

Reproducibility and Galaxy



Yvan Le Bras, Marine biology station MNHN Concarneau
Grenoble, 2016 December 6th



Scheddule

- What is Galaxy?
- Why this “tool” for reproducibility?
- An example: essential Biological question
 - Presentation
 - Demo
- Workflow system
- Companion tools
- Levels of reproducibility

Loss of analysis skills

*“One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their **discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows.**”*

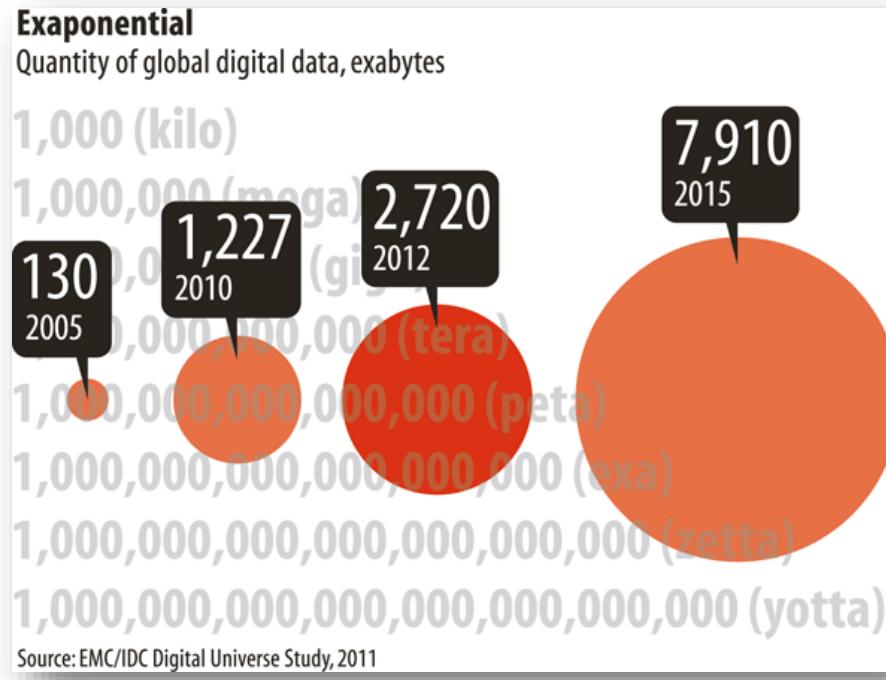
FORCE11 Future of Research Communication & e-Scholarship (2016) *The FAIR principles* <https://www.force11.org/group/fairgroup/fairprinciples>

Loss of analysis skills

“people with the skills to analyse data are scarce & will become scarcer”

Ludwig Siegele
Welcome to the yotta world
The economist, 2012

- Facilitate data analysis
 - Skills transfer by Training
 - Accessibility : More usable tools
 - Reproducibility : Peer review / re-analysis / Portability
 - Transparency : Public funding / citizens & society



Galaxy

A web platform dedicated to analyse data in Biology

Originally focuses on “large-scale genome analysis” (Giardine et al., 2005 - doi: [10.1101/gr.408650](https://doi.org/10.1101/gr.408650)) then Next Generation Sequencing

But... the general Galaxy “concept”/“Framework” can be applied to others Life sciences domains

“a framework that makes advanced computational tools usable by non experts”
(Afgan et al., 2016 - doi: [10.1093/nar/gkw343](https://doi.org/10.1093/nar/gkw343))

A real Universe and an engaged international community

Galaxy

A web platform dedicated to analyse data in Biology

Originally focuses on “large-scale genome analysis” (Giardine et al., 2005 - doi: [10.1101/gr.408650](https://doi.org/10.1101/gr.408650)) then Next Generation Sequencing

But... the general Galaxy “concept”/“Framework” can be applied to others Life sciences domains

“a framework that makes advanced computational tools usable by non experts”
(Afgan et al., 2016 - doi: [10.1093/nar/gkw343](https://doi.org/10.1093/nar/gkw343))

Intergalactic Utilities Commission

A real

The Intergalactic Utilities Commission (IUC) was established in 2012 to enable the pervasive use of the main Galaxy ToolShed by ensuring the repositories available in the tool shed include contents that are functionally correct and optimized for installation into local Galaxies. When appropriate, guidance is provided to repository owners so that they can understand what changes are necessary for their repository to be approved.

