Galaxy in MetaCentrum

(introduction)

Pavel Fibich

MetaCentrum, CESNET

12-12-2014

CESNET

Galaxy, our instance

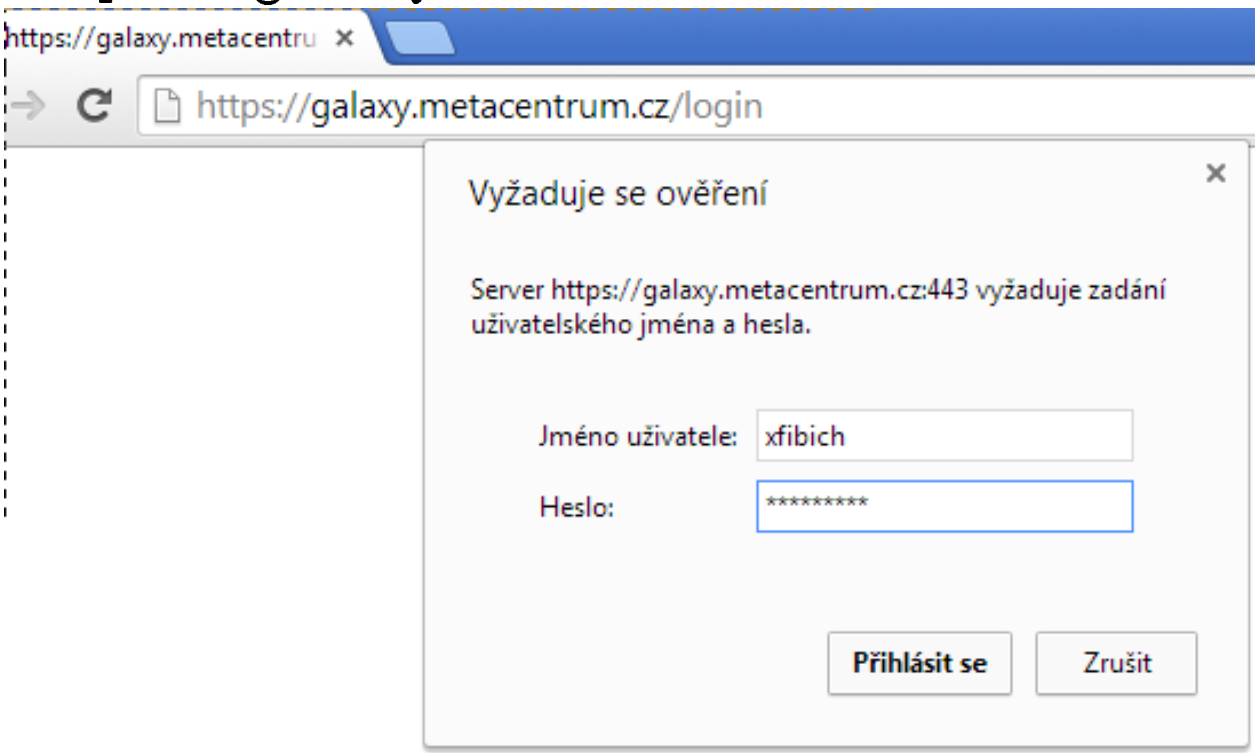
- open source scientific work-flow system
- web server based front-end for bioinformaticians (you can run computations through web browser)
- lightweight, modularly written in python
- many tools for bio data manupulations, managed by wrappers and allow pipe-lining (workflows)
- running analyses in the background
- primarily designed for single user and run in VM

MetaCentrum instance

- sends jobs to the grid under real user's accounts
- `https://wiki.metacentrum.cz/wiki/Galaxy_application`
- login: `https://galaxy.metacentrum.cz`

Login

HTTPs authentication; use std. MetaCentrum login/password for `https://galaxy.metacentrum.cz`



