deSTAIR Projects

Release 0.0.1

Mar 24, 2020

Contents

1	de.STAIR	1
2	GALAXY2.1Galaxy workflow generator for assisted RNA-Seq and BS/RRBS-Seq analyses2.2Plugin2.3Atoms2.4Workflows2.5Tools	7 10 12
3	RNA-Seq-Scrips 3.1 de.STAIR RNA-Seq scripts	15 15
4	Materials 4.1 Training material 4.2 Talks and posters	17 17 18
5	Indices and tables	19

de.STAIR

Please use the following link to reach our documentation: de.STAIR DOCS

de.STAIR is a consortium consisting of three scientific groups, Prof. Steve Hoffmann, Prof. Olaf Wolkenhauer, as well as Prof. Wolfgang R. Hess. The consortium provides training courses in bioinformatics and offers a Galaxy plugin called "Galax workflow generator". The latter mentioned plugin leads non-bioinformaticians through an interactive workflow builder. At the beginning of each session, the user can choose from a set of predefined topics such as "RRBS/BS-Seq analysis" or "Differential gene expression analysis". If a session has started, the user can choose from a well-selected set of tools per step. Furthermore, we provide additional information for each tool like "ease of use," "user base," "sensitivity," "speed," "configurability," as well as a detailed description.

Public Galaxy instance with our interactive tour builder: de.STAIR Galaxy

GALAXY

Build Status Docker Repository on Quay

2.1 Galaxy workflow generator for assisted RNA-Seq and BS/RRBS-Seq analyses

With contributions from a growing community of developers and users, the number of *alternative* Galaxy tools addressing the same questions has steadily risen. For this reason, it has become harder for the user to make informed decisions on the selection of tools, and their parameterization. The Galaxy training material has helped sharing bestpractices where novice and expert users can learn new methods to carry out their analyses. However, tools are preselected by the community, and the absence of a systematic overview of the available alternative tools of a Galaxy instance, does not train users on how to evaluate alternative algorithms and parameterizations.

The **Galaxy workflow generator** provides an overview of the available alternative tools that users can use to build RNA-Seq and BS-Seq analyses. Here, users are presented with new interactive pop-ups to compare and evaluate the suitability of each tool during the course of the analysis.

- How it works
- Run the instance
 - Installation requirements
 - Run the container
 - Login credentials
 - Howto update
- Tools
 - Quality control
 - Mapping
 - RNA-Seq

- BS/RRBS-Seq
- Utilities
- Support
 - Contributing
 - Contributors
- MIT license

2.1.1 How it works

The **Galaxy workflow generator** presents a different way of how to assist in the creation of Galaxy workflows. Here, single or multiple alternative Galaxy tools are provided as an **atom**: an interactive tour that illustrates tool options and parameterizations within the context of a selected analysis. In this Galaxy instance, we provide sample analyses organized into their logical **tasks**. For example, an RNA-Seq analysis is organized into:

- Task 1: Quality control and data preprocessing
- Task 2: Genome alignment
- Task 3: Transcript quantification and differential gene expression

Each task can be completed by means of one or more alternative atoms. The Galaxy workflow generator embeds a **plugin** that extends the Galaxy interface with novel interactive dialogs, which compare the alternative atoms that can be used to solve the current task in the selected analysis. Doing so, we are able to inform users about alternative tools and parameterizations.

The Figure below illustrates how the process of selecting atoms and building workflows takes place. Here, a user is carrying out an RNA-Seq analysis, which is divided into its three tasks: 1) Quality control and data preprocessing, 2) Genome alignment, 3) Transcript quantification and differential gene expression. For each task, the user selects the desired atom in the front-end interface, and navigates the corresponding interactive tour. The interactive tour drives the execution of each underlying tool in the Galaxy back-end. Once each task has been completed, the resulting workflow (traced in red), will thus be composed of the series of alternative tools of each selected atom, parameterized according to the interactive tour.