community

Galaxy & Reproducibility: a love story

Ten Simple Rules for Reproducible Computational Research (PlosOne)

- ▶ 1: For Every Result, [Keep Track](#) of How It Was Produced
 - ▶ 2: [Avoid Manual](#) Data Manipulation Steps
 - ▶ 3: [Archive](#) the Exact Versions of All External Programs Used
 - ▶ 4: [Version Control](#) All Custom Scripts
 - ▶ 5: Record [All Intermediate Results](#), When Possible in Standardized Formats
 - ▶ 6: For Analyses That Include Randomness, [Note Underlying Random Seeds](#)
 - ▶ 7: Always [Store Raw Data](#) behind Plots
 - ▶ 8: Generate Hierarchical Analysis Output, [Allowing Layers of Increasing Detail to Be Inspected](#)
 - ▶ 9: [Connect Textual Statements to Underlying Results](#)
 - ▶ 10: [Provide Public Access](#) to Scripts, Runs, and Results
- Several ways to follow them
- More or less complex (from manually to fully automatically)
 - More or less time-consuming (repeat, reproduce,, reuse)

Ten Simple Rules for Reproducible Computational Research

Geir Kjetil Sandve^{1,2*}, Anton Nekrutenko³, James Taylor⁴, Eivind Hovig^{1,5,6}

Galaxy & Reproducibility: a love story

Ten Simple Rules for Reproducible Computational Research (PlosOne)

- ▶ 1: For Every Result, [Keep Track](#) of How It Was Produced
 - ▶ 2: [Avoid Manual](#) Data Manipulation Steps
 - ▶ 3: [Archive](#) the Exact Versions of All External Programs Used
 - ▶ 4: [Version Control](#) All Custom Scripts
 - ▶ 5: Record [All Intermediate Results](#), When Possible in Standardized Formats
 - ▶ 6: For Analyses That Include Randomness, [Note Underlying Random Seeds](#)
 - ▶ 7: Always [Store Raw Data](#) behind Plots
 - ▶ 8: Generate Hierarchical Analysis Output, [Allowing Layers of Increasing Detail to Be Inspected](#)
 - ▶ 9: [Connect Textual Statements to Underlying Results](#)
 - ▶ 10: [Provide Public Access](#) to Scripts, Runs, and Results
- Several ways to follow them
- More or less complex (from manually to fully automatically)
 - More or less time-consuming (repeat, reproduce,, reuse)

Ten Simple Rules for Reproducible Computational Research

Geir Kjetil Sandve^{1,2*}, Anton Nekrutenko³, James Taylor⁴, Eivind Hovig^{1,5,6}



Speakers



Jeremy Goecks

Galaxy Project, George Washington University



Anton Nekrutenko

Galaxy Project, Penn State University



James Taylor

Johns Hopkins University

S3 T1 GCC 2016 Galaxy Update [PDF](#)

Galaxy & Reproducibility: a love story

2005

Genome Res. 2005 Oct; 15(10): 1451–1455.
doi: [10.1101/gr.4086505](https://doi.org/10.1101/gr.4086505)

Galaxy: A platform for interactive large-scale genome analysis

Belinda Giardine,¹ Cathy Riemer,¹ Ross C. Hardison,¹ Richard Burhans,¹ Laura Elnitski,² Prachi Shah,^{1,2} Yi Zhang,¹ Daniel Blankenberg,¹ Istvan Albert,¹ James Taylor,¹ Webb Miller,¹ W. James Kent,³ and Anton Nekrutenko^{1,4}

PMCID: PMC1240089

2010

Genome Biol. 2010; 11(8): R86.
Published online 2010 Aug 25. doi: [10.1186/gb-2010-11-8-r86](https://doi.org/10.1186/gb-2010-11-8-r86)

PMCID: PMC2945788

Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences

Reviewed by Jeremy Goecks,¹ Anton Nekrutenko,^{1,2} and James Taylor,^{1,3} The Galaxy Team team@galaxyproject.org

2016

The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update

Enis Afgan¹, Dannon Baker¹, Marius van den Beek², Daniel Blankenberg³, Dave Bouvier³, Martin Čech³, John Chilton³, Dave Clements¹, Nate Corao³, Carl Eberhard¹, Björn Grüning^{4,5}, Aysam Guerler¹, Jennifer Hillman-Jackson³, Greg Von Kuster⁶, Eric Rasche⁷, Nicola Soranzo⁸, Nitesh Turaga¹, James Taylor^{1,*}, Anton Nekrutenko^{3,*} and Jeremy Goecks^{9,*}

