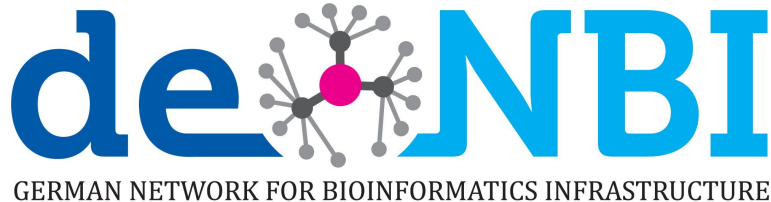


Galaxy Basics

Dilmurat Yusuf, Bora Uyar
Dan Munteanu, Ricardo Wurmus
BMSB Galaxy Server



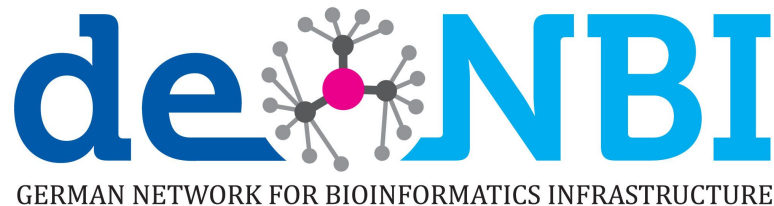
The format of the course

- Introduction of basics (Dilmurat Yusuf)
- Introduction of workflow (Bora Uyar)
- **Learning-by-doing**

