de.STAIR Galaxy flavor for RNA-Seq analysis

K. Riege[1] and S. Hoffmann[1,2]

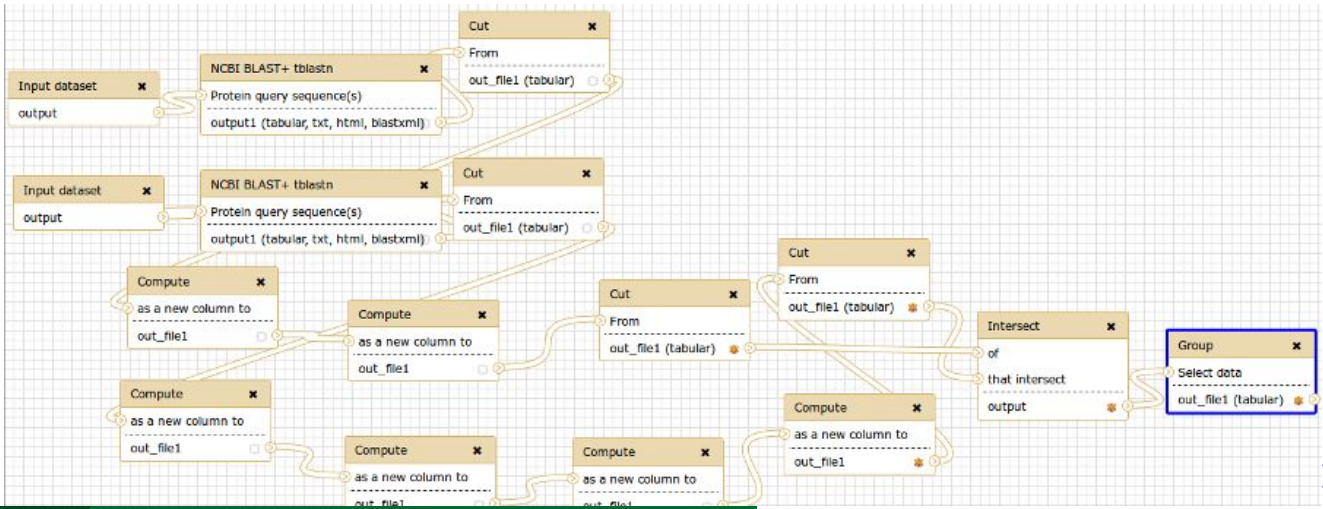
1 Leibniz Institute on Aging – Fritz Lipmann Institute

2 FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

TBI Winterseminar 2018

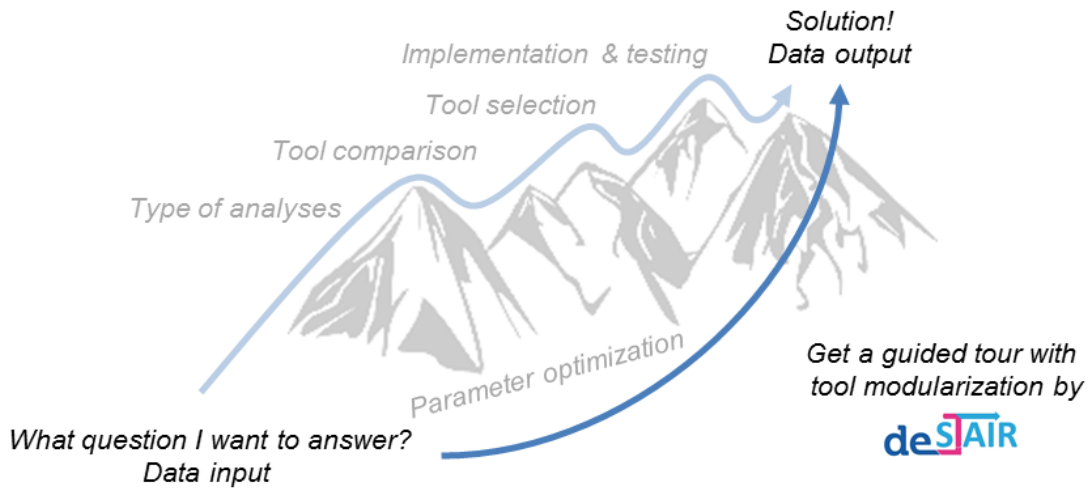
Motivation

- 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)