Nucl. Acids Res. (2016)
doi: [10.1093/nar/gkw343](https://doi.org/10.1093/nar/gkw343)
First published online: May 2, 2016

This article is Open Access

Abstract [Free](#)

» Full Text (HTML) [Free](#)

Full Text (PDF) [Free](#)

All Versions of this Article:
[gkw343v1](#)

[44/W1/W3](#) most recent

Enhancing pre-defined workflows with ad hoc analytics using Galaxy, Docker and Jupyter

Björn Grüning, Eric Rasche, Boris Rebolledo Jaramillo, Carl Eberhard, Torsten Houwaart, John Chilton, Nathan Corao, Rolf Backofen, James Taylor, Anton Nekrutenko
doi: <http://dx.doi.org/10.1101/075457>

Feature	Galaxy	Jupyter	Galaxy/Jupyter
Low barrier of entry for a naive user	*		*
Versatility of available tools		*	*
Provenance tracking	*	*	*
Hardware backend for processing of large datasets	*		*
Attractive to experimentalists	*		*
Attractive to bioinformaticians and data scientists		*	*

Table 1: Congruence between Galaxy and Jupyter as a function of their pros and cons.

« Galaxy seeks to make data-intensive research more accessible, transparent and reproducible by providing a Web-based environment in which users can perform computational analyses and have all of the details automatically tracked for later inspection, publication, or reuse »

Galaxy & Reproducibility: a love story

Genome Res. 2005 Oct; 15(10): 1451–1455.
doi: [10.1101/gr.4086505](https://doi.org/10.1101/gr.4086505)

Galaxy: A platform for interactive large-scale genome analysis

Belinda Giardine,¹ Cathy Riemer,¹ Ross C. Hardison,¹ Richard Burhans,¹ Laura Elnitski,² Prachi Shah,^{1,2} Yi Zhang,¹ Daniel Blankenberg,¹ Istvan Albert,¹ James Taylor,¹ Webb Miller,¹ W. James Kent,³ and Anton Nekrutenko^{1,4}

PMCID: PMC1240089

2005

Genome Biol. 2010; 11(8): R86.
Published online 2010 Aug 25. doi: [10.1186/gb-2010-11-8-r86](https://doi.org/10.1186/gb-2010-11-8-r86)

2010

Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences

Reviewed by Jeremy Goecks,¹ Anton Nekrutenko,^{1,2} and James Taylor,^{1,4} The Galaxy Team team@galaxyproject.org

PMCID: PMC2945788

2016

The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update

Enis Afgan¹, Dannon Baker¹, Marius van den Beek², Daniel Blankenberg³, Dave Bouvier³, Martin Čech³, John Chilton³, Dave Clements¹, Nate Corao³, Carl Eberhard¹, Björn Grüning^{4,5}, Aysam Guerler¹, Jennifer Hillman-Jackson³, Greg Von Kuster⁶, Eric Rasche⁷, Nicola Soranzo⁸, Nitish Turaga¹, James Taylor^{1,*}, Anton Nekrutenko^{3,*} and Jeremy Goecks^{9,*}

Nucl. Acids Res. (2016)
doi: [10.1093/nar/gkw343](https://doi.org/10.1093/nar/gkw343)
First published online: May 2, 2016

This article is Open Access

Abstract [Free](#)

» Full Text (HTML) [Free](#)

Full Text (PDF) [Free](#)

All Versions of this Article:
[gkw343v1](#)