BIMSB Galaxy Server

<https://galaxy.mdc-berlin.net/>

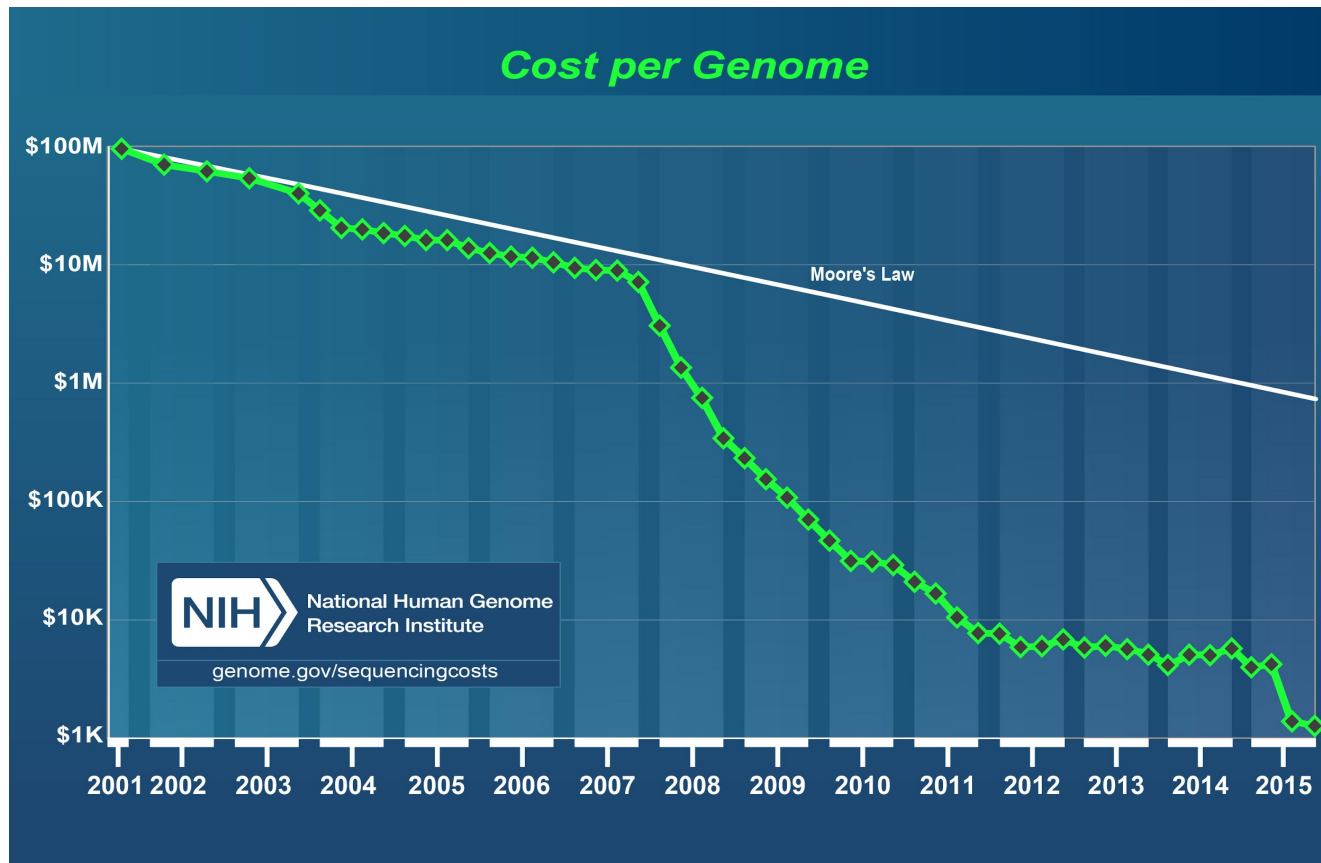
Contact: galaxy@mdc-berlin.de



Our roles

- Bioinformatics researchers
 - Part of BMSB bioinformatics platform
- Galaxy administrators
 - Maintenance
 - Training
 - Help desk
 - **Online forum (under development)**
 - **Dedicated walk-in hours**
 - Well... we are not the IT guys

Cheap data



Galaxy



Galaxy / CIBB

Workflow Canvas | Clone of "Send export to bam QA" shared by 'delhomme@cibb.de' | Options

Tools

- FASTA manipulation
- NGS: Format conversion
- NGS: Quality Assurance
- NGS: QC and manipulation
- NGS: Mapping
- NGS: RNA Analysis
- NGS: SAM Tools
- NGS: Peak Calling
- Transfer: Transfer a file to its final location, renaming it as part of the process
- Expand gzip compressed files
- Uses the gunzip command to expand gz files
- Testing tools

Workflow control

Items

- Input dataset

Workflow Canvas

Input Dataset → Expand gzip compressed files → QA Bundle → Export to fastq → Map with Bowtie for Illumina → SAM-to-BAM

Details

Tool: Export to fastq

Export file
Data tool: 'exportfile' (export)

Edit Step Actions

Remove Dataset
[fastqfile] Create

Add actions to this step; actions are applied when this workflow step completes.

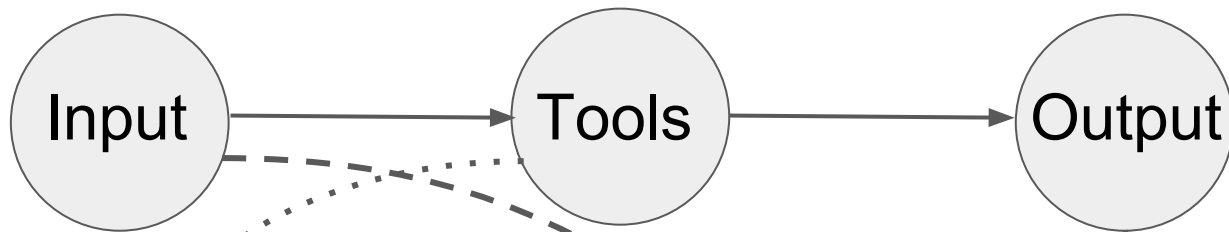
Edit Step Attributes

Annotation / Notes:

Add an annotation or notes to this step; annotations are available when a workflow is viewed.

Program: export_to_fastq (v0.1)
Author: Nicolas Delhomme (delhomme@cibb.de)
Summary: Converts export files to FASTQ format.
Usage: export_to_fastq.sh input_directory [threshold] [tag_file]
Note: Here the threshold is set on 0.

```
drwxr-xr-x. 2 root root 4096 May 18 16:03 opt
drwxr-xr-x. 2 root root 4096 May 18 16:03 preserve
drwxr-xr-x. 2 root root 4096 Jul 1 22:11 report
drwxrwxrwx. 1 root root 4096 May 14 00:12 run -> .../run
drwxr-xr-x. 14 root root 4096 May 18 16:08 swap
drwxrwxrwt. 4 root root 4096 Sep 12 23:50 [redacted]
drwxr-xr-x. 2 root root 4096 May 18 16:03 yp
[root@localhost var]# yum search wkli
Loaded plugins: langpacks, presto, refresh-packagekit, remove-with-leaves
rpmfusion-free-updates | 2.7 kB | 00:00
rpmfusion-free-updates/primary_db | 206 kB | 00:04
rpmfusion-nonfree-updates | 2.7 kB | 00:00
updates/metalink | 5.5 kB | 00:00
updates | 4.7 kB | 00:00
updates/primary_db | 73% [=====] | 62 kB/s | 2.6 MB | 00:15 ETA
```



The screenshot shows the Galaxy web interface. At the top, a navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. Below this is a green banner for the 'BIOINFORMATICS PLATFORM' provided by MDC and BIMS. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories such as 'Get Data', 'Text Manipulation', 'Statistics', and 'Workflows'. The main content area displays a message about the Galaxy project's support. On the right, a 'History' panel lists recent data collections and their associated tools, such as 'DESeq2 plots on data 62, data 69, and others' and 'TopHat on collection 23: splice junctions'. A large grey circle labeled 'Display' is positioned in the lower center of the interface.