Motivation

- 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

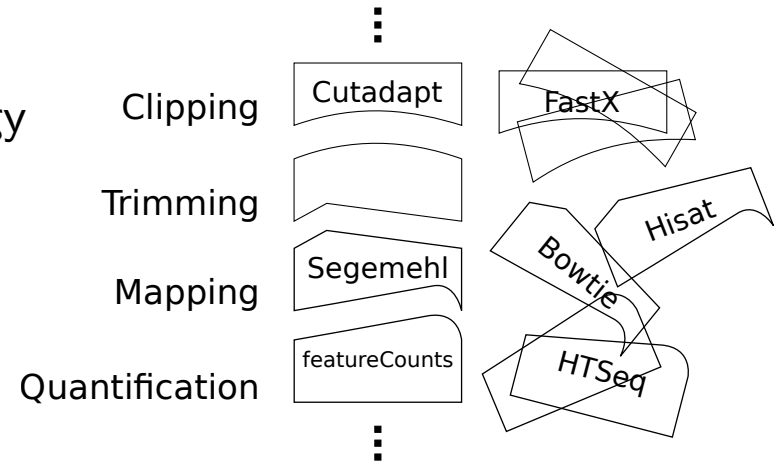


Guided construction of personalized workflows

Static workflows? No!

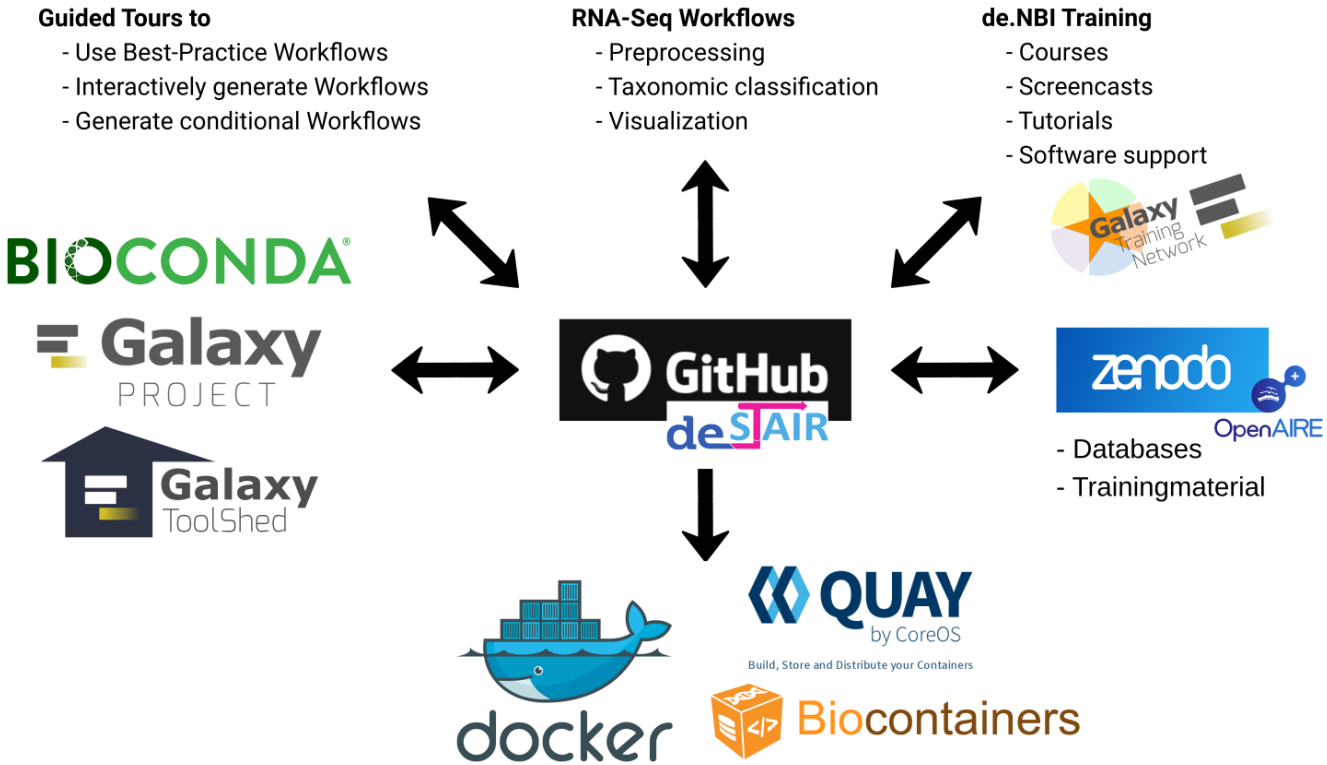
Given: Guided tours for workflows

- 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

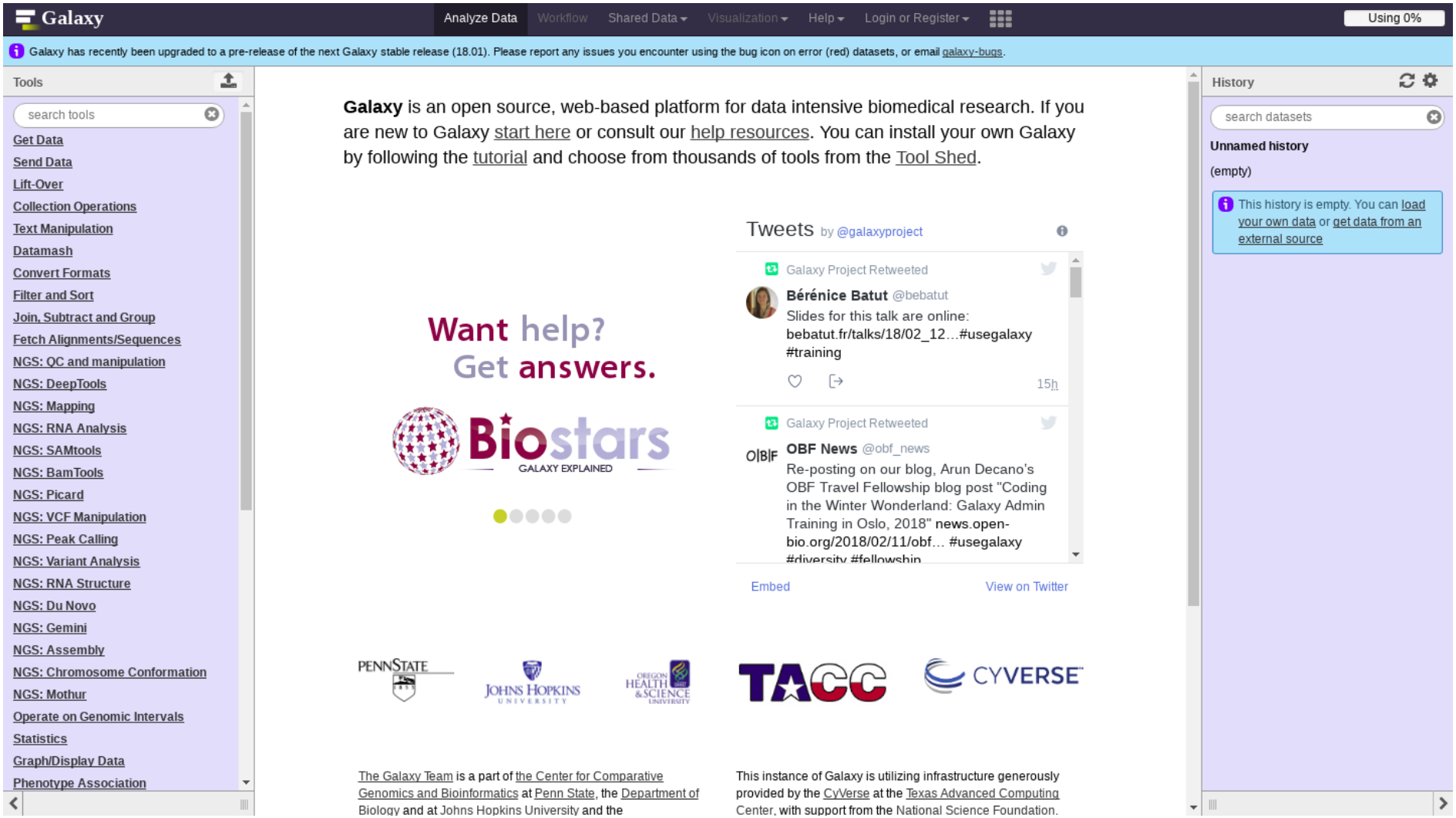


de.STAIR workspace