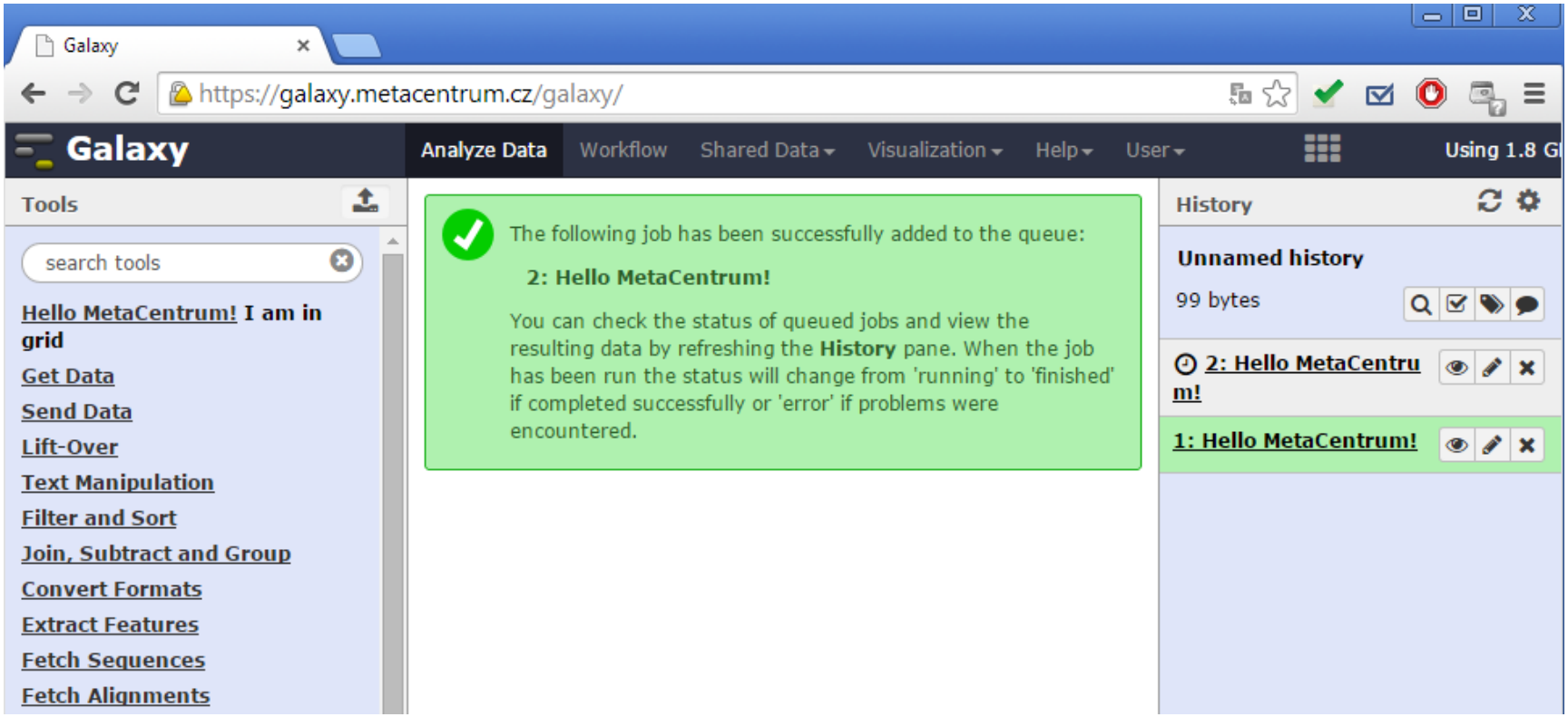
Use login page to renew user's tickets (allow access to the infrastructure) every day!

In Galaxy

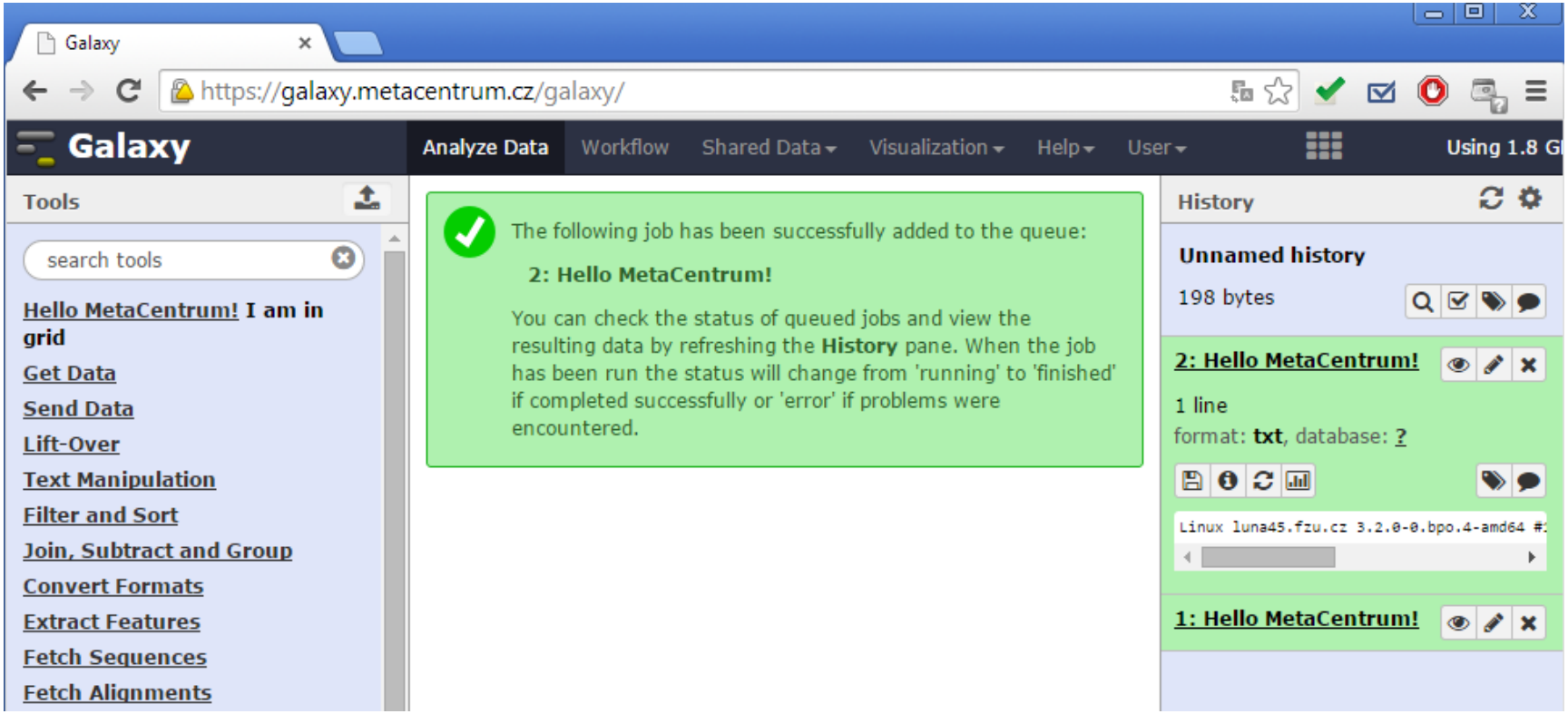
(tools menu) (body) (history)

Hello MetaCentrum!

