

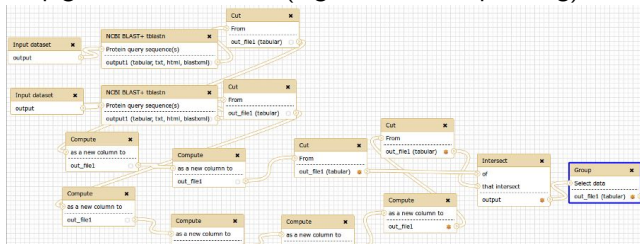
de.STAIR Galaxy flavor for RNA-Seq analysis

K. Riege¹ and S. Hoffmann^{1,2}

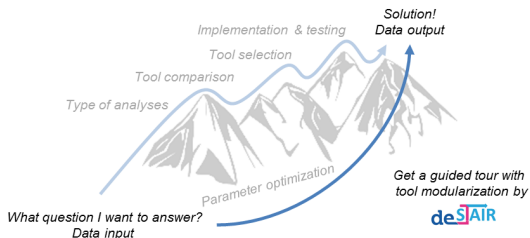


TBI Winterseminar 2018

- de.STAIR @ de.NBI
- Services for Structured Analysis and Integration of RNA-Seq experiments
- Hand over RNA-Seq data analysis back to experimentalists
 - Preprocessing, rRNA depletion, error correction
 - Mapping and file type homogenization
 - Quantification, expression analysis and GO enrichment
 - Differential splicing
 - Transcript classification (taxonomic, RNA families)
 - Prediction of RNA interaction
 - Integration of epigenetic information (e.g. Bisulfite Sequencing)



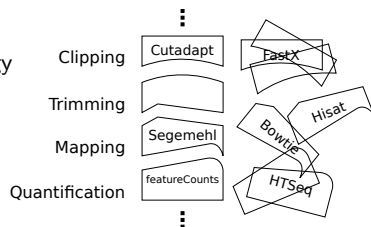
- de.STAIR
 - Tool selection
 - Parameterization
 - Adaption to different pro- and eukaryotic NGS protocols
 - RNA-Seq (total RNA mRNA)
 - smallRNA-Seq
 - miRNA-Seq
 - dRNA-Seq
 - mdRNA-Seq
 - metaRNA-Seq
 - Visualization
 - Statistics
 - Reports
 - Advices for troubleshooting



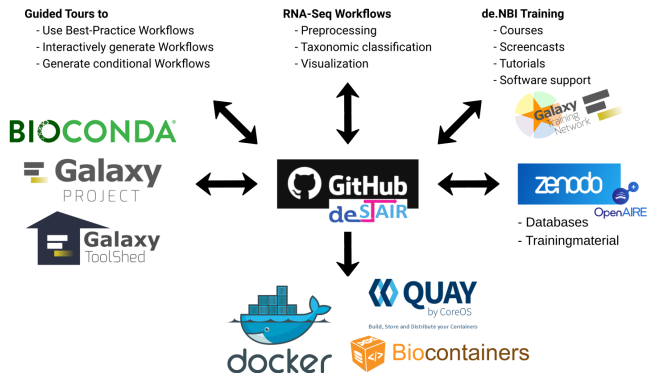
Static workflows? No!

Given: Guided tours for workflows

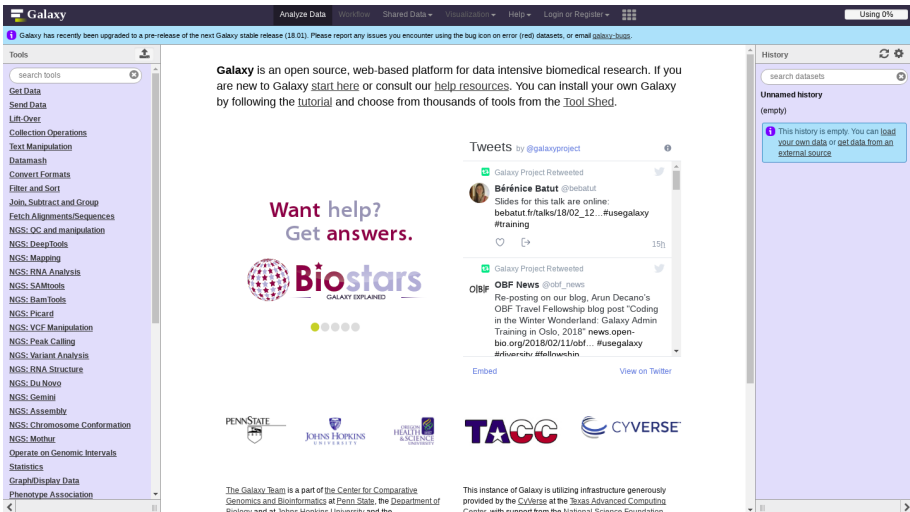
- de.STAIR: **one tool** → **one workflow** → **one guided tour**
 - Crafting workflows and guided tours
 - Setup assembly-framework
 - Recommendation system
 - Most suitable tools
 - Most widely used tools
 - Previously used tools
 - Suggestions by file type ontology