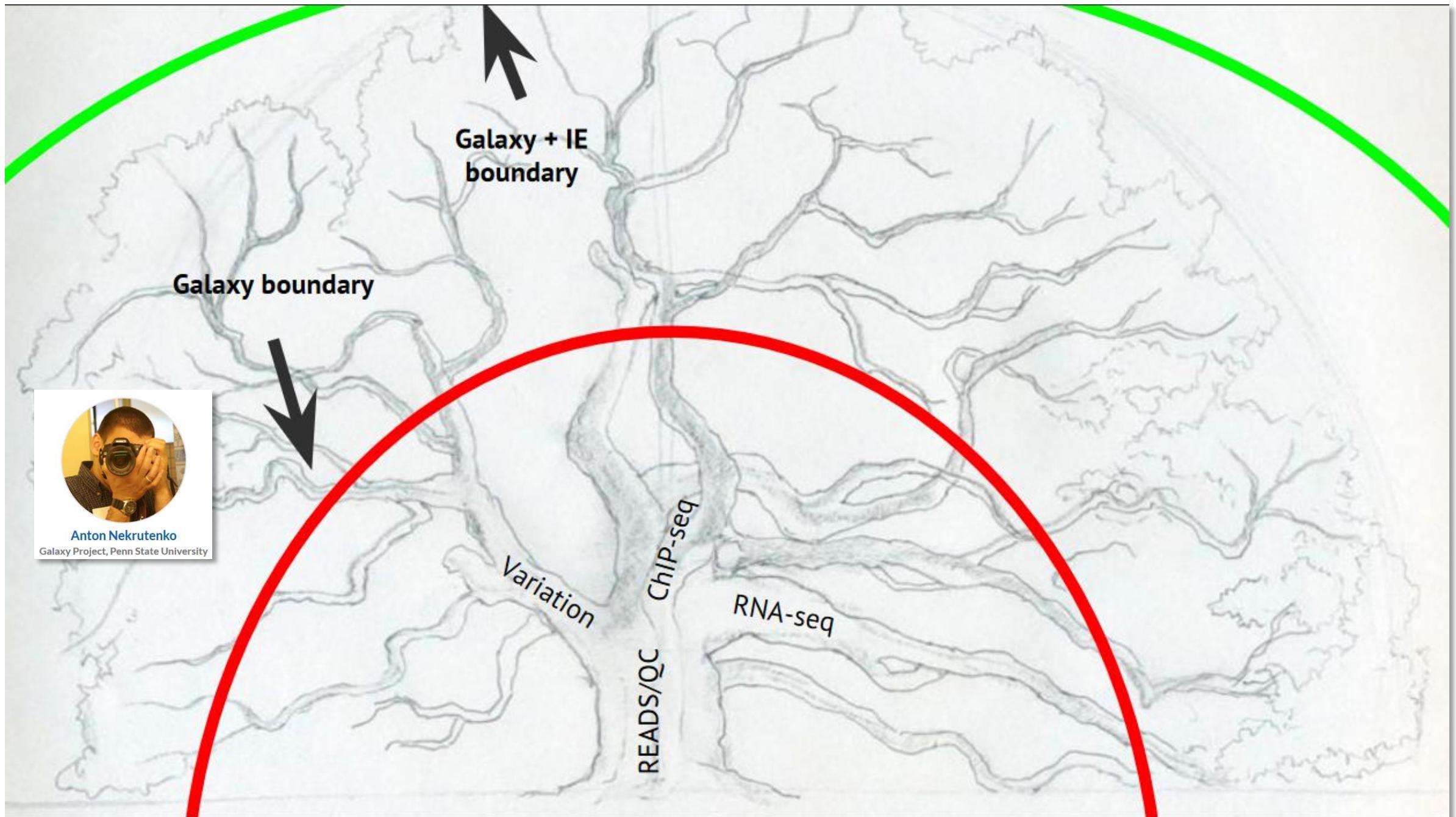
[44/W1/W3](#) most recent

Enhancing pre-defined workflows with ad hoc analytics using Galaxy, Docker and Jupyter

Björn Grüning, Eric Rasche, Boris Rebolledo Jaramillo, Carl Eberhard, Torsten Houwaart, John Chilton, Nathan Corao, Rolf Backofen, James Taylor, Anton Nekrutenko
doi: <http://dx.doi.org/10.1101/075457>

« Galaxy seeks to make data-intensive research more accessible, transparent and reproducible by providing a Web-based environment in which users can perform computational analyses and have all of the details automatically tracked for later inspection, publication, or reuse »

« As we traverse the tree, the set of established analysis tools becomes increasingly sparse and it is up to an individual researcher to come up with statistical and visualization approaches necessary to reach the leaves (or fruits) representing conclusive, publishable results »



Feedback after several years in France

- Is there really (in term of reproducibility) any instance used in production ?
 - Docker, Interactive environments functionalities, how many instances... 1 ?
 - Use of workflows in a production manner ?
 - Rennes IFB GenOuest INRIA core facility (INRA RapsoSNP/ CNRS UR1 easyOTU)
 - Galaxy updates
 - Storage: Quotas limitation
- Don't get out of Galaxy
 - “you can use IGV”
 - “you can use Excel”...

