

# Galaxy in MetaCentrum

(introduction)

Pavel Fibich

MetaCentrum, CESNET

12-12-2014



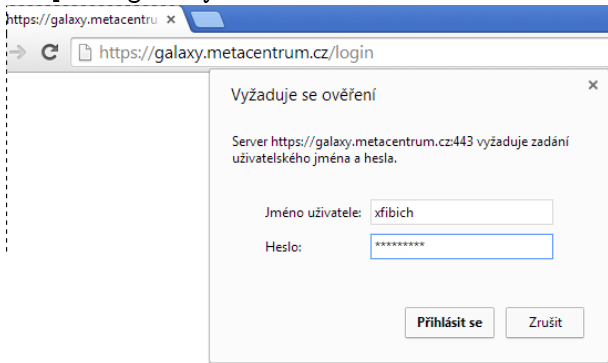


- ▶ open source scientific work-flow system
- ▶ web server based front-end for bioinformaticians (you can run computations through web browser)
- ▶ lightweight, modularly written in python
- ▶ many tools for bio data manipulations, managed by wrappers and allow pipe-lining (workflows)
- ▶ running analyses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

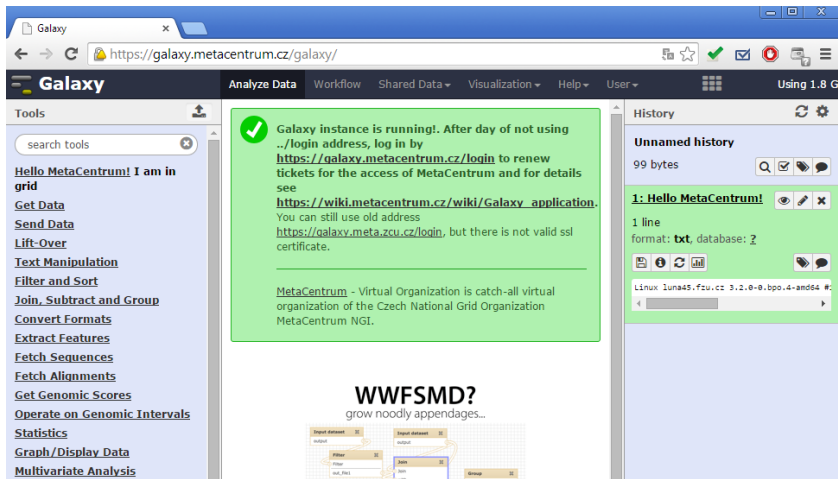
### MetaCentrum instance

- ▶ sends jobs to the grid under real user's accounts
- ▶ [https://wiki.metacentrum.cz/wiki/Galaxy\\_application](https://wiki.metacentrum.cz/wiki/Galaxy_application)
- ▶ **login:** <https://galaxy.metacentrum.cz>

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!**



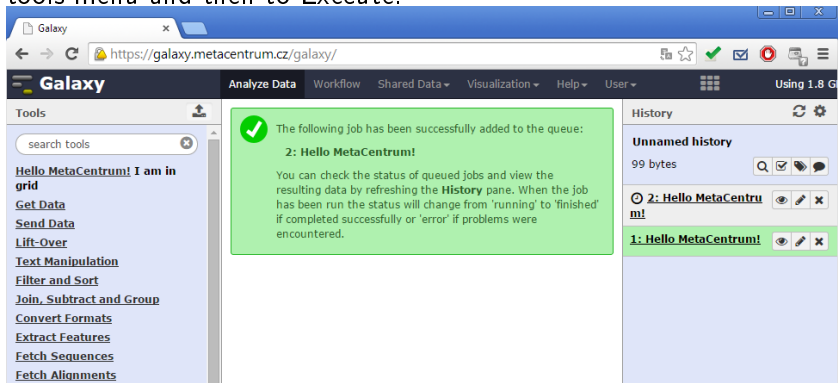
The screenshot shows the Galaxy web interface. The browser address bar displays `https://galaxy.metacentrum.cz/galaxy/`. The interface includes a top navigation bar with 'Galaxy' and 'Using 1.8 G'. A left sidebar contains a 'Tools' menu with a search box and various tool categories. The main content area (body) features a green status message with a checkmark icon, indicating the Galaxy instance is running and providing instructions for login and ticket renewal. Below the message is a diagram titled 'WWFSMD?' with the subtitle 'grow noodly appendages...' and a workflow diagram showing data flow between 'Input dataset', 'Filter', 'Join', and 'Group' tools. The right sidebar shows a 'History' panel with an 'Unnamed history' entry, 99 bytes, and a single line of output: 'format: txt, database: 2'. Below the history entry, the Linux command `Linux luna45.fzu.cz 3.2.0-0.bpo.4-amd64 #:` is visible.

