar with 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'User', and a star icon. Below this is a panel titled 'Build Rules for Applying to Existing Collection'. It contains a text area for rules and a table of data.

Build Rules for Applying to Existing Collection

Use this form to describe rules for import datasets. At least one column should be defined to a source to fetch data from (URLs, FTP files, etc). Specify a column as a list identifier – specify more to create nested list structures. Specify a column to serve as "collection name" to group datasets.

Rules

- Add column for identifier0. ☒ ✕
- Add new column using (.*)_.* applied to column A ☒ ✕
- Set columns B, C, and A as List Identifier(s) ☒ ✕

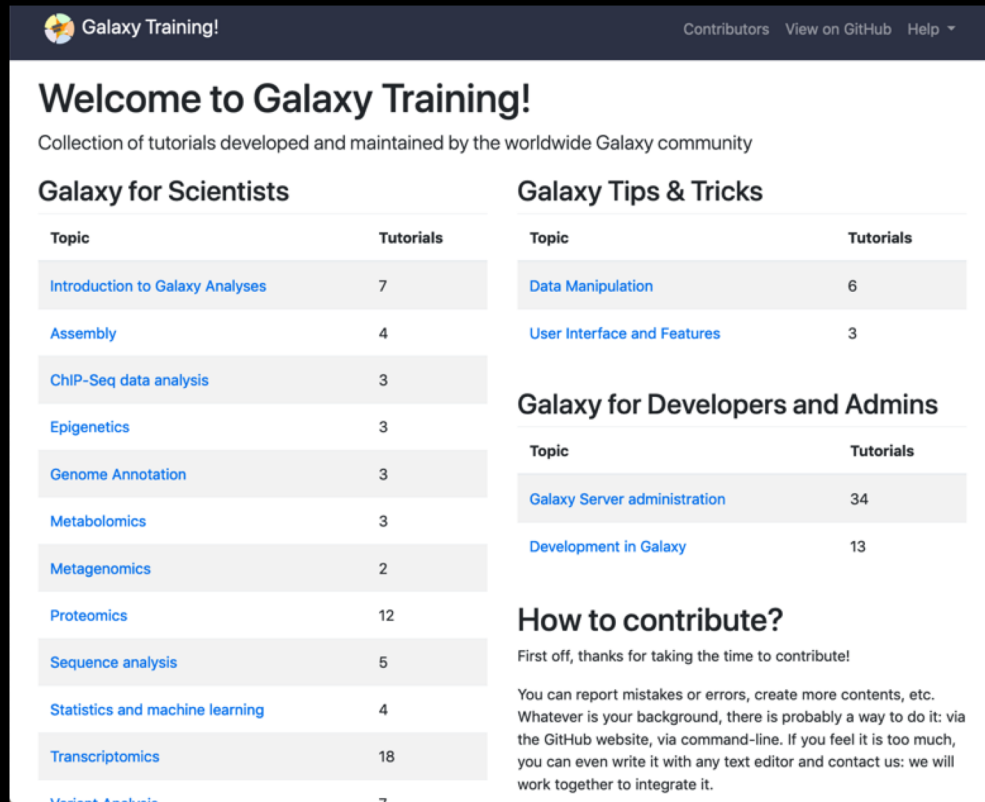
A (List Identifier)	B (List Identifier)	
treated_single_1	treated	sin
treated_paired_2	treated	pa
treated_paired_3	treated	pa
untreated_single_4	untreated	sin
untreated_single_5	untreated	sin
untreated_paired_6	untreated	pa
untreated_paired_7	untreated	pa

+ Rules ▲ + Filter ▲ + Column ▲

Image: John Chilton, Helena Rasche
from Collections: Rule Based Uploader Tutorial

Training Library

- *Galaxy Training Network Library*
- Hands-on tutorials, slides, datasets (in Zenodo), Docker images, workflows, ...
- Galaxy Admin training added this year



Galaxy Training! Contributors View on GitHub Help

Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

Galaxy for Scientists

Topic	Tutorials
Introduction to Galaxy Analyses	7
Assembly	4
ChIP-Seq data analysis	3
Epigenetics	3
Genome Annotation	3
Metabolomics	3
Metagenomics	2
Proteomics	12
Sequence analysis	5
Statistics and machine learning	4
Transcriptomics	18
Variant Analysis	7

Galaxy Tips & Tricks

Topic	Tutorials
Data Manipulation	6
User Interface and Features	3

Galaxy for Developers and Admins

Topic	Tutorials
Galaxy Server administration	34
Development in Galaxy	13

How to contribute?

First off, thanks for taking the time to contribute!

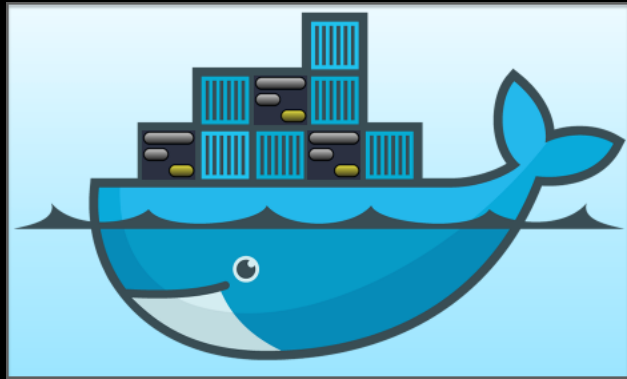
You can report mistakes or errors, create more contents, etc. Whatever is your background, there is probably a way to do it: via the GitHub website, via command-line. If you feel it is too much, you can even write it with any text editor and contact us: we will work together to integrate it.

<https://training.galaxyproject.org/>

Containerization, Conda & Clouds

Containerization greatly simplifies

- wrapping tools for Galaxy
- distributing Galaxy instances
- running Galaxy on cloud infrastructures
- AWS, Google Cloud, OpenStack supported
- Azure support coming



kubernetes



Google Cloud

UseGalaxy.* & Platforms Directory

- UseGalaxy.org Project run server, hosted at TACC
- UseGalaxy.eu European server, hosted by U Freiburg
- UseGalaxy.org.au Australian server, hosted by U Queensland

- galaxyproject.org/use/ Links to over 150 platforms for using Galaxy
right now



Thanks

Dave Clements

Galaxy Project

Johns Hopkins University

clements@galaxyproject.org