- Q/A system via webhooks
- Galaxy instance as webservice for workflow generation
- Export of workflows and CLI commands
- Use for training events



Proof of concept



Galaxy Analyze Data Workflow Shared Data Visualization Help Login or Register Using 0%

Galaxy has recently been upgraded to a pre-release of the next Galaxy stable release (18.01). Please report any issues you encounter using the bug icon on error (red) datasets, or email galaxy-bugs.

Tools

- Get Data
- Send Data
- Life-Over
- Collection Operations
- Text Manipulation
- Datamash
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- NGS: QC and manipulation
- NGS: DeepTools
- NGS: Mapping
- NGS: RNA Analysis
- NGS: SAMtools
- NGS: Bam Tools
- NGS: Picard
- NGS: VCF Manipulation
- NGS: Peak Calling
- NGS: Variant Analysis
- NGS: RNA Structure
- NGS: Du Novo
- NGS: Gemini
- NGS: Assembly
- NGS: Chromosome Conformation
- NGS: Mother
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#). You can install your own Galaxy by following the [tutorial](#) and choose from thousands of tools from the [Tool Shed](#).

Want help? Get answers.

BioStars
GALAXY EXPLAINED

Tweets by @galaxyproject

- Galaxy Project Retweeted
Bérénice Batut @bebatut
Slides for this talk are online: bebatut.fr/talks/18/02_12...#usegalaxy
#training
- Galaxy Project Retweeted
OBF News @obf_news
Re-posting on our blog, Arun Decano's OBF Travel Fellowship blog post "Coding in the Winter Wonderland: Galaxy Admin Training in Oslo, 2018" news.openbio.org/2018/02/11/obf...#usegalaxy
#rivercity #followshin

Embed **View on Twitter**

Penn State **Johns Hopkins University** **Open Health & Science** **TACC** **CYVERSE**

The [Galaxy Team](#) is a part of the [Center for Comparative Genomics and Bioinformatics at Penn State](#), the [Department of Biology](#) and at [Johns Hopkins University](#) and the

This instance of Galaxy is utilizing infrastructure generously provided by the [CYVerse](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).

History

search datasets

Unnamed history (empty)

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

Proof of concept

Galaxy

Analyze Data | Shared Data | Workflow | Help | Login or Register

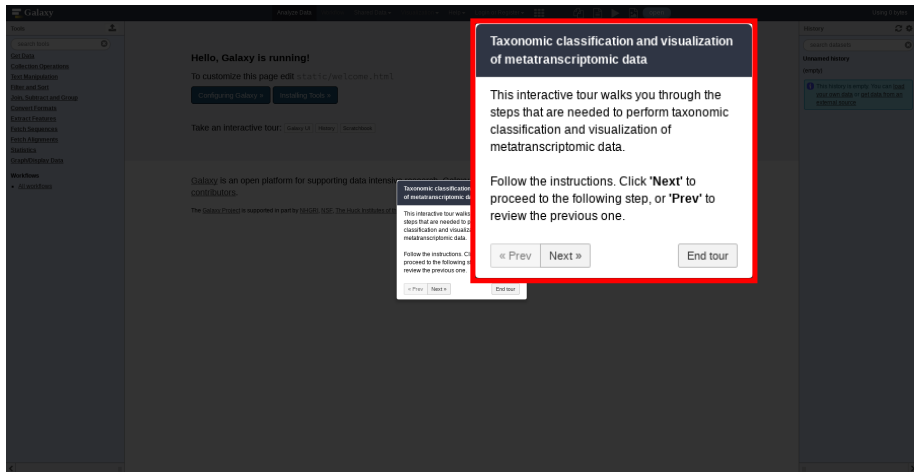
Using 0 bytes

close

Please select:

- destar.training.preprocessing_mt_data
- destar.training.visualization_mt_data

Submit



Taxonomic classification and visualization of metatranscriptomic data

This interactive tour walks you through the steps that are needed to perform taxonomic classification and visualization of metatranscriptomic data.

Follow the instructions. Click 'Next' to proceed to the following step, or 'Prev' to review the previous one.

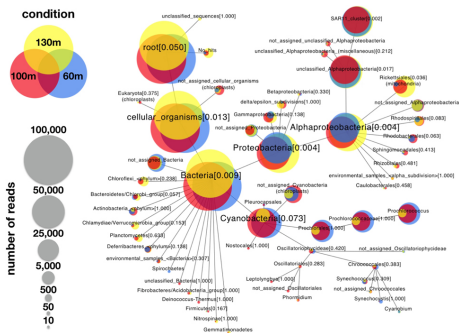
« Prev Next » End tour

Proof of concept

The screenshot shows the Galaxy web interface. A central dialog box titled "Download from web or upload from disk" is open. It has tabs for "Regular", "Composite", and "Collection". Below the tabs, it says "You added 1 file(s) to the queue. Add more files or click 'Start' to proceed." There is a table with columns: Name, Size, Type, Genome, Settings, and Status. One row is visible: "New File", "84 b", "Auto-detect", and "Additional Species". Below the table, there is a text input field containing a URL: "https://zenodo.org/record/574367/files/preprocessed_metatranscriptome_dataset.fasta". At the bottom of the dialog, there are buttons for "Choose local file", "Paste/Fetch data", "Pause", "Reset", and "Start".

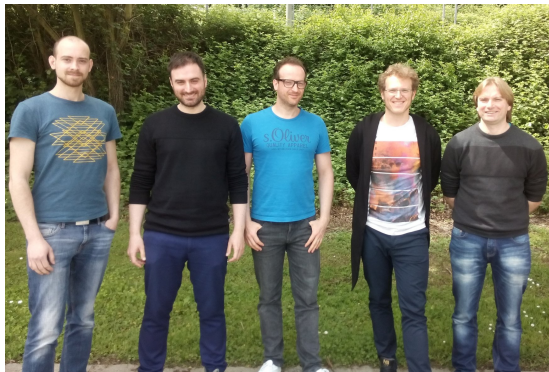
A tooltip box is overlaid on the dialog, titled "Import the metatranscript data". It contains the text: "Transcriptomic data are contained in a FASTA file that must have been pre-processed for high quality reads. Load our sample dataset by clicking 'Next'." Below the text are three buttons: "« Prev", "Next »", and "End tour".

- Training Freiburg, 2017
- Training Jena, **June 2018 - announced soon!**
 - Pitfalls in RNA sequencing
 - Galaxy UI
 - Guided tours
 - Data processing
 - Taxonomic classification
 - CoVennTree visualization



- Go online
- Integrate new tools into Galaxy toolshed and BioConda
 - Segemehl v3 - enhanced short read mapping
 - DIEGO - detection of differential alternative splicing
 - Trapline - DGE analysis and miRNA target prediction
 - GLASSgo - sRNA prediction
 - CopraRNA - sRNA target prediction
 - CoVennTree - weighted Venn based visualization of populations from metaRNA-Seq data

Acknowledgments



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA



Bundesministerium
für Bildung
und Forschung

- Steve Hoffmann
- Olaf Wolkenhauer, Andrea Bagnacani, Markus Wolfien
- Wolfgang Hess, Steffen Lott
- Björn Grüning



- [1] Lott S, Wolfien M, Riege K, Bagnacani A et.al.
Customized workflow development and data modularization concepts for RNA-Sequencing and metatranscriptome experiments.
J Biotechnol, 261:85-96, November, 2017.