(tools menu)

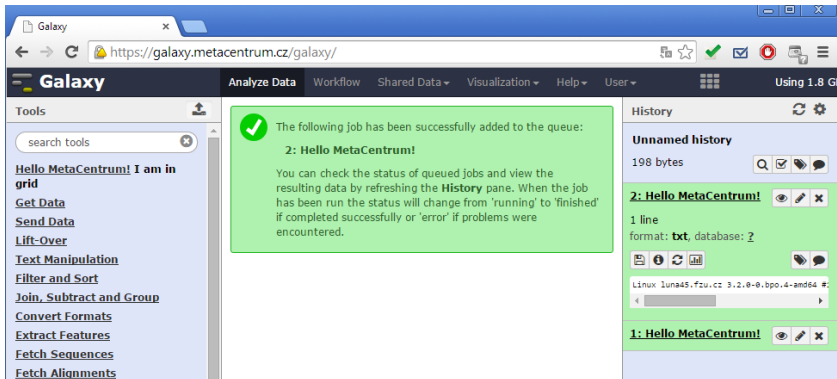
(body)

(history)

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



The screenshot shows the Galaxy web interface in a browser window. The address bar displays `https://galaxy.metacentrum.cz/galaxy/`. The main navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A 'Tools' sidebar on the left lists various tools, with 'Hello MetaCentrum! I am in grid' at the top. A central green notification box contains a checkmark and the text: 'The following job has been successfully added to the queue: 2: Hello MetaCentrum! You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' The 'History' pane on the right shows a list of jobs, with '2: Hello MetaCentrum!' and '1: Hello MetaCentrum!' visible.



The screenshot shows the Galaxy web interface in a browser window. The address bar displays `https://galaxy.metacentrum.cz/galaxy/`. The main navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. On the left, a 'Tools' sidebar lists various analysis tools. A central green notification box with a checkmark icon states: 'The following job has been successfully added to the queue: 2: Hello MetaCentrum! You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' On the right, the 'History' pane shows an 'Unnamed history' with 198 bytes. Below it, two job entries are visible: '2: Hello MetaCentrum!' and '1: Hello MetaCentrum!'. Each job entry includes icons for viewing, tagging, downloading, and editing.

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

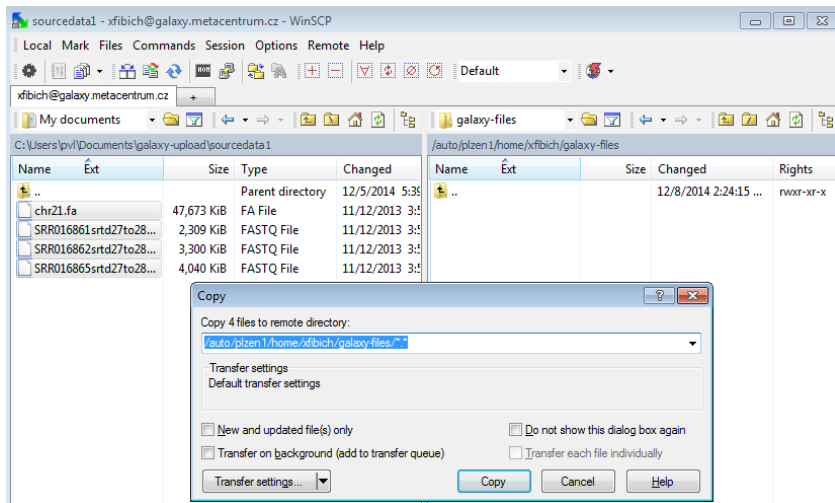
## Uploading files:

- ▶ 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>



sourcedata1 - xfibich@galaxy.metacentrum.cz - WinSCP

Local Mark Files Commands Session Options Remote Help

xfibich@galaxy.metacentrum.cz

My documents galaxy-files

C:\Users\pvl\Documents\galaxy-upload\sourcedata1

Name	Ext	Size	Type	Changed
..			Parent directory	12/5/2014 5:35
chr21.fa		47,673 KiB	FA File	11/12/2013 3:5
SRR016861srtid27to28...		2,309 KiB	FASTQ File	11/12/2013 3:5
SRR016862srtid27to28...		3,300 KiB	FASTQ File	11/12/2013 3:5
SRR016865srtid27to28...		4,040 KiB	FASTQ File	11/12/2013 3:5

/auto/plzen1/home/xfibich/galaxy-files

Name	Ext	Size	Changed	Rights
..			12/8/2014 2:24:15 ...	rwxr-xr-x

Copy

Copy 4 files to remote directory:

/auto/plzen1/home/xfibich/galaxy-files/

Transfer settings  
Default transfer settings

New and updated file(s) only

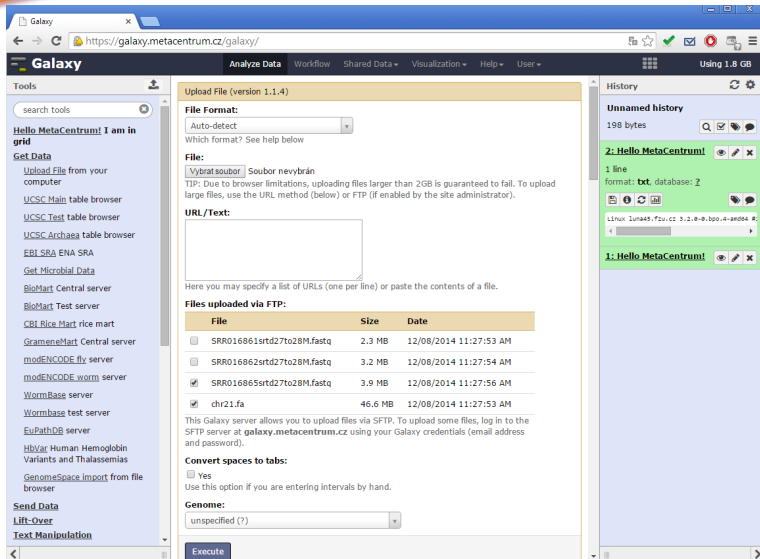
Transfer on background (add to transfer queue)

Do not show this dialog box again

Transfer each file individually

Transfer settings... Copy Cancel Help





The screenshot shows the Galaxy web interface at <https://galaxy.metacentrum.cz/galaxy/>. The main content area displays the 'Upload File (version 1.1.4)' tool configuration page.

**File Format:** Auto-detect (dropdown menu)

Which format? See help below

**File:** Vybrat soubor | Soubor nevybrán

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

**URL/Text:** [Empty text area]

Here you may specify a list of URLs (one per line) or paste the contents of a file.

**Files uploaded via FTP:**

File	Size	Date
<input type="checkbox"/> SRR016861srtid27to28M.fastq	2.3 MB	12/08/2014 11:27:53 AM
<input type="checkbox"/> SRR016862srtid27to28M.fastq	3.2 MB	12/08/2014 11:27:54 AM
<input checked="" type="checkbox"/> SRR016865srtid27to28M.fastq	3.9 MB	12/08/2014 11:27:56 AM
<input checked="" type="checkbox"/> chr21.fa	46.6 MB	12/08/2014 11:27:53 AM

This Galaxy server allows you to upload files via SFTP. To upload some files, log in to the SFTP server at [galaxy.metacentrum.cz](https://galaxy.metacentrum.cz) using your Galaxy credentials (email address and password).