Galaxy / Uni Freiburg Analyze Data Workflow Shared Data Visualization Help User Using 45.8 MB

Tools

search tools



- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Unix Tools
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Genomic Scores
- Operate on Genomic Intervals**
 - Concatenate two datasets into one dataset
 - Complement intervals of a dataset
 - Cluster the intervals of a dataset
 - Base Coverage of all intervals
 - Profile Annotations for a set of genomic intervals
 - Merge BED files (mergeBed)
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools

Concatenate (version 1.0.1)

Concatenate:  

8: Intersect on data 5 and data 7

First dataset

with:  

8: Intersect on data 5 and data 7

Second dataset

Both datasets are same filetype?: ☒

If unchecked Second dataset will be forced into format of First dataset

Execute

TIP: If your dataset does not appear in the pulldown menu -> it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.

Screencasts!

See Galaxy Interval Operation [Screencasts](#) (right click to open this link in another window).

Syntax

Both datasets are exactly the same filetype will preserve all extra fields in both files. Leaving this unchecked will force the second dataset to use the same column assignments for chrom, start, end and strand, but will fill extra fields with a period (.). In both cases, the output fields are truncated or padded to the fields type to obtain a fully tabular output.

Example

First dataset

Second dataset

Concatenated intervals

History

Galaxy Introduction - Part 1
35.5 MB

13: Charts on data 9

8: Intersect on data 5 and data 7

7: Get flanks on data 2

6: Convert Genomic Intervals To BED on data 5

5: Find and Replace on data 4

4: Find and Replace on data 3

3: awk on data 1

2: UCSC Main on Mouse: refGene (genome)