An essential biological question ;)

- Which coding part of the human genome have the biggest diversity ?
- So in Bioinformatics: Which exon have the biggest SNP density

An essential biological question ;) Demo

- Which coding part of the human genome have the biggest diversity ?
- So in Bioinformatics: Which exon have the biggest SNP density
- Workflow steps:
 - Download exon information from Human genome: 1 dataset
 - Download SNP information from Human genome: 1 dataset
 - Join and sort these two datasets on an “integrated dataset”
 - Count SNP occurrences by exon
 - Select the 100 Exons presenting the highest SNP diversity
 - Compare original datasets and final filtered dataset

And Visualization

An essential biological question ;) Demo

- Which coding part of the human genome have the biggest diversity ?
- So in Bioinformatics: Which exon have the biggest SNP density
- Workflow step
 - Download SNP information from Human genome: 1 dataset
 - Join and sort these two datasets on an “integrated dataset”

Demo time!!!!!!

And Visualization

<https://usegalaxy.org/>

<https://galaxy.genouest.org/>

Workflow system

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 52%

Tools

search tools

DATA TRANSFER

[Get Data](#)

[Send Data](#)

[Archive tools](#)

SYMBIOSE

[Metagenomics benchmark](#)

[Text Manipulation](#)

[Fasta to OTUs](#)

[NGS: 454 data Manipulation](#)

[Motif](#)

[NGS: Alignment](#)

[NGS: Assembly](#)

[NGS: Genome manipulation](#)

[Systems biology](#)

[Data management](#)

[GenOuest: Mothur toolkit](#)

[KLAST](#)

COLIBREAD

[Colib'read tool suite](#)

BIOGENOUEST COMMUNITY

[QTL Detection](#)

GENOMICS

[Alignment](#)

GENETICS

GenOuest Galaxy server is running

For any question or tools request, please send an email at support@genouest.org

History

human exon

6 shown

8.31 MB

6: Compare two Datasets on data 5 and data 1

5: Selection des 100 meilleurs SNPs

4: Count on data 3

3: Join on data 2 and data 1

2: UCSC Main on Human:.snp147Common (chr22:1-50818468)

1: UCSC Main on Human: refGene (chr22:1-50818468)

GenOuest Bioinformatics platform

Development, expertise and resources for bioinformatics

IRISA

Inria

Biogenouest

Réseau BRETAGNE

Workflow system: How to Edit

Galaxy / GenQuest

Analyze Data Workflow Shared Data Visualization Help User Using 52%

Tools search tools Inputs DATA TRANSFER Get Data Send Data Archive tools SYMBIOSE Metagenomics benchmark Text Manipulation Fasta to OTUs NGS: 454 data Manipulation Motif NGS: Alignment NGS: Assembly NGS: Genome manipulation Systems biology Data management GenOuest: Mothur toolkit KLAST COLIBREAD Colib'read tool suite BIOGENQUEST COMMUNITY QTL Detection

Workflow Canvas | Workflow constructed from history 'human exon'

Compare two Datasets to find common or distinct rows (Galaxy Version 1.0.2)

Compare Data input 'input1' (tabular)

Using column 4

against Data input 'input2' (tabular)

and column 2

To find Matching rows of 1st dataset

Annotation / Notes

Add an annotation or note for this step. It will be shown with the workflow.

Email notification Yes No

An email notification will be sent when the job has completed.

Output cleanup Yes No

Delete intermediate outputs if they are not used as input for another job.

Configure Output: 'out_file1'

```
graph LR; A[Input dataset] --> B[Join]; B --> C[Count]; C --> D[Select first]; D --> E[Compare two Datasets];
```

Workflow system: How to share

Galaxy / GenOuest Analyze Data Workflow Shared Data ▾ Visualization ▾ Help ▾ User ▾ Using 52%

[Go back to Workflows List](#)

Workflow ' Workflow constructed from history 'human exon''

Share

This workflow is currently restricted so that only you and the users listed below can access it.

[Make Workflow Accessible via Link](#)

Generates a web link that you can share with other people so that they can view and import the workflow.

[Make Workflow Accessible and Publish](#)

Makes the workflow accessible via link (see above) and publishes the workflow to Galaxy's [Published Workflows](#) section, where it is publicly listed and searchable.

You have not shared this workflow with any users yet.

[Share with a user](#)

Export

[Download](#) workflow as a file so that it can be saved or imported into another Galaxy server.