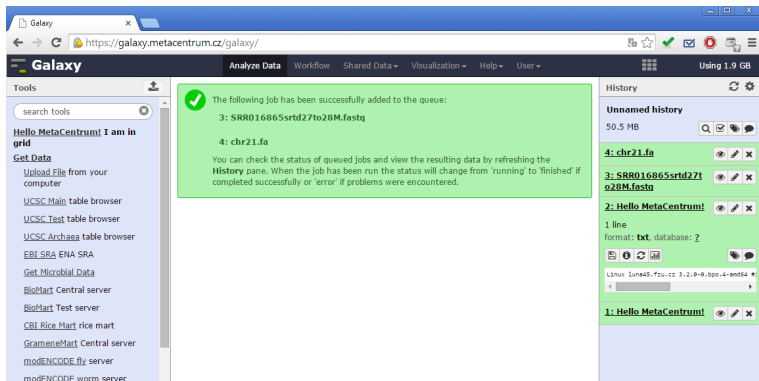
**Convert spaces to tabs:**  Yes  
Use this option if you are entering intervals by hand.

**Genome:** unspecified (?) (dropdown menu)

[Execute button]

The right sidebar shows the 'History' panel with 'Unnamed history' containing 198 bytes. The history list includes:

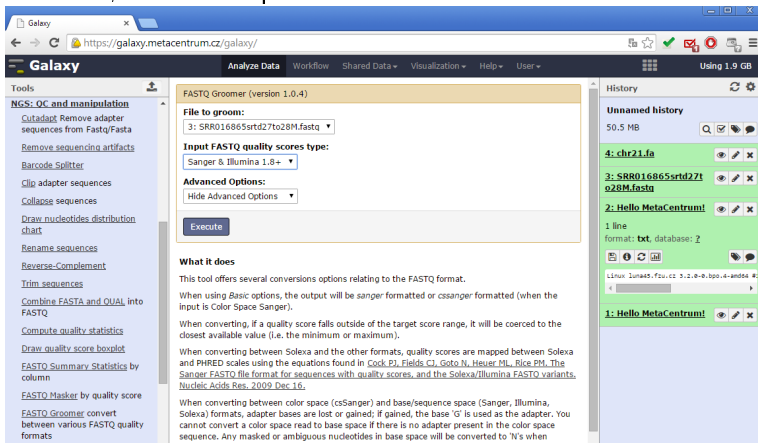
- 2: Hello MetaCentrum! (1 line, format: txt, database: Z)
- 1: Hello MetaCentrum!



The screenshot shows the Galaxy web interface. A green notification box in the center states: "The following job has been successfully added to the queue: 3: SRR016865srtcd27to28M.fastq 4: chr21.fa". Below this, it provides instructions on how to check the status of the job in the History pane. On the right, the History panel shows an "Unnamed history" containing three items: "4: chr21.fa", "3: SRR016865srtcd27to28M.fastq", and "2: Hello MetaCentrum!". Each item has a green background and icons for viewing, editing, and deleting. The interface also shows a "Tools" sidebar on the left and a navigation bar at the top.

Files were successfully uploaded in the history.

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



The screenshot shows the Galaxy web interface. The main panel displays the **FASTQ Groomer (version 1.0.4)** tool configuration. The **File to groom:** field is set to `3: SRR016865srttd27to28M.fastq`. The **Input FASTQ quality scores type:** is set to `Sanger & Illumina 1.8+`. The **Advanced Options:** dropdown is set to `Hide Advanced Options`. An **Execute** button is visible at the bottom of the tool panel.