Hello World example: Click on HelloMetaCentrum tool in the tools menu and then to Execute.



Hello MetaCentrum!



In the history: you can view, tag, download and edit attributes of the result(s), you can re-run the analysis too.

Uploading files

Uploading files:

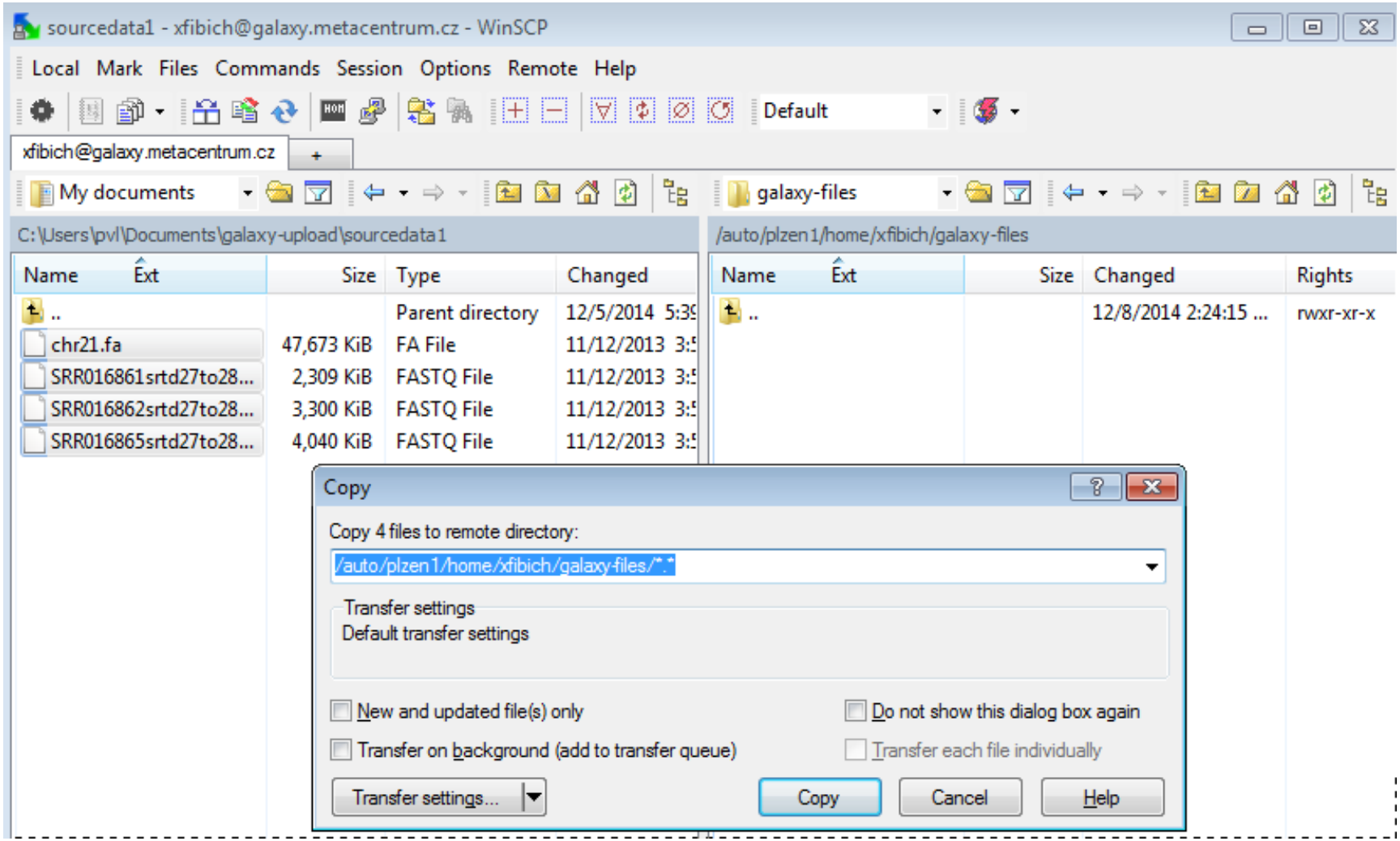
- directly in the tools menu: Get Data - Upload File
- large files are often uploaded by SFTP and later you see these files in the tools menu: Get Data - Upload File
  - connect to galaxy.metacentrum.cz machine by std. sftp protocol, eg. WinSCP program
  - use std. MetaCentrum user/password
  - copy your files to some directory and send the name of the directory, you will use for galaxy uploads, to the `meta@cesnet.cz`

Later you will see uploaded files in the history.