This workflow must be accessible. Please use the option above to "Make Workflow Accessible and Publish" before receiving a URL for importing to another Galaxy.

[Create image](#) of workflow in SVG format

Export to the [www.myexperiment.org](#) site.

myExperiment username:
ylebras

myExperiment password:

[Export to myExperiment](#)

Workflow system: Workflow file

```
"a_galaxy_workflow": "true",
"annotation": "",
"format-version": "0.1",
"name": "Workflow constructed from history 'human exon'",
"steps": {
    "0": {
        "annotation": "",
        "content_id": null,
        "id": 0,
        "input_connections": {},
        "inputs": [
            {
                "description": "",
                "name": "Exons"
            }
        ],
        "label": null,
        "name": "Input dataset",
        "outputs": [],
        "position": {
            "left": 205,
            "top": 201
        },
        "tool_errors": null,
        "tool_id": null,
        "tool_state": "{\"name\": \"Exons\"}",
        "tool_version": null,
        "type": "data_input",
        "uuid": "ed889fb3-77fa-4aae-aeb4-6b42d3ab52e6",
        "workflow_outputs": []
    },
    "1": {
        "annotation": "",
        "content_id": null,
        "id": 1,
        "input_connections": {},
        "inputs": [
            {
                "description": "",
                "name": "SNPs"
            }
        ],
        "label": null,
        "name": "Input dataset",
        "outputs": [],
        "position": {
            "left": 205,
            "top": 321
        },
        "tool_errors": null,
        "tool_id": null,
        "tool_state": "{\"name\": \"SNPs\"}",
        "tool_version": null,
        "type": "data_input",
        "uuid": "f48bc5e1-bef5-4cee-94b2-be01a120638b",
        "workflow_outputs": []
    }
}
```

A .ga file

```
"1": {
    "annotation": "",
    "content_id": null,
    "id": 1,
    "input_connections": {},
    "inputs": [
        {
            "description": "",
            "name": "SNPs"
        }
    ],
    "label": null,
    "name": "Input dataset",
    "outputs": [],
    "position": {
        "left": 205,
        "top": 321
    },
    "tool_errors": null,
    "tool_id": null,
    "tool_state": "{\"name\": \"SNPs\"}",
    "tool_version": null,
    "type": "data_input",
    "uuid": "f48bc5e1-bef5-4cee-94b2-be01a120638b",
    "workflow_outputs": []
},
"2": {
    "annotation": "",
    "content_id": "toolshed.g2.bx.psu.edu/repos/devteam/join/gops_join_1/1.0.0",
    "id": 2,
    "input_connections": {
        "input1": {
            "id": 0,
            "output_name": "output"
        },
        "input2": {
            "id": 1,
            "output_name": "output"
        }
    },
    "inputs": [
        {
            "description": "runtime parameter for tool Join",
            "name": "input2"
        }
    ],
    "label": null,
    "name": "Join"
}
```

Workflow system: Repositories

my experiment Home Users Groups Workflows Files Packs Search Go Log in Register

Home / Workflows

Workflows

« Previous 1 2 3 4 5 ... 8 Next » Sort by: Most downloaded Results per page: 10

Showing 80 results. Use the filters on the left and the search box below to refine the results.

Search filter terms Search

Filter by type

- Taverna 2 1558
- Taverna 1 566
- RapidMiner 280
- Galaxy 80
- KNIME 52
- Kepler 48
- Bioclipse ... 44
- LONI Pipe... 26
- GWorkflow... 24
- BioExtract... 19
- More...

Filter by tag

- ngs 23
- BLAST 17
- metageno... 17
- de novo a... 14
- metavisitor 14
- virus diag... 14
- virus disco... 14
- galaxy 12
- chip-seq 5
- filtering 5
- More...