The **History** panel on the right shows the execution history. The top entry is **4: chr21.fa** (50.5 MB). Below it is the current job: **3: SRR016865srttd27to28M.fastq** (format: `txt, database: 2`). Below that is **2: Hello MetaCentrum!** (1 line, format: `txt, database: 2`). The bottom entry is **1: Hello MetaCentrum!**.

**What it does**

This tool offers several conversions options relating to the FASTQ format.

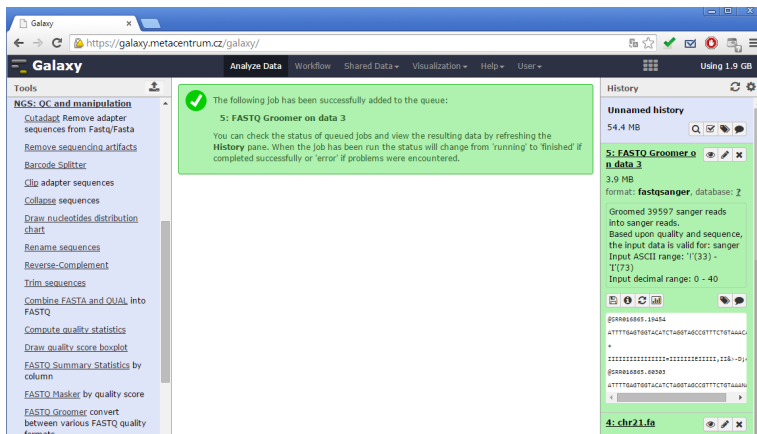
When using **Basic** options, the output will be `sanger` formatted or `cssanger` formatted (when the input is Color Space Sanger).

When converting, if a quality score falls outside of the target score range, it will be coerced to the closest available value (i.e. the minimum or maximum).

When converting between Solexa and the other formats, quality scores are mapped between Solexa and PHRED scales using the equations found in [Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2009 Dec 16.](#)

When converting between color space (csSanger) and base/sequence space (Sanger, Illumina, Solexa) formats, adapter bases are lost or gained; if gained, the base 'G' is used as the adapter. You cannot convert a color space read to base space if there is no adapter present in the color space sequence. Any masked or ambiguous nucleotides in base space will be converted to 'N's when

If you need a tool, not in menu, send request to [meta@cesnet.cz](mailto:meta@cesnet.cz).



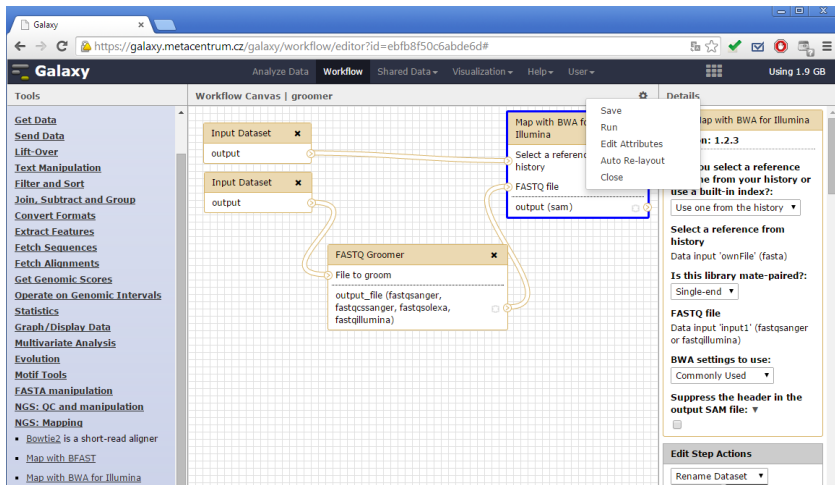
The screenshot shows the Galaxy web interface. A green notification box in the center states: "The following job has been successfully added to the queue: 5: FASTQ Groomer on data 3. You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." The left sidebar lists tools under "NGS: QC and manipulation", including "FASTQ Groomer". The right sidebar shows the "History" panel with a job entry: "5: FASTQ Groomer on data 3" (3.9 MB). The job details indicate it groomed 39597 sanger reads into sanger reads, based on quality and sequence, with valid input data for sanger. The input ASCII range is '!(33) - T!(73)' and the input decimal range is 0 - 40. Below the details, a preview of sanger reads is shown, including headers like "@GRR016865.18454" and sequence lines with quality scores.

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

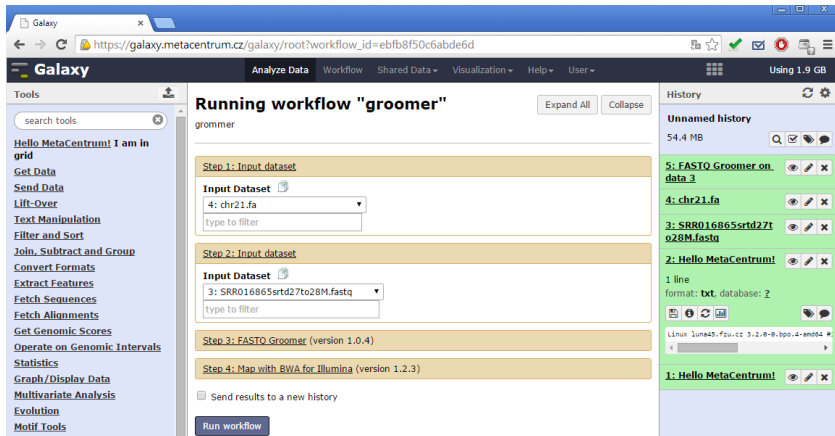
You can easily join several analyses together (output of one as input of other)

- ▶ 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.



The screenshot shows the Galaxy workflow editor interface. The browser address bar displays `https://galaxy.metacentrum.cz/galaxy/workflow/editor?id=ebfb8f50c6abde6d#`. The interface includes a top navigation bar with 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A left sidebar lists various tool categories such as 'Get Data', 'Text Manipulation', and 'NGS: Mapping'. The main 'Workflow Canvas | groomer' area contains three workflow steps: two 'Input Dataset' steps, a 'FASTQ Groomer' step, and a 'Map with BWA for Illumina' step. A context menu is open over the 'Map with BWA for Illumina' step, listing options: 'Save', 'Run', 'Edit Attributes', 'Auto Re-layout', and 'Close'. The right sidebar shows the configuration for the selected step, including options for 'Select a reference from history', 'FASTQ file', 'BWA settings to use', and 'Suppress the header in the output SAM file'.



The screenshot shows the Galaxy web interface for running a workflow named "groomer". The browser address bar shows the URL: `https://galaxy.metacentrum.cz/galaxy/root?workflow_id=ebfb8f50c6abde6d`. The interface includes a top navigation bar with "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "User" menus. A left sidebar contains a "Tools" section with a search bar and a list of tool categories such as "Get Data", "Text Manipulation", and "Statistics".

The main workflow area is titled "Running workflow 'groomer'" and contains four steps:

- Step 1: Input dataset**: "Input Dataset" dropdown set to "4: chr21.fa".
- Step 2: Input dataset**: "Input Dataset" dropdown set to "3: SRR016865srted27to28M.fastq".
- Step 3: FASTQ Groomer (version 1.0.4)**: No input parameters shown.
- Step 4: Map with BWA for Illumina (version 1.2.3)**: No input parameters shown.

At the bottom of the workflow area, there is a checkbox for "Send results to a new history" and a "Run workflow" button.

The right sidebar shows the "History" section, which is currently "Unnamed history" (54.4 MB). It lists five workflow steps:

- 5: FASTQ Groomer on data 3
- 4: chr21.fa
- 3: SRR016865srted27to28M.fastq
- 2: Hello MetaCentrum! (1 line, format: `txt`, database: `2`)
- 1: Hello MetaCentrum!

The terminal output for step 2 is visible, showing the command: `Linux lun45.fzu.cz 3.2.0-0-bpo.4-amd64 #`.

- ▶ 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`