1: GSE37268_mof3_out.hpeak.txt

dataset

Upload data

Click category name to expand

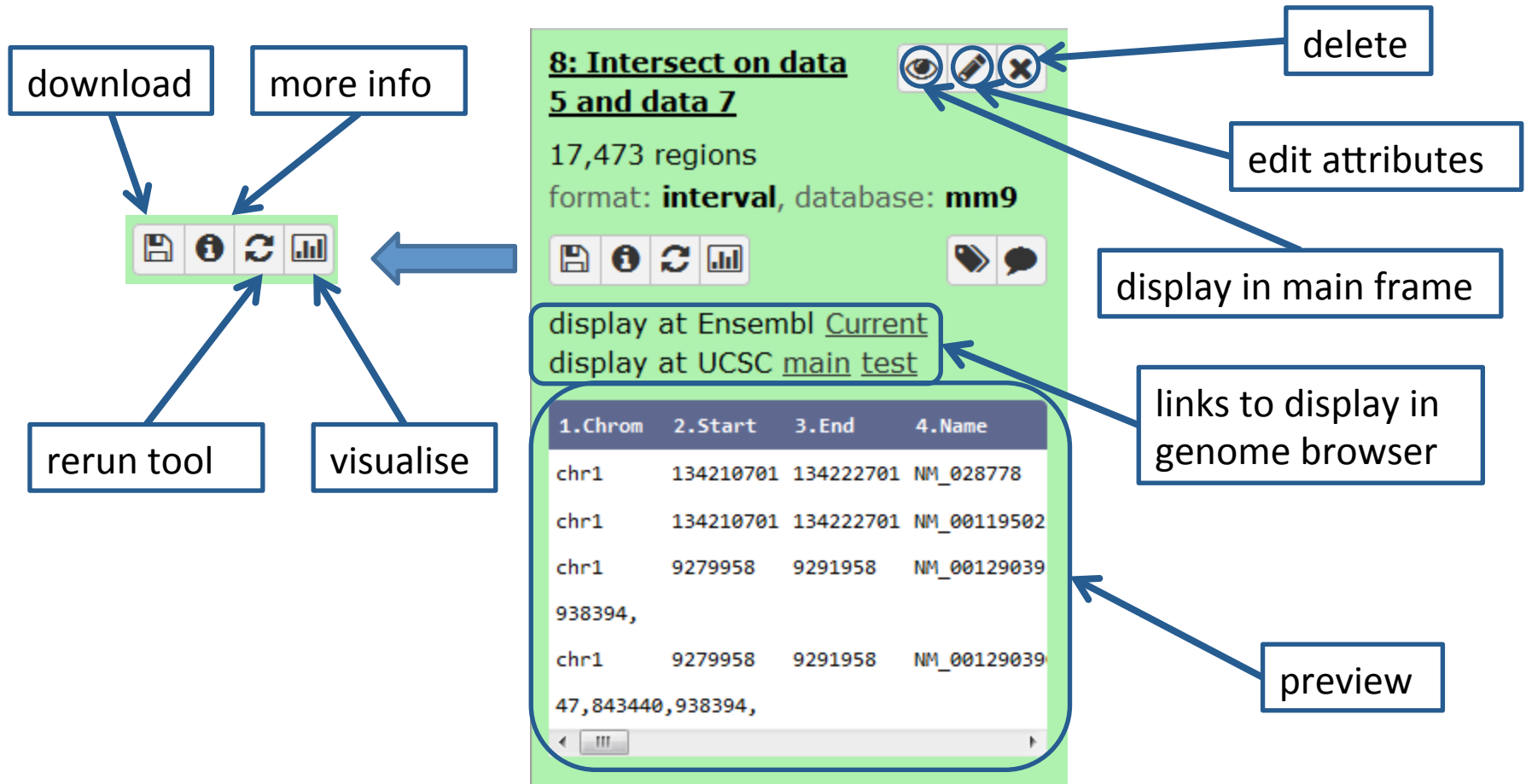
Click tool name to use

Tool options

Tool help

History options





Datasets









Click dataset name to expand





Dataset states

- Waiting to be run
- Running
- Successfully finished
- Failed

 **6: Compute on data 1**   

 **7: Compute on data 1**   




7: Compute on data 1   

 **9: Bar chart on data 8**   

error

An error occurred with this dataset:

```
Traceback (most recent call last):
  File "/usr/local/galaxy/galaxy-dist/tools/pl
import Gnuplot, Gnuplot.funcutils
ImportError: No module named Gnuplot
```