Filter by user

- Ylebras 22
- Christophe... 14
- Björn Grü... 8
- jason limb... 7
- Portia Holl... 7

Type: Galaxy Remove all filters

Detrprok (4) Download

Uploader: Clairetn In the case study of stranded and prokaryotic RNAseq data, the DetRprok workflow detects candidates of 3 kinds of non coding RNA: 5UTRs, antisense RNAs, and small RNAs. Inputs: i) an mapping file (bam format) containing one valid alignment by read, ii) a feature file (gff format) annotating the genomic sequences used for the mapping. Dependencies (from the Galaxy toolshed): "s_mart", "detrprok_scripts"

Created: 2013-05-20 | Last updated: 2015-11-03

Basic RNA-Seq Analysis - Differential Expr... (1) Download

Uploader: David De Roure From the RNA-Seq analysis tutorial during the Functional Genomics Workshop 2012 https://caps.osu.edu/pfg-workshop Workflow published by mejia-guerra on Galaxy Jun 22, 2012 imported to myExperiment Jul16, 2012 during demonstration of Galaxy-myExperiment integration

Created: 2012-07-16 | Last updated: 2012-07-16

NGS : RNA-seq differential expression anal... (1) Download

Uploader: Ylebras Analyse paired-end RNA-seq reads of 2 different individuals with FastQC, Tophat2, Cufflinks and Cuffdiff using a reference annotation in gtf format

Created: 2014-02-27 | Last updated: 2014-06-10
Credits: Ylebras

Perform QC on FastQ data (1) Download

Uploader: Björn Grü... This Galaxy pipeline performs quality control analyses on FastQ data. A report is generated which provides various

Workflow system: Repositories

Galaxy Tool Shed [Repositories](#) [Groups](#) [Help](#) [User](#)

4300 valid tools on Nov 09, 2016

Search

- [Search for valid tools](#)
- [Search for workflows](#)

Valid Galaxy Utilities

- [Tools](#)
- [Custom datatypes](#)
- [Repository dependency definitions](#)
- [Tool dependency definitions](#)

All Repositories

- [Browse by category](#)
- [Login to create a repository](#)

Available Actions

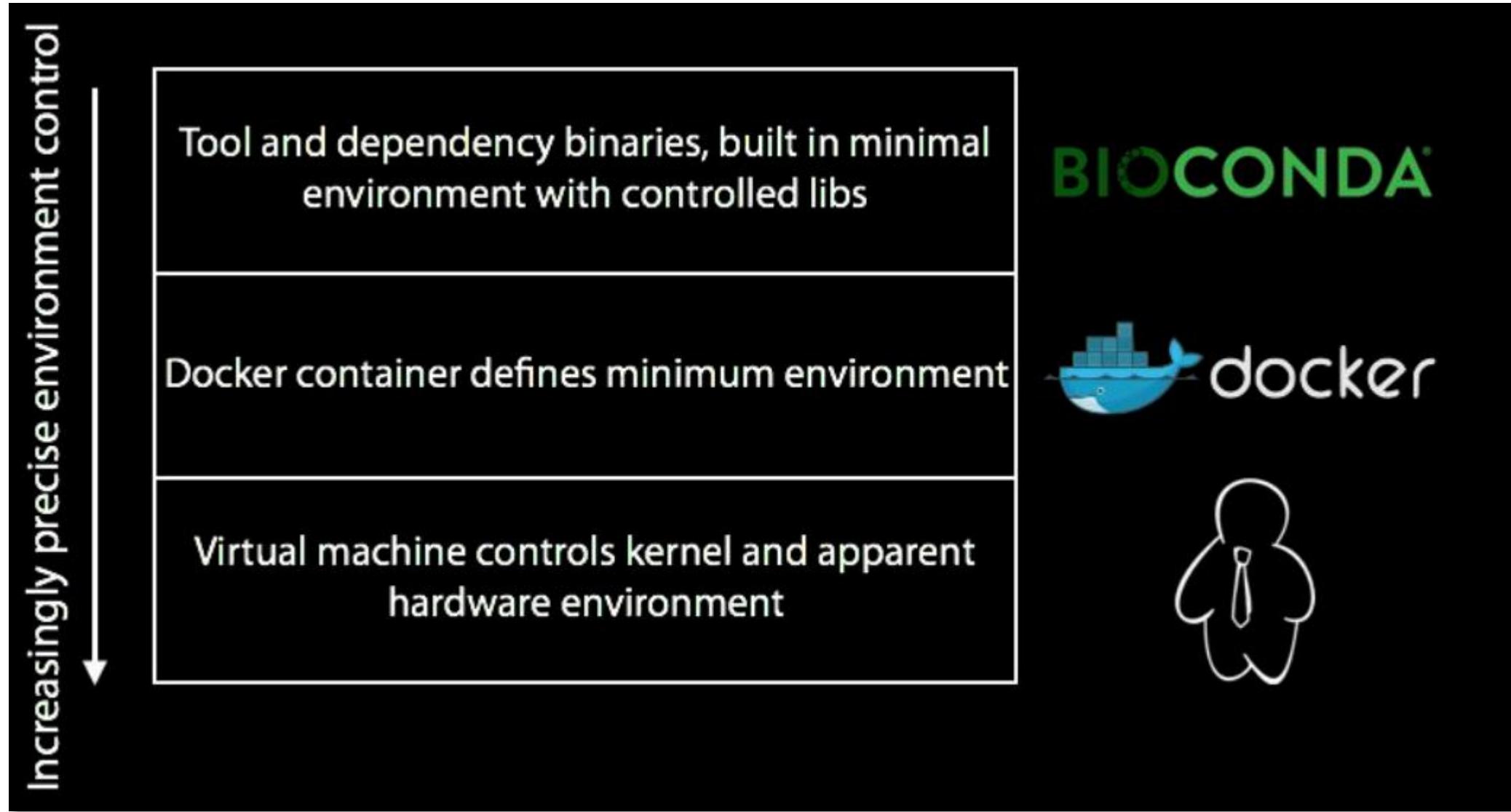
- [Login to create a repository](#)