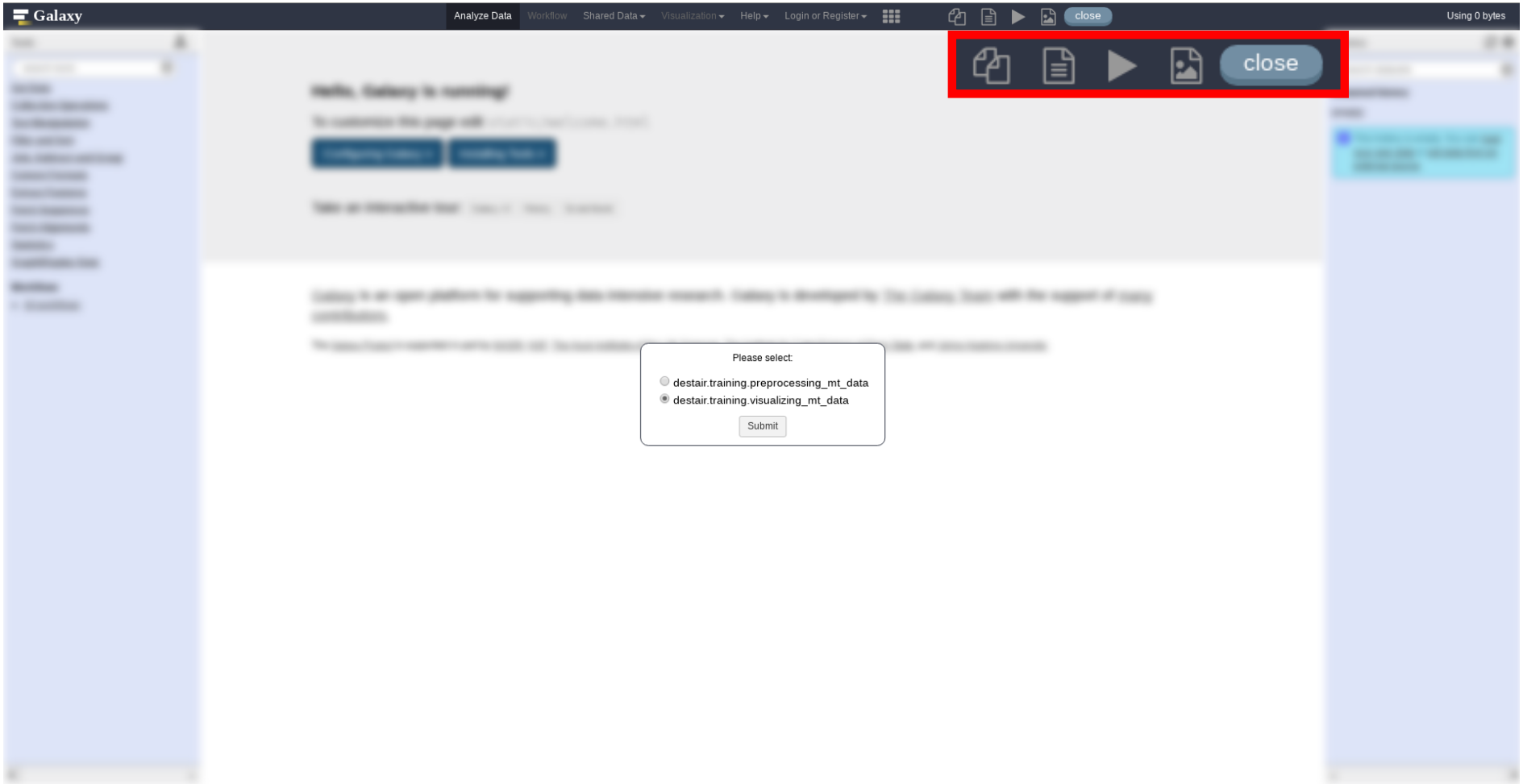
- Q/A system via webhooks
- Galaxy instance as webservice for workflow generation
- Export of workflows and CLI commands
- Use for training events