The Galaxy workflow generator collects alternative atoms to carry out RNA-Seq and BS/RRBS-Seq analyses.

2.1.2 Run the instance

The Galaxy workflow generator can run on you machine. The following sections will help you set up and run the instance. Optional, have a look here for more detailed information.

Installation requirements

The only requirement is Docker, which can be installed in different ways depending on the underlying system:

- Unix users should follow the Docker installation for Linux
- MacOS 10.12+ users should follow the Docker installation for Mac
- Windows 10+ users, should follow the Docker installation for Windows
- Non-unix users, whose operative system version is older than the aforementioned one, can rely on Kitematic

Run the container

Users not relying on Kitematic can open a terminal, or a Windows PowerShell, and type:

To allow the use of multiple threads, please prefix the aforementioned command in the following way:

```
$ docker run -e "GALAXY_CONFIG_PARALLEL_SLURM_PARAMS=--ntasks=8" \
    -e "GALAXY_CONFIG_PARALLEL_LOCAL_NTASKS=8" ...
```

To store users and data permanently, mount a directory on your harddrive into the containers /export directory by adding the following parameter:

\$ -v /absolute/path/to/local/directory:/export

To use an other user account rather than the default admins, to allow for anonoymous workflow generation prefix the command the following way and create the new user afterwards:

Kitematic users can launch the Galaxy instance by following these instructions.

After running the container, the Galaxy instance can be accessed from the local web browser, at the address localhost:8080.We recommend using Google Chrome, Chromium, or Mozilla Firefox.

Login credentials

To be able to analyse data, users need to be logged in.

- Galaxy users can create an account by clicking on the Login or Register button, on the top header
- Galaxy administrators can use the default credentials username: admin, password: admin, and then change settings later on.

Update to a newer image

Stop the running container.

\$ docker stop destair

Pull the latest image

\$ docker pull quay.io/destair/galaxy-workflow-generator:latest

Run the new image as described above

In case a bind mount (-v parameter, see above) was used, upgrades on atoms and tours can be found, compared and copied from the following locations.

Now restart Galaxy

\$ docker exec destair supervisorctl restart galaxy:

2.1.3 Tools

The tools catalog is bound to be modified and expanded during the development of the Galaxy workflow generator. The following list provides an overview of the tools that have been installed so far.

Quality control

Tool | Description | Reference :—: | :— | :— Cutadapt | Error-tolerant adapter removal tool for High-Throughput Sequencing reads | Martin 2011 FastQC | A quality control tool for high throughput sequence data | Andrews PRINSEQ | A quality control and data preprocessing tool for metagenomic data | Schmieder and Edwards 2011 Trim Galore! | Quality control tool for read trimming and filtering of NGS data | Trimmomatic | Quality control tool for read trimming and filtering of Illumina NGS data | Bolger et al. 2014

Mapping

Tool | Description | Reference :--: | :-- | :-- BWA | Burrows-Wheeler Aligner for mapping low-divergent sequences against a large reference genome | Li and Durbin 2010 bwameth | Fast and accurate aligner of BS-Seq reads | Brent et al. 2014 HISAT2 | Hierarchical indexing for spliced alignment of transcripts | Pertea et al. 2016 segement | Short sequence read to reference genome mapper | Otto et al. 2014 STAR | Rapid spliced aligner for RNA-seq data | Dobin et al. 2013

RNA-Seq

Tool | Description | Reference :---: | :--- | :--- BWA | Burrows-Wheeler Aligner for mapping low-divergent sequences against a large reference genome | Li and Durbin 2010 DESeq2 | Differential gene expression analysis based on the negative binomial distribution | Love et al. 2014 featureCounts | Genomic feature read count tool for summarising of genes, exons, and promoter counts | Liao et al. 2014 HISAT2 | Hierarchical indexing for spliced alignment of transcripts | Pertea et al. 2016 HTSeq-count | Tool for counting reads in features | Anders et al. 2015 Rcorrector | A kmer-based error correction method for RNA-seq data | Song et al. 2015 RSeQC | An RNA-seq Quality Control Package | Wang et al. 2012 SortMeRNA | A tool for filtering, mapping and OTU-picking NGS reads in metatranscriptomic and -genomic data | Kopylova et al. 2011 STAR | Rapid spliced aligner for RNA-seq data | Dobin et al. 2013