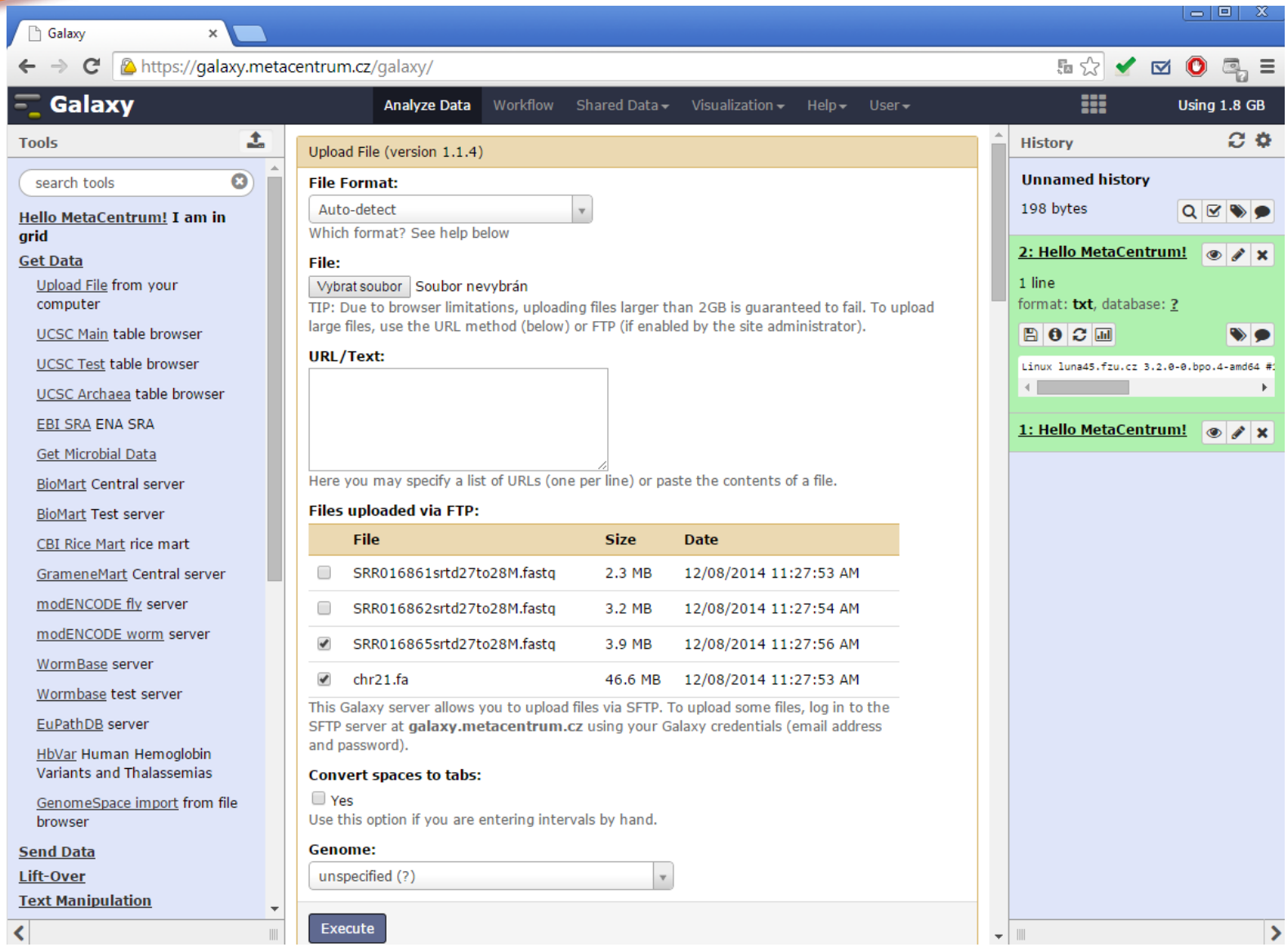
Example data: download and unzip for upload