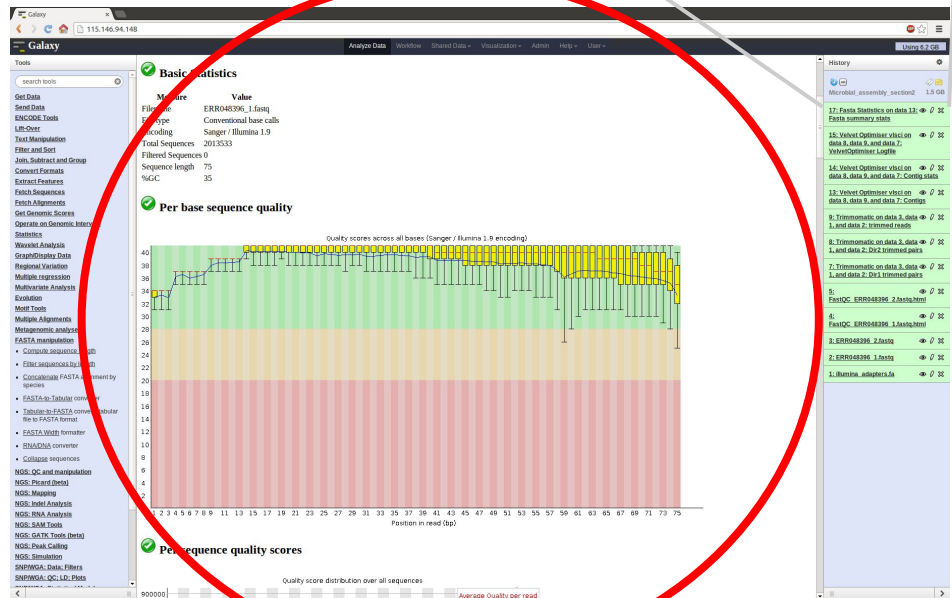
  

Send bug report

Viewing data / tool output

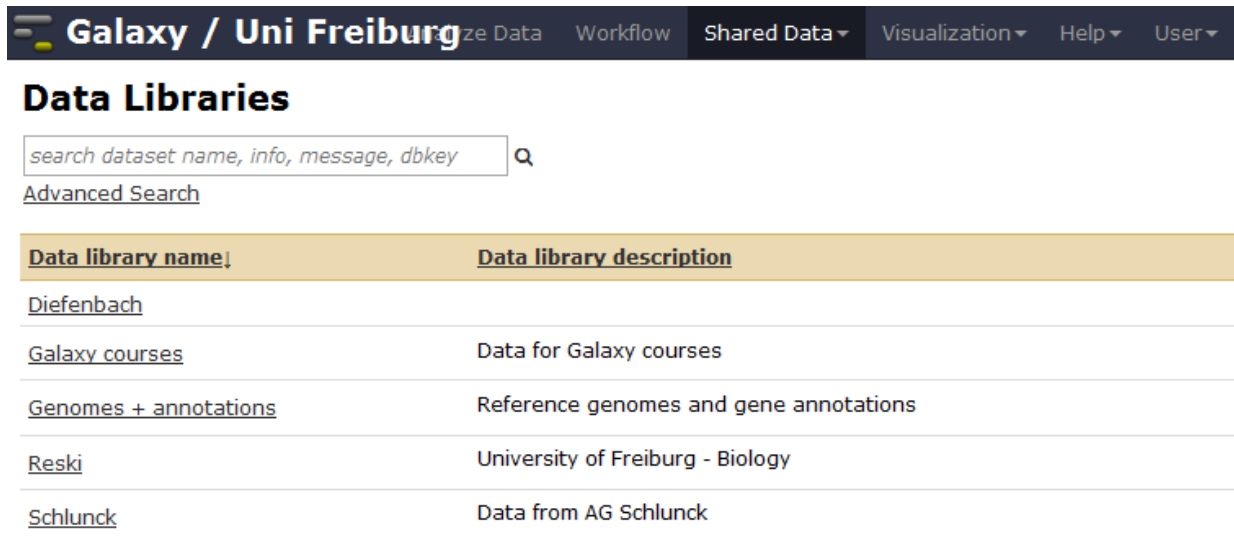
View data in centre

- Click on the “eye” symbol on a data file
- Its contents appear in the centre



Data Libraries

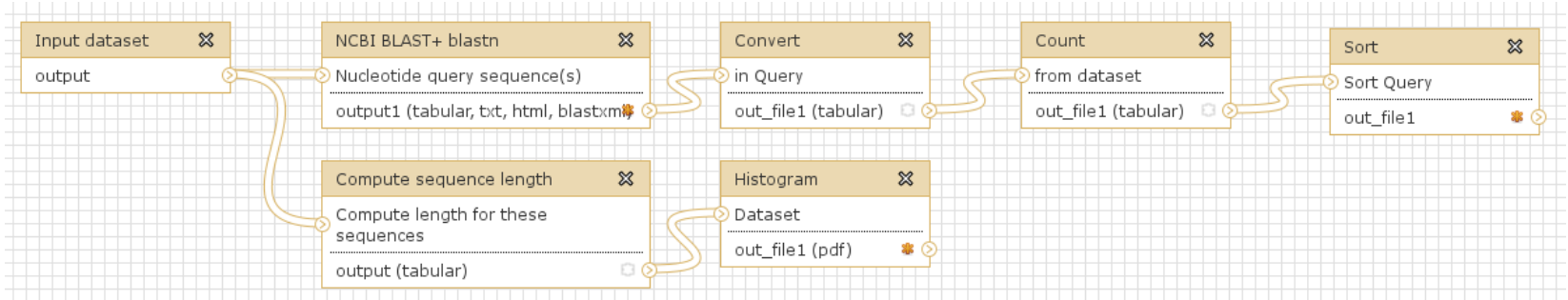
- Top menu -> Shared Data -> Data Libraries
- Access restricted by permissions
- Deposit for generally useful data
- Place for sharing



The screenshot shows the Galaxy / Uni Freiburg interface. At the top is a dark navigation bar with the logo and text 'Galaxy / Uni Freiburg' on the left, and a series of menu items: 'Analyze Data', 'Workflow', 'Shared Data' (which is highlighted with a dropdown arrow), 'Visualization', 'Help', and 'User'. Below the navigation bar, the page title 'Data Libraries' is displayed. Underneath the title is a search bar with the placeholder text 'search dataset name, info, message, dbkey' and a magnifying glass icon. Below the search bar is a link for 'Advanced Search'. The main content area features a table with two columns: 'Data library name' and 'Data library description'. The table lists five data libraries: 'Diefenbach', 'Galaxy courses', 'Genomes + annotations', 'Reski', and 'Schlunck', each with its corresponding description.

<u>Data library name</u>	<u>Data library description</u>
<u>Diefenbach</u>	
<u>Galaxy courses</u>	Data for Galaxy courses
<u>Genomes + annotations</u>	Reference genomes and gene annotations
<u>Reski</u>	University of Freiburg - Biology
<u>Schlunck</u>	Data from AG Schlunck

Workflows



- Automate repeating analysis
- Help to stay organized
- Share them
- Use other people's (e.g. ones we provide you)

Warnings

- Watch your disk usage!
- For some tools or tasks it's perfectly normal to run for a day or more
- You cannot upload data > 2 GB through your browser