BS/RRBS-Seq

Tool | Description | Reference :--: | :-- | :-- Bismark | A program to map bisulfite treated sequencing reads to a genome of interest and perform methylation calls | Krueger et al. 2011 bwameth | Fast and accurate aligner of BS-Seq reads | Brent et al. 2014 MethylDackel | A tool to extract per-base methylation metrics from coordinate-sorted and indexed BS-seq alignments | Ryan segement | Short sequence read to reference genome mapper | Otto et al. 2014

Utilities

Tool | Description | Reference :--: | :-- | :-- SAMtools | Utilities for manipulating alignments in the SAM format | Heng et al. 2009 BEDTools | Utilities for genome arithmetic | Quinlan et al. 2010

2.1.4 Support

If you have questions, or don't know how to solve a problem, please contact us here, or file an issue.

Contributing

New contributions are always welcome. Please read these instructions before proceeding in doing so.

Contributors

- Andrea Bagnacani
- Bérénice Batut
- Björn Grüning
- Steffen Lott
- Konstantin Riege
- Markus Wolfien

2.1.5 MIT license

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PAR-TICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFT-WARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

2.2 Plugin

This repository contains the de.STAIR Galaxy plugin, which is incorporated in the Galaxy workflow generator.The plugin is based on the Galaxy webbooks, which extend the Galaxy interface with buttons and pop-up dialogues.

- How it works
- How to contribute
 - Set up the Galaxy environment
 - Set up the de.STAIR atoms
 - Set up the de.STAIR plugin
 - Run the Galaxy framework

2.2.1 How it works

The plugin provides:

- A new button placed on the Galaxy header
- New pop-up dialogues to
 - ask users which analysis they want to carry out
 - present alternative Galaxy tools to complete each task in the desired analysis

The button placed in the Galaxy header is used to start the interactive guide, which is implemented by means of a series of pop-ups. Each pop-up presents a list of alternative Galaxy tools, which can be adopted to complete the current task in the selected analysis. These are the *atoms*: interactive tours that illustrate one or more Galaxy tools. **Figure a** below shows how a plugin pop-up presents the set of alternative atoms to carry out the Quality control and data preprocessing.

By selecting an atom, a Galaxy webbook triggers the corresponding interactive tour, which the user navigates to execute each of its underlying Galaxy tools.Before selecting one, users can compare each alternative atom by reading their description page. These help pages explain the underlying Galaxy tool (or series thereof), and provide literature-based and empirical rankings.**Figure b** above shows the pop-up and description page of the selected atom.

The plugin defines new interactions to build Galaxy workflows:

- 1. Galaxy interface presents a pop-up to ask the user which analysis to carry out
- 2. User selects the type of analysis
- 3. Galaxy interface presents alternative atoms to carry out the 1st task of the selected analysis
- 4. Galaxy interface triggers the selected atom
- 5. User navigates the atom's interactive tour until the end, therefore completing the first task of the selected analysis
- 6. Repeat from 3, this time for 2nd task of the selected analysis

The loop ends when the last task of the Galaxy workflow is completed.

2.2.2 How to contribute

The plugin can be tested, modified, and further extended. To do so, you will need to install the atoms as well. The following sections will help you setting up the Galaxy framework to contribute to the plugin.

Option 1: Docker bind mount method

Run the Galaxy Docker container by creating a bind mount

for more parameters and further help, consult these instructions.