`http://nihlibrary.ors.nih.gov/bioinfo/ngs/sourcedata1.zip`

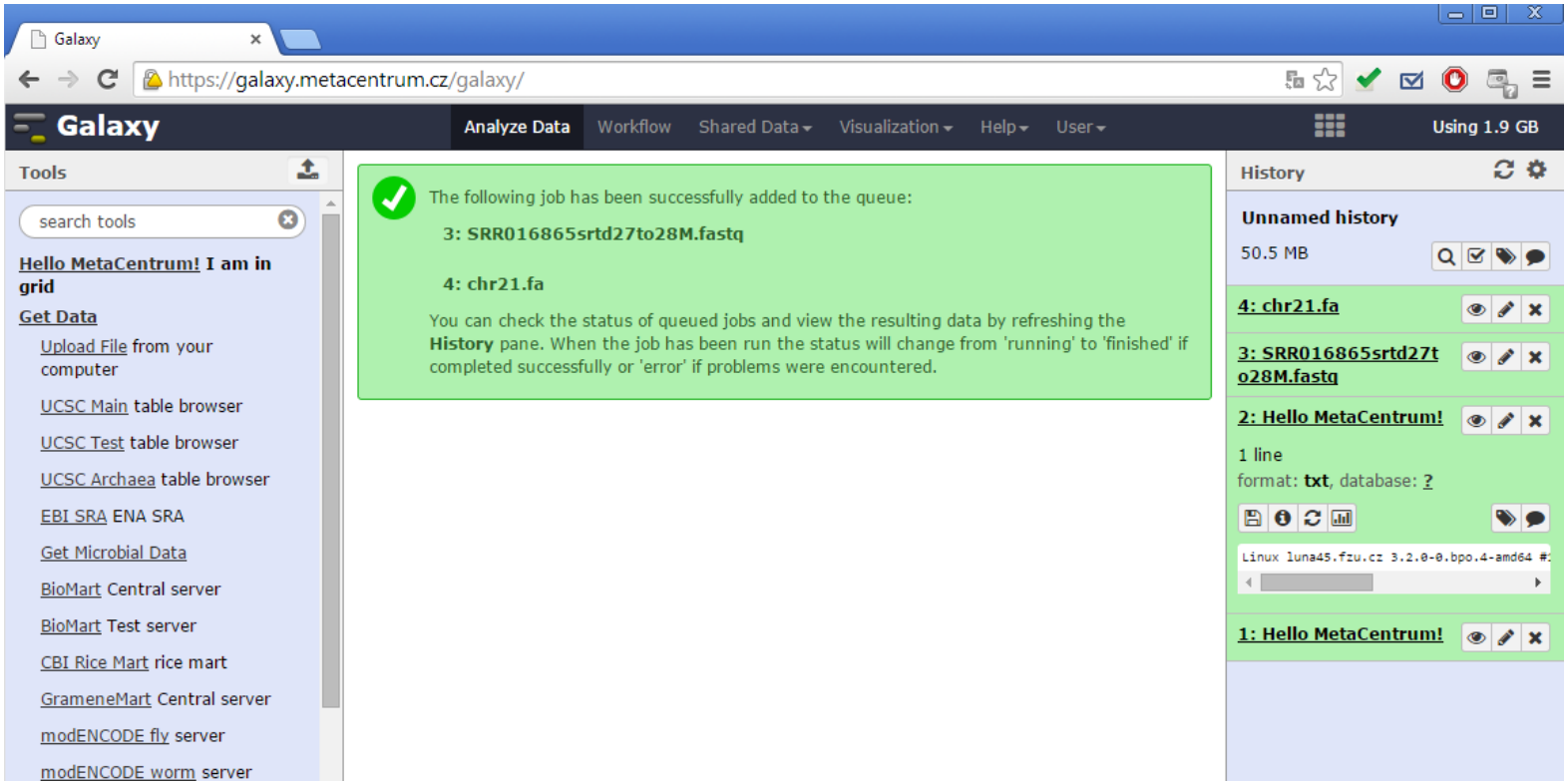
Uploading files SFTP



Uploading files SFTP



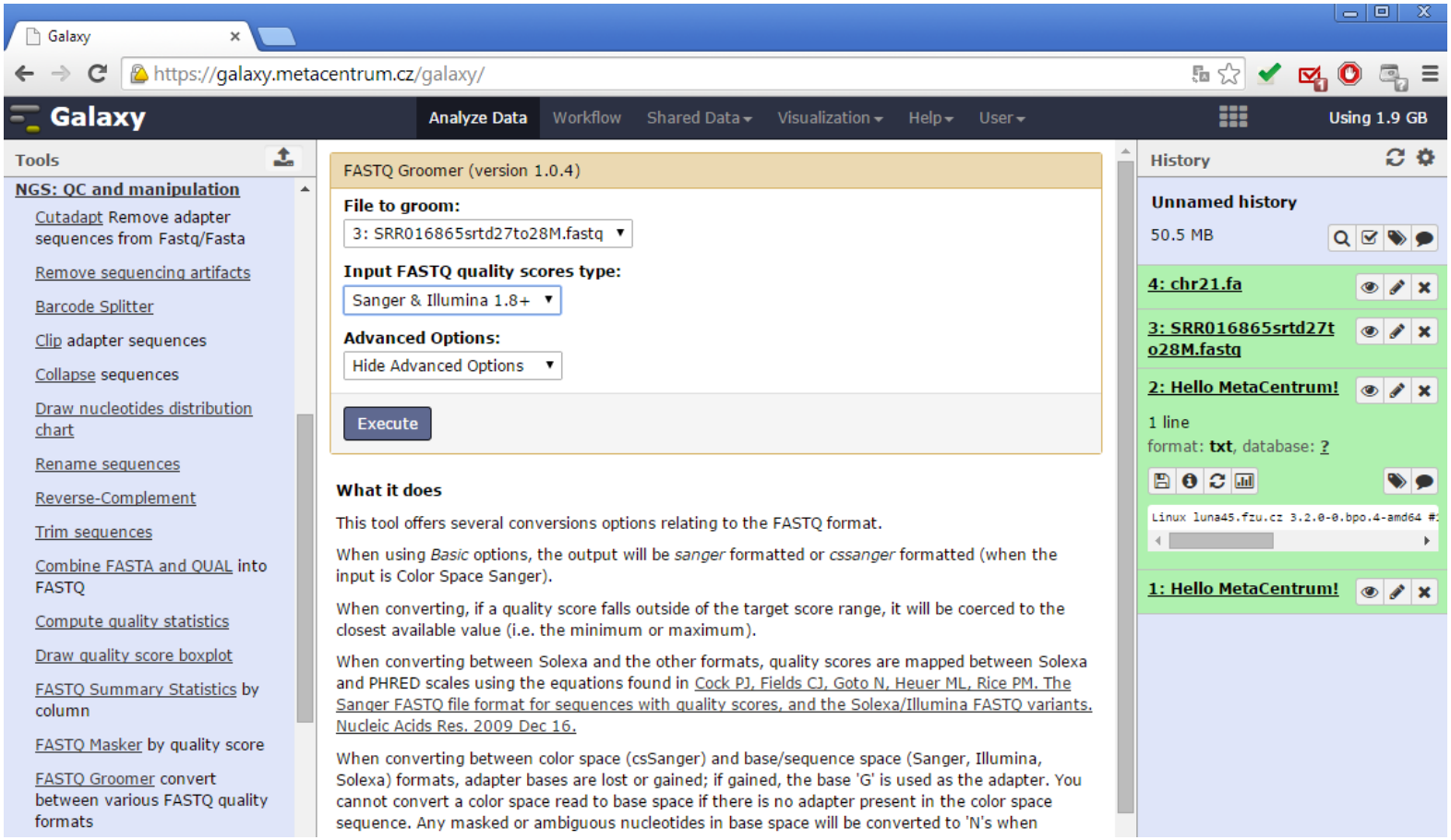
Uploading files SFTP



Files were successfully uploaded in the history.

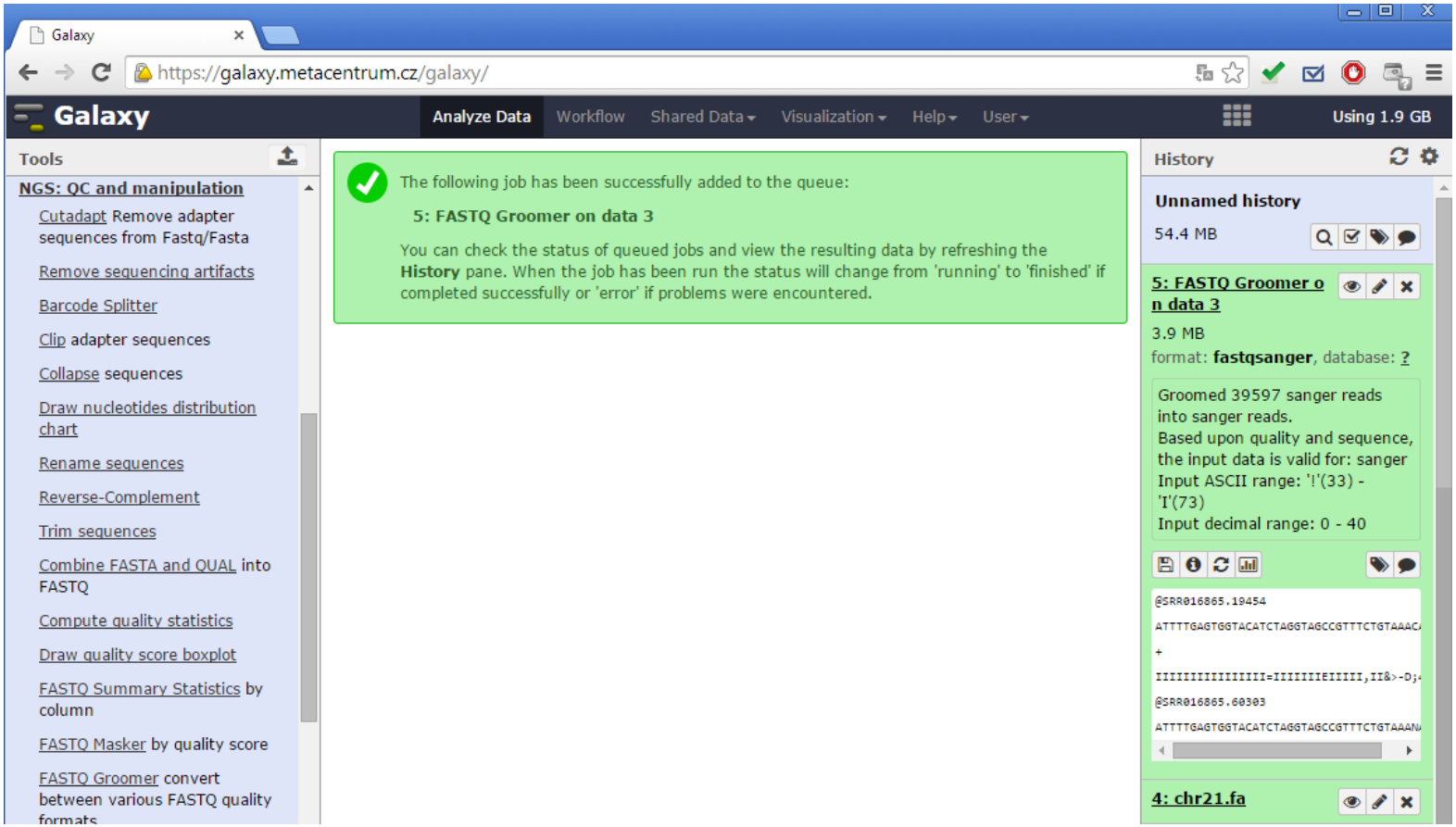
FASTQ Groomer

Click in the tools menu: NGS: QC and manipulation - FASTQ Groomer, choose fasq file and click on Execute.



If you need a tool, not in menu, send request to `meta@cesnet.cz`.

FASTQ Groomer



Analysis finished: download results, re-run analysis or use outputs for other analyses.