Repositories with matching workflows

workflow name: **workflow**
exact matches only: **False**

<input type="checkbox"/> Repository name	Synopsis	Revision	Owner
<input type="checkbox"/> asmsTutorial_2016	ASMS 2016 Tutorial Workflows	673718067538	galaxyp
<input type="checkbox"/> blockclust_workflow	BlockClust Workflow	70170338b4f7	rnameam
<input type="checkbox"/> caddsuite_workflows	CADDSuite workflows for use in galaxy	39def4dc6718	marcel
<input type="checkbox"/> caddsuite_workflows	CADDSuite workflows for use in galaxy	e047a2629ecf	marcel
<input type="checkbox"/> gandalfWorkflow	gandalf Workflow	3ca0f78f1102	urgi-team
<input type="checkbox"/> gandalfWorkflow	gandalf Workflow	9c2cbeb63ff0	urgi-team
<input type="checkbox"/> gandalfWorkflow	gandalf Workflow	7557b71a92e7	urgi-team
<input type="checkbox"/> gandalfWorkflow	gandalf Workflow	1fbe965e22e2	urgi-team
<input type="checkbox"/> gcc2016Tutorial	GCC2016 Tutorial Workflows	ec32f7039106	galaxyp
<input type="checkbox"/> mdust	mdust	034d5b66e97a	urgi-team
<input type="checkbox"/> metavisitor_workflows	A collection of workflows using the Metavisitor tools	ba15c770fd40	drosofff

Companion tools: Dependencies & VM



Companion tools: Dependencies & VM

- ToolShed

The screenshot shows the Galaxy Tool Shed interface. The top navigation bar includes links for 'Repositories', 'Groups', 'Help', and 'User'. A 'Repository Actions' button is located in the top right corner. The main content area displays the 'stacks_denovomap' repository, which has been cloned from https://toolshed.g2.bx.psu.edu/repos/iuc/stacks_denovomap. The repository type is listed as 'Unrestricted'. The 'Contents' section shows a directory structure with files like 'macros.xml', 'stacks_denovomap.xml', and a 'test-data/' folder. Below the directory listing is the XML code for the pipeline definition.

```
<?xml version="1.0"?>
<repositories description="Stacks tools for building loci from short-read sequences, such as RAD-seq"> <repository changeset_revision="ae705d244e28" name="stacks_stacks">
<repository changeset_revision="a62f7e799494" name="stacks_cstacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="998dfec32bde" name="stacks_denovomap" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="969a09779e3c" name="stacks_genotypes" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="dbb325dfe508" name="stacks_populations" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="c2733bcc51f7" name="stacks_procrad" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="77e29e646101" name="stacks_pstacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="827c8c6ec4c1" name="stacks_re fmap" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="fec69a2d92d7" name="stacks_rxstacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="679f97ca0dd5" name="stacks_sstacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="ff04b3748654" name="stacks_ustacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
</repositories>
```

The screenshot shows the Galaxy Tool Shed interface. The top navigation bar includes links for 'Repositories', 'Groups', 'Help', and 'User'. A 'Repository Actions' button is located in the top right corner. The main content area displays the 'suite_stacks' repository, which has been cloned from https://toolshed.g2.bx.psu.edu/repos/iuc/suite_stacks. The repository type is listed