Once the container is ready, you will be able to modify our de.STAIR webbook backend python script:

You might also modify the client-side javascripts found in the src directory and compile via npm

```
/absolute/path/to/local/directory/export/galaxy-central/config/plugins/webhooks/

→switchtour/src

npm install

npm run build
```

Now restart the running Galaxy instance and access it via a webbrowser from the following address localhost:8080.

\$ docker exec destair supervisorctl restart galaxy:

Option 2: Local Galaxy setup method

Create a directory for the Galaxy repository

\$ export GALAXY_ROOT=/path/to/galaxy

Clone the Galaxy repository into the directory

\$ git clone https://github.com/galaxyproject/galaxy.git \$GALAXY_ROOT

Enter in the Galaxy repository, and checkout version 19.01

```
$ cd $GALAXY_ROOT
$ git checkout v19.01
```

ș git checkout vi9.0

Set up the de.STAIR atoms

Create a directory for the de.STAIR Galaxy atoms repository

\$ export DESTAIR_GALAXY_ATOMS=/path/to/destair-galaxy-atoms

Clone the de.STAIR Galaxy atoms repository

\$ git clone https://github.com/destairdenbi/galaxy-atoms.git \$DESTAIR_GALAXY_ATOMS

Set up the de.STAIR atoms into the cloned Galaxy repository

\$ \$DESTAIR_GALAXY_ATOMS/setup.sh

Set up the de.STAIR plugin

Create a directory for the de.STAIR Galaxy plugin repository

\$ export DESTAIR_GALAXY_PLUGIN=/path/to/destair-galaxy-plugin

Clone the de.STAIR Galaxy plugin repository and modify our de.STAIR switchtour client site javascript and backend python script according to your needs

\$ git clone https://github.com/destairdenbi/galaxy-webhooks.git \$DESTAIR_GALAXY_PLUGIN

Set up the de.STAIR plugin into the cloned Galaxy repository

\$ \$DESTAIR_GALAXY_PLUGIN/setup.sh

Run the Galaxy framework

Run the Galaxy launch script to verify the Galaxy framework setup and the inclusion of the new tools

```
$ $GALAX_ROOT/run.sh
```

2.3 Atoms

This repository contains the de.STAIR atoms that are incorporated in the Galaxy workflow generator, to assist users in the analysis of RNA-Seq and BS/RRBS-Seq data.

- How it works
- How to contribute
 - Set up Docker
 - When some tools are missing
 - When every tool is there
 - Test new atoms

2.3.1 How it works

Atoms are interactive Galaxy tours that illustrate one or more tools within the context of an experimental setup. Alternative atoms are therefore provided to describe the usage and parameterization of alternative tools to solve the same biocomputational problem.

In the Galaxy workflow generator, each atom describes a tool (or series thereof) to solve a *task* of the desired data analysis.

For instance, to carry out the tasks of an RNA-Seq analysis:

- 1. Quality control and data preprocessing
- 2. Genome alignment
- 3. Transcript quantification and differential gene expression

alternative atoms can be built using the following Galaxy tools:

Doing so, users can be informed about the availability of alternative strategies to carry out the desired analysis, and select the most appropriate atom within the context of their experimental setup.

2.3.2 How to contribute

Atoms can be tested, modified, and extended. The following sections will help you set up the Galaxy workflow generator to contribute with new atoms.

Set up Docker

Set up Docker by following the same installation requirements needed to run the Galaxy workflow generator.

When some tools are missing

New tools can be added in the Galaxy workflow generator by installing them...

- 1. From the admin panel (may load to unexpected errors upon execution)
- 2. Via re-building the docker image

Clone the Galaxy workflow generator:

\$ git clone https://github.com/destairdenbi/galaxy-workflow-generator.git

Enter the cloned repository, and edit the tools.yaml file. New entries are described in the Markdown format.The minimum set of metadata to be provided for each tool comprise:

- name
- owner
- tool_panel_section_label
- revisions

If unsure, please refer to the Galaxy Toolshed.

Once edited, you can build the Docker container locally:

\$ docker build -t destair-local:latest .

Now run your local image and access it via a web browser:

for more parameters and further help, consult these instructions.

When every tool is there

Atoms are interactive tours that illustrate one or more Galaxy tools.For ease of use within the context of specific experimental setups, we suggest to build atoms that solve entire tasks of a target analysis, *i.e.* Adapter clipping, Quality control, Genome alignment, Transcript quantification, and so on.

You can create atoms by leveraging on the Galaxy Tour Builder. However, be aware that the Tour Builder cannot identify elements with a missing tour_id attribute, which means that the resulting tour's bubbles will not always be placed on the HTML elements of interest. This limitation is particularly visible on those elements that require an explicit user input, such as dropdown menu selectors, checkboxes, etc.To overcome this problem, we defined some templates that can be used to place the interactive tour's bubble on the actual HTML elements of interest. Check our TEMPLATES.md file for examples and use cases.

If your atom needs to be placed within an analysis which is not yet included in the Galaxy workflow generator, you need to create a new *linker file*. Look at the examples in the repository.Otherwise, name the contributed atom with the label dgea_Xx.yaml, or bs_Xx.yaml. Where X is the task number, and x an identifier of your choice.

Test new atoms

New atoms can be tested by running a Galaxy Docker container using a *bind mount*, or by running the Galaxy workflow generator using the interactive plugin.

Option 1: Docker bind mount method

Run the Galaxy Docker container by creating a bind mount:

for more parameters and further help, consult these instructions.

Once the container is ready, you will be able to copy the new atom into

/absolute/path/to/local/directory/export/galaxy-central/config/plugins/tours/

Now, either restart the running Galaxy instance and access it via a web browser locally localhost:8080:

\$ docker exec destair supervisorctl restart galaxy:

or login to your Galaxy instance as administrator and run our de.STAIR plugin once with Update tours DB admin option checked.

New atoms will be automatically offered by our plugin and can be found by navigating the Galaxy interface under the header section Help -> Interactive Tours.

Option 2: Local Galaxy setup method

- Set up the Galaxy environment
- Set up the de.STAIR atoms
- Contribute new atoms as stated above
- Set up the de.STAIR plugin
- Run the Galaxy framework

2.4 Workflows

This repository contains the de.STAIR workflows that are incorporated in the Galaxy workflow generator.

2.5 Tools

This repository contains the de.STAIR tools that are incorporated in the Galaxy workflow generator, to assist users in the analysis of RNA-Seq and BS/RRBS-Seq data.

- How to contribute
 - Set up the Galaxy environment
 - Set up the de.STAIR tools

- Run the Galaxy framework

2.5.1 How to contribute

Tools can be tested, modified, and extended. The following sections will help you set up the Galaxy framework to contribute to the tools.

Set up the Galaxy environment

Create a directory for the Galaxy repository

\$ export GALAXY_ROOT=/path/to/galaxy

Clone the Galaxy repository into the directory

\$ git clone https://github.com/galaxyproject/galaxy.git \$GALAXY_ROOT

Enter in the Galaxy repository, and checkout version 19.01

```
$ cd $GALAXY_ROOT
$ git checkout v19.01
```

Set up the de.STAIR tools

Create a directory for the de.STAIR Galaxy tools repository

\$ export DESTAIR_GALAXY_TOOLS=/path/to/destair-galaxy-tools

Clone the de.STAIR Galaxy tools repository

\$ git clone https://github.com/destairdenbi/galaxy-tools.git \$DESTAIR_GALAXY_TOOLS

Set up the de.STAIR tools into the cloned Galaxy repository

\$ \$DESTAIR_GALAXY_TOOLS/setup.sh

Run the Galaxy framework

Run the Galaxy launch script to verify the Galaxy framework setup and the inclusion of the new tools

\$ \$GALAX_ROOT/run.sh

RNA-Seq-Scrips

3.1 de.STAIR RNA-Seq scripts

Place for some of our scripts.

For an introduction to git you might want to explore a Github tutorial

Feel free to contribute, fork or add! For editing mark down, please see cheat sheet

3.1.1 Source code

R code written and executed in RStudio (R version 3.4.3) Python code written in Spyder3 (Python version 3.6) Unix stuff executed with (Ubuntu 16)

3.1.2 de.STAIR

See project details at de.STAIR project page

3.1.3 Tools

- Demultiplexing flexbar
- Quality check with good old fastqc or qualimap
- Pre-processing fastx alternatively trimmomatic or trimgalore
- Mapping kallisto
 - See for reference genomes: RefSeq or Ensembl
- DE Analysis and statistics with kallisto sub-sequent tool sleuth

- Preferred setup for larger data sets, because kallisto is incredibly fast
- Sleuth needs R/Rstudio
- Alternatively using traditional mappers like Star or Hisat2
- TRAPLINE only for showcases =P

Materials

4.1 Training material

This repository collects the training materials used during our de.NBI training events, on-site support trainings, and on-line tutorials.

- de.NBI training events
- On-site support trainings
- Video tutorials
 - Gene enrichment analysis with Enrichr
 - Running a Galaxy Docker container using Kitematic

4.1.1 de.NBI training events

The following table collects the list of de.NBI training events that we organized to introduce data analysis concepts to undergraduates, PhD students, postdocs, and Life Science scholars.

I Date | Location | | :---: | :--- || 2016-10-26 | de.NBI training at EASyM 2016 Berlin || 2017-03-29 | de.NBI training at CASyM 2017 Ljubljana || 2017-09-18 | de.NBI training at GMDS 2017 Oldenburg || 2017-10-04 | de.NBI training at Albert-Ludwigs-Universität Freiburg || 2018-03-07 | de.NBI training at Christian-Albrechts-Universität Kiel || 2018-06-27 | de.NBI training at Leibniz-Institut für Alternsforschung (FLI) Jena || 2018-09-04 | de.NBI training at GMDS 2018 Osnabrück || 2019-03-06 | de.NBI training at Universität Rostock || 2019-09-08 | de.NBI training at GMDS 2019 Dortmund || 2019-10-09 | de.NBI training at Albert-Ludwigs-Universität Freiburg |

4.1.2 On-site support trainings

The following table collects the list of on-site support training events that we organized to introduce data analysis concepts to undergraduates, PhD students, postdocs, and Life Science scholars.

| Date | Location | | :---: | :--- | | 2017-03-20 | On-site support training at Universität Rostock | | 2017-09-29 | Onsite support training at Leibniz-Institut für Nutztierbiologie (FBN) Dummerstorf | | 2017-10-10 | On-site support training at Universität Rostock | | 2018-05-21 | On-site support training at Universität Rostock | | 2019-01-07 | On-site support training at Leibniz-Institut für Nutztierbiologie (FBN) Dummerstorf | | 2019-01-08 | On-site support training at Universität Rostock | | 2019-01-29 | On-site support training at Universität Rostock | | 2019-04-04 | On-site support training at Leibniz-Institut für Alternsforschung (FLI) Jena |

4.1.3 Video tutorials

In the following sections we provide some video tutorials, which we created to illustrate tools and data analysis concept to the participants of our training sessions.

Gene enrichment analysis with Enrichr

In this tutorial we show the use of Enrichr (Chen et al. 2013, Kuleshov et al. 2016) to carry out the enrichment analysis of a list of differtially expressed genes, which we obtained from the RNA-Seq analysis of two NGS Human breast cancer data sets.

Running a Galaxy Docker container using Kitematic

In this tutorial we show how to set up your system in order to run the Galaxy RNA workbench (Grüning et al. 2017, Fallmann et al. 2019) as a Docker container, using the Kitematic desktop application.

4.2 Talks and posters

This repository collects all demo materials, such as slides and posters, of de.STAIR talks, conferences, symposia, etc.

Indices and tables

- genindex
- modindex
- search