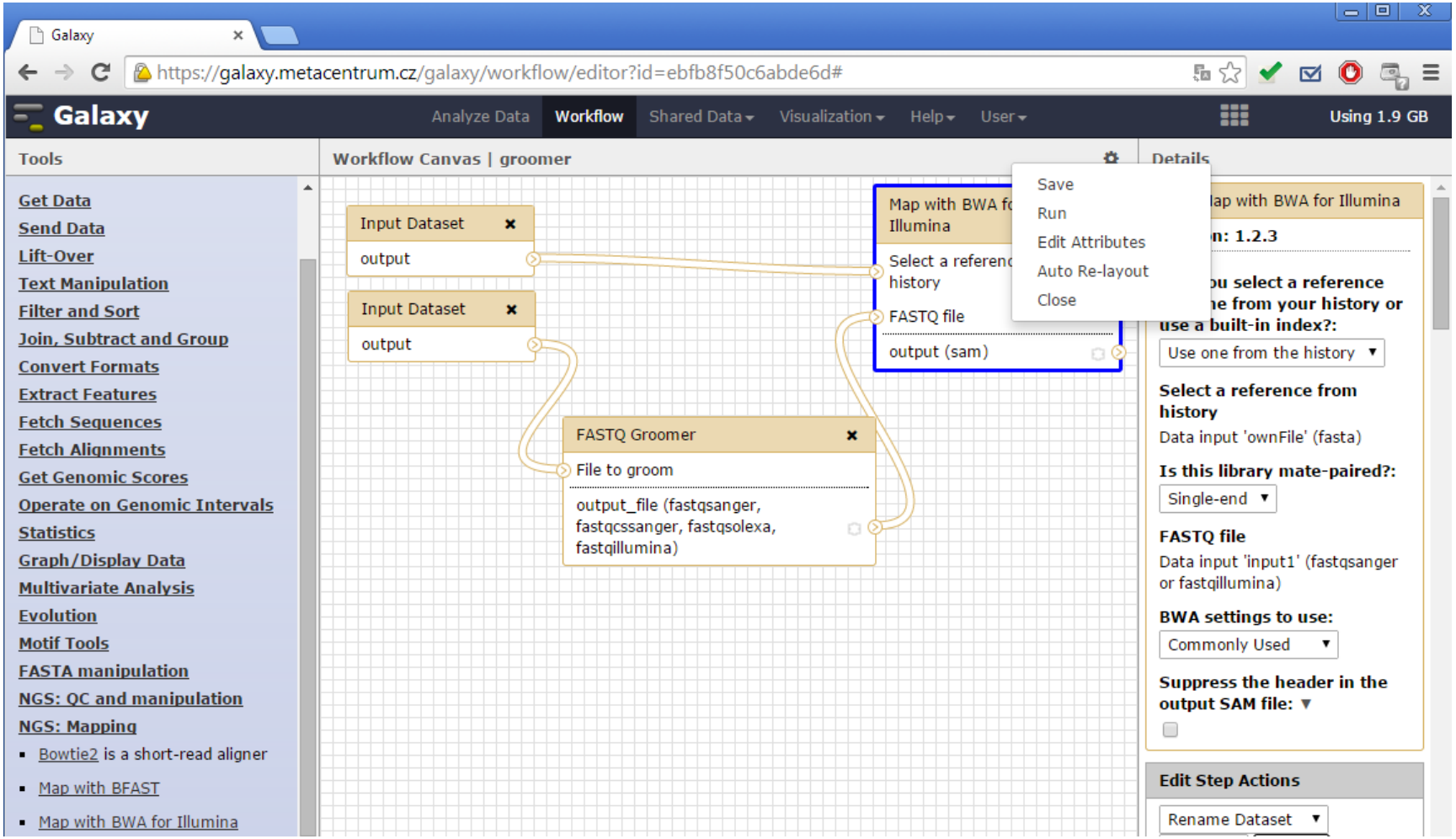
Workflows

You can easily join several analyses together (output of one as input of other)

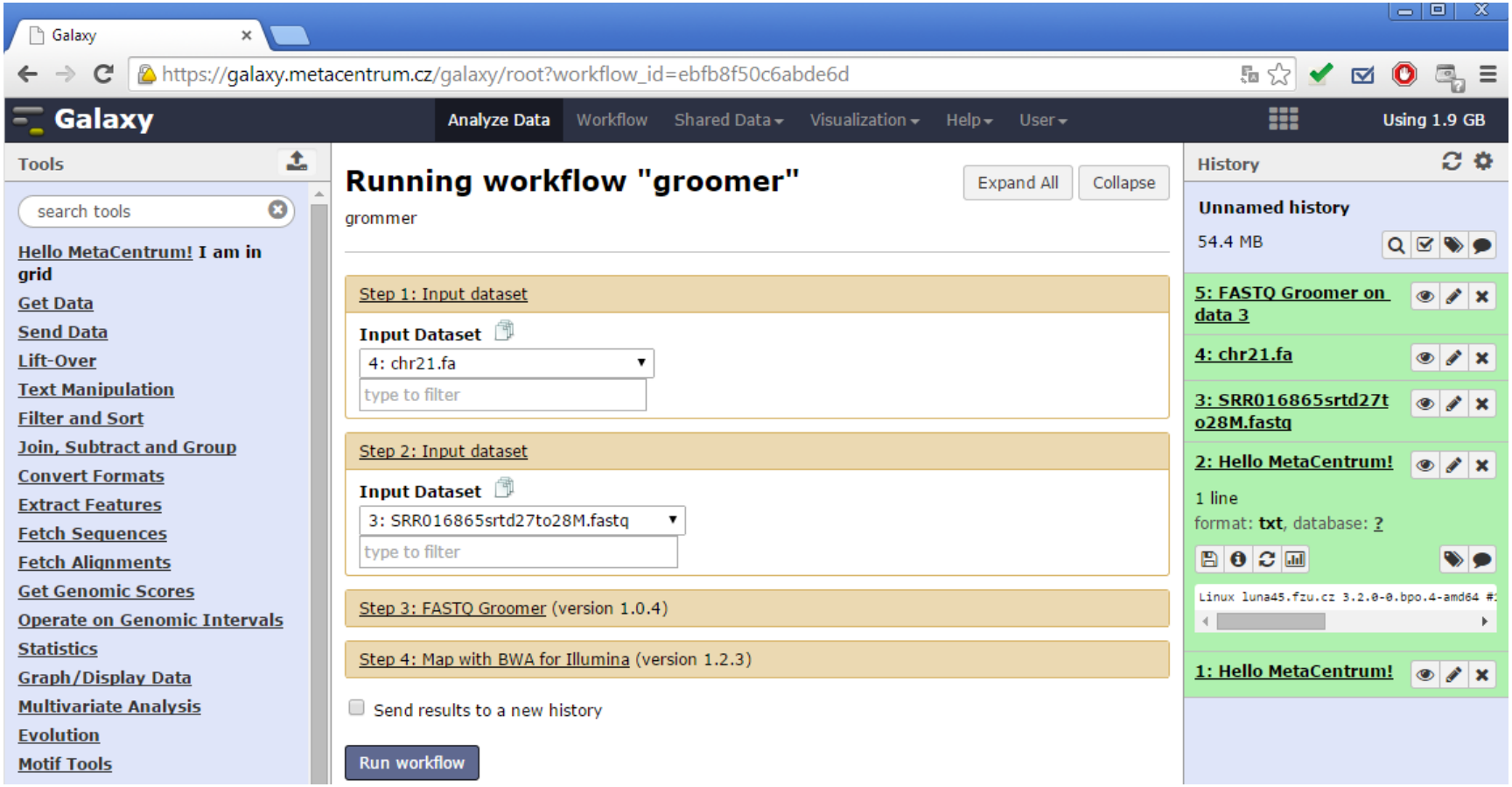
- click on Workflow (upper menu) - Create new workflow and fill some name, click on Create
- click on your new workflow and choose Edit
- add two Input Datasets, FASTQ Gromer (NGS: QC and manipulation) and Map with BWA for Illumina (NGS: Mapping) and connect them as it is on the following slide
- click on wheel and choose Save
- click on Workflow, choose your one and click Run
- choose input Datasets and click Run workflow

You can change inputs and re-run workflow or share/publish it.

Workflows


Galaxy
https://galaxy.metacentrum.cz/galaxy/workflow/editor?id=ebfb8f50c6abde6d#
Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 1.9 GB
Tools
Get Data
Send Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Graph/Display Data
Multivariate Analysis
Evolution
Motif Tools
FASTA manipulation
NGS: QC and manipulation
NGS: Mapping
Bowtie2 is a short-read aligner
Map with BFAST
Map with BWA for Illumina
Workflow Canvas | groomer
Input Dataset
output
Input Dataset
output
FASTQ Groomer
File to groom
output_file (fastqsanger, fastqcssanger, fastqsolexa, fastqillumina)
Map with BWA for Illumina
Select a reference history
FASTQ file
output (sam)
Save
Run
Edit Attributes
Auto Re-layout
Close
Details
Map with BWA for Illumina
n: 1.2.3
u select a reference
e from your history or
use a built-in index?:
Use one from the history
Select a reference from history
Data input 'ownFile' (fasta)
Is this library mate-paired?:
Single-end
FASTQ file
Data input 'input1' (fastqsanger or fastqillumina)
BWA settings to use:
Commonly Used
Suppress the header in the output SAM file:
Edit Step Actions
Rename Dataset

Workflows


Galaxy
https://galaxy.metacentrum.cz/galaxy/root?workflow_id=ebfb8f50c6abde6d
Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 1.9 GB
Tools
search tools
Hello MetaCentrum! I am in grid
Get Data
Send Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Graph/Display Data
Multivariate Analysis
Evolution
Motif Tools
Running workflow "groomer"
grommer
Expand All
Collapse
Step 1: Input dataset
Input Dataset
4: chr21.fa
type to filter
Step 2: Input dataset
Input Dataset
3: SRR016865srtd27to28M.fastq
type to filter
Step 3: FASTQ Groomer (version 1.0.4)
Step 4: Map with BWA for Illumina (version 1.2.3)
Send results to a new history
Run workflow
History
Unnamed history
54.4 MB
5: FASTQ Groomer on data 3
4: chr21.fa
3: SRR016865srtd27to28M.fastq
2: Hello MetaCentrum!
1 line
format: txt, database: ?
Linux luna45.fzu.cz 3.2.0-0.bpo.4-amd64 #
1: Hello MetaCentrum!

Final notes

- read
  `https://wiki.metacentrum.cz/wiki/Galaxy_application`
- always (every day) login through
  `https://galaxy.metacentrum.cz`
- clean (delete) history items if you will not use them anymore
- ask for help or report problem at the `meta@cesnet.cz`

Questions/comments?

`pavel.fibich@cesnet.cz`