



Bioinformatic Resources at the GWGDG

bioinfo.gwdg.de

Martin Paleico

Hosted: Workflow Tools

Workflow Tools:

- Galaxy: Open-source scientific workflow, data integration, and analysis persistence platform, mostly for bioinformatics (genetic sequence analysis)
galaxy.gwdg.de

The screenshot displays the Galaxy GWDG web interface. The main content area shows the configuration for a workflow named "Bowtie2 - map reads against reference genome (Galaxy Version 2.3.4.2)". The configuration includes several sections:

- Input:** A dropdown menu for "FASTA/Q file" with a search icon and a button to "No fastqsanger, fastqsanger.gz, fastqsanger.b2 or fasta dataset available". Below it, a note states "Must be of datatype 'fastqsanger' or 'fasta'".
- Output Options:** Two sections for "Write unaligned reads (in fastq format) to separate file(s)" and "Write aligned reads (in fastq format) to separate file(s)", each with "Yes" and "No" radio buttons.
- Parameters:** Two text input fields for "-un/-us-conc" and "-al/-al-conc" with explanatory text below them.
- Reference Genome:** A section titled "Will you select a reference genome from your history or use a built-in index?" with a dropdown menu set to "Use a built-in genome index". Below this, a note says "Built-ins were indexed using default options. See 'Indexes' section of help below". A "Select reference genome" dropdown is set to "Arabidopsis thaliana TAIR10".
- Read Groups:** A section titled "Set read groups information?" with a dropdown menu set to "Do not set".
- Analysis Mode:** A section titled "Select analysis mode" with a note: "Specifying read group information can greatly simplify your downstream analyses by allowing combining multiple datasets."

The right sidebar contains a "History" section with a search bar and a list of datasets. The list includes:

- 20: gffread on data 1: gtf
- 19: GFF-to-GTF on data 1: Converted
- 18: Htseq-count on data 8 and data 17 (no feature)
- 17: Htseq-count on data 8 and data 17
- 16: Htseq-count on data 75 and data 17 (no feature)
- 15: Htseq-count on data 75 and data 17
- 14: Htseq-count on data 83 and data 17 (no feature)
- 13: Htseq-count on data 83 and data 17
- 12: Wrapper for md5sum checksum on data 2
- 11: Wrapper for md5sum checksum on data 1

Hosted: Programming and IDEs

Programming and IDEs:

- RStudio: Interactive IDE for the R programming language rstudio.gwdg.de

The screenshot shows the RStudio IDE interface. The main window is divided into several panes:

- 1) Code Editor:** The central pane where R code is written. It contains a script with various R commands including `plot`, `arr`, `seq`, and `arr` functions.
- 2) History and Env.:** The upper right pane showing the execution history and the current environment.
- 3) Interactive Console and Terminal:** The lower left pane showing the output of the R code, including data frames and console messages.
- 4) Plots and Misc.:** The lower right pane displaying a scatter plot of the data generated by the code, showing a parabolic distribution.

- JupyterHub: Interactive Jupyter Notebooks in Python

Hosted: Windows Server, Misc.

Windows Server:

- WinTSBio: Windows server hosting a variety of applications
docs.gwdg.de/doku.php?id=de:services:server_services:compute_server:windows_terminal_server
- Various Statistics Software: SAS, STATA, Statistica, SPSS (some also available as individual licenses for free or at reduced prices)

Miscellaneous:

- Scientific Compute Cluster (SCC): Dozens of bioinformatics programs, plus computing power
- MASCOT: Protein identification using mass spectrometry data

Hosted: Misc.

Miscellaneous:

- RShiny Server: **NEW SERVICE** Hosting of user generated web applets in the RShiny framework rshiny.gwdg.de

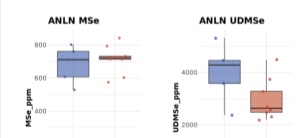
Currently hosted apps

MPI-NAT

Proteome and transcriptome data access for

Gargareta, Reuschenbach, Siems, Sun et al., elife (2022)

Myelin proteome quantification by MSe and UDMSe



Check Gene/Protein

Please type in a target gene symbol

Note: all symbols will be capitalized

Check

Color Scheme

Paper style

rshiny.gwdg.de/apps/CNS-myelin

It's Alive!

Number of bins: 1 50 90

Shiny Doc

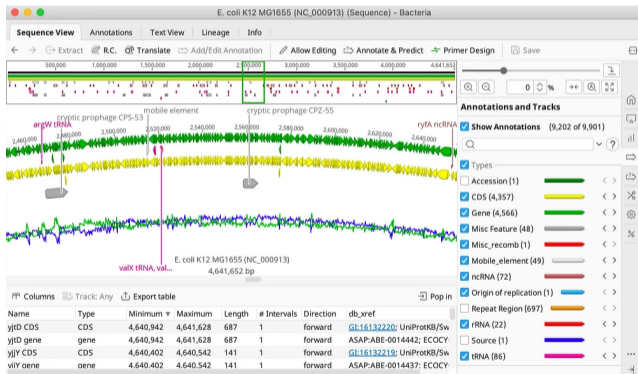
Region: N_Amer

It's Alive!

Number of bins: 1 50 90

Licenses

- XWin: Windows X server for accessing Linux based applications
- Geneious Prime: Platform of molecular biology and sequence analysis tools



Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

Figure: Source: Geneious Prime website

Community/Communication Channels

- Website: bioinfo.gwdg.de (out of date, will be replaced)
- Rocket Chat: **bioinformatics**
- Mailing List: listserv.gwdg.de/mailman/listinfo/hpc-in-lifescience

Questions?

Comments?

Suggestions for new services
or requests for particular tutorials?