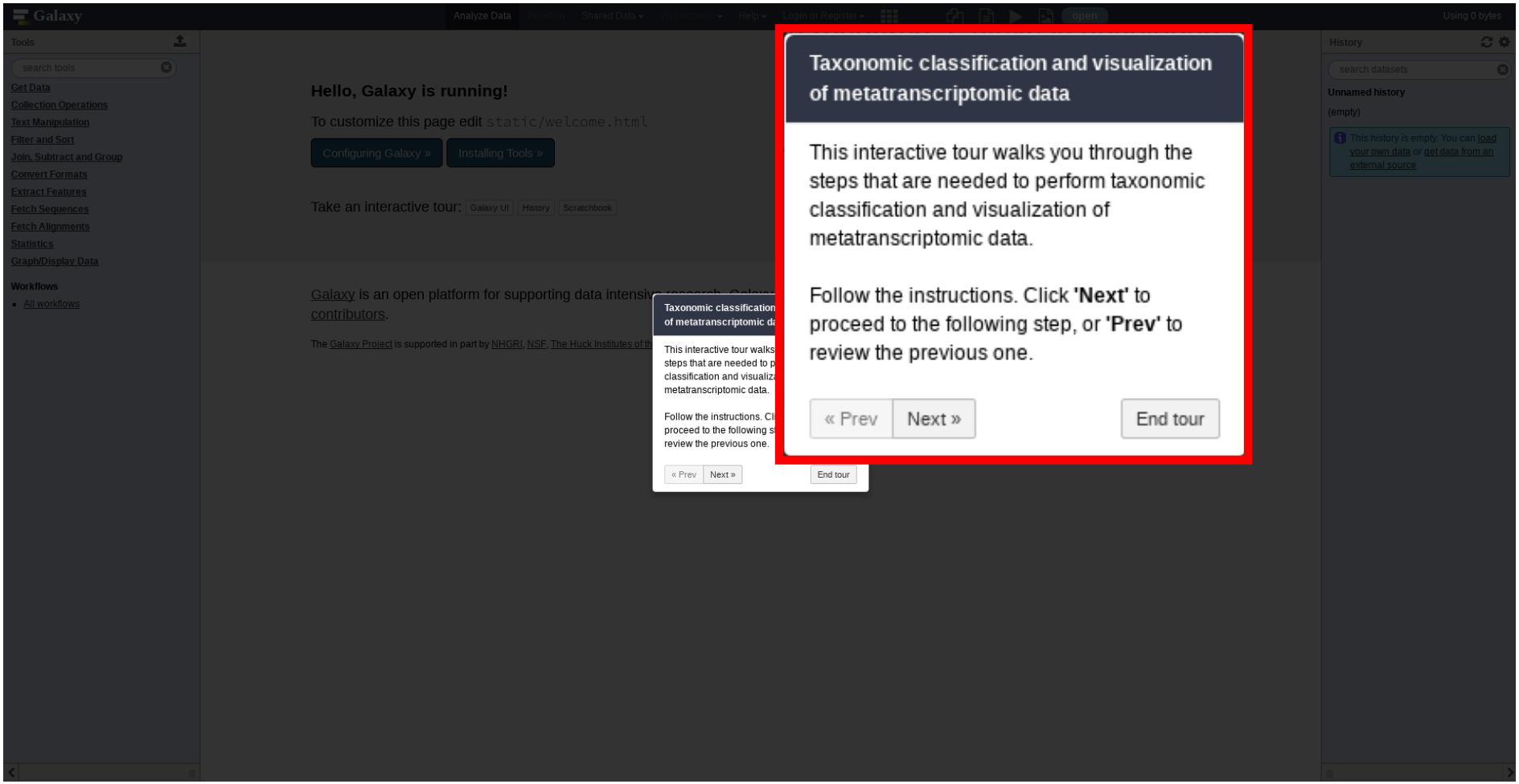
Proof of concept



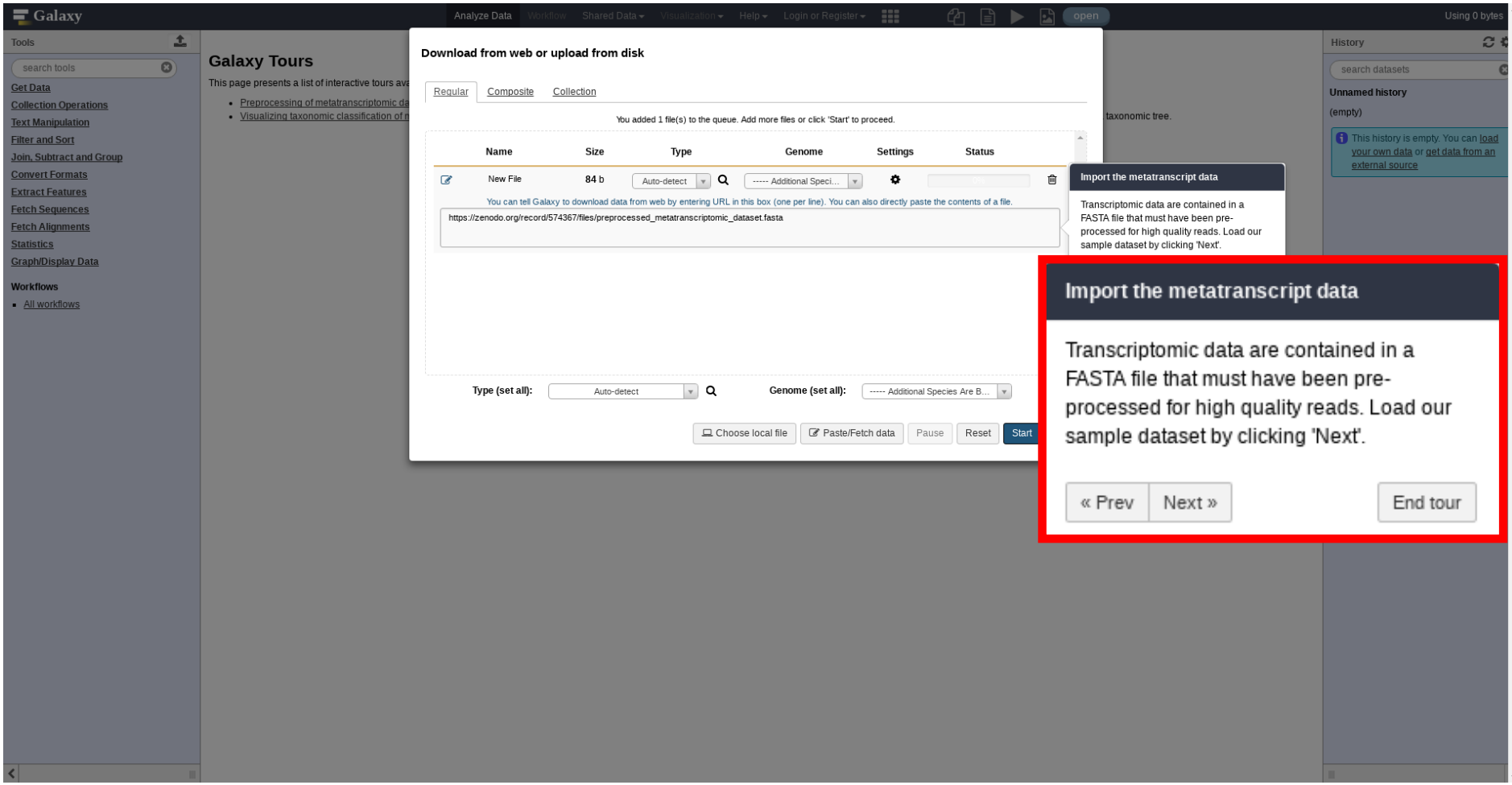
Proof of concept



Proof of concept

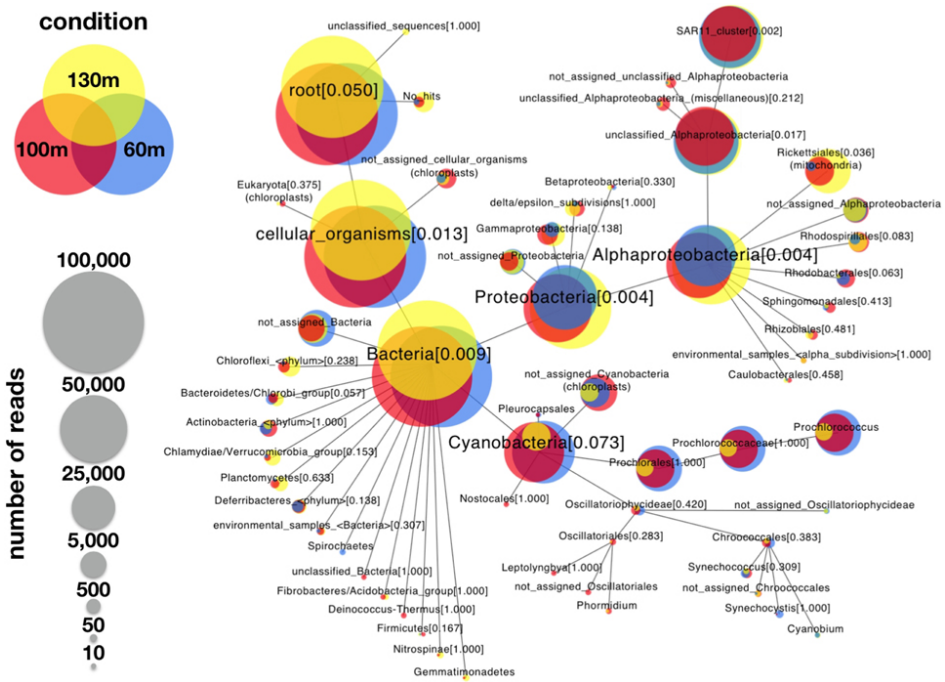


Proof of concept



Trainings

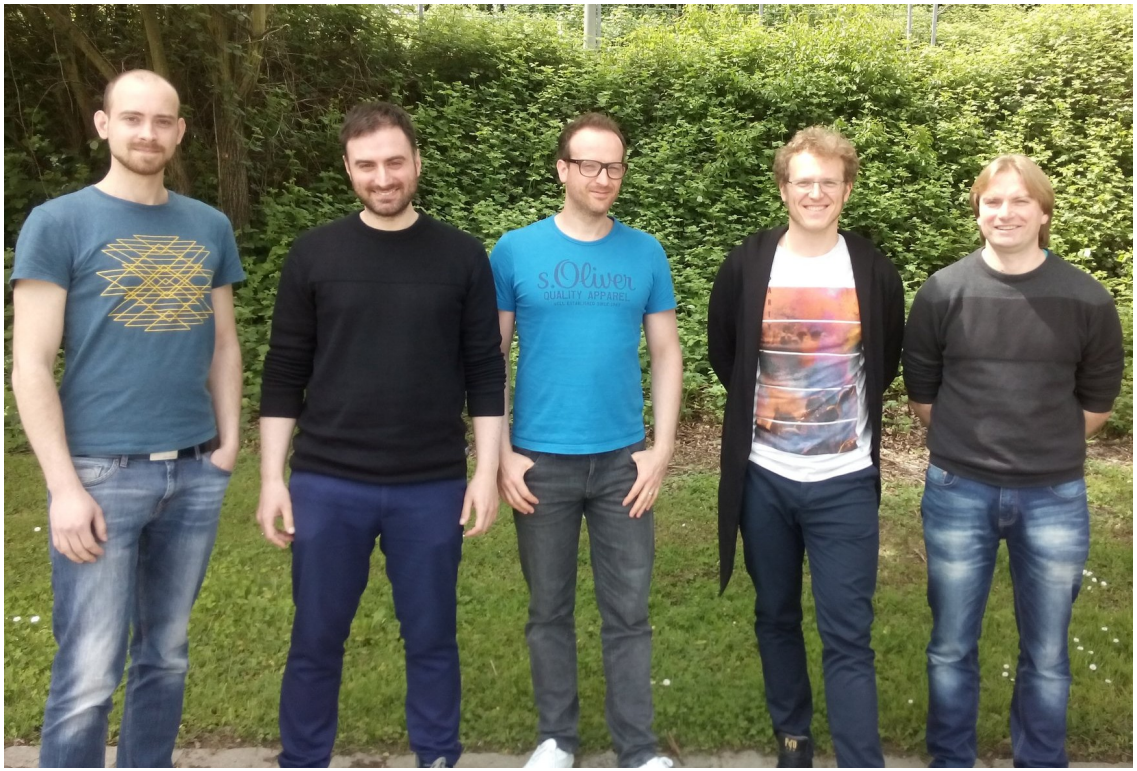
- Training Freiburg, 2017
- Training Jena, June 2018 - announced soon!
  - Pitfalls in RNA sequencing
  - Galaxy UI
  - Guided tours
  - Data processing
  - Taxonomic classification
  - CoVennTree visualization



Perspectives

- 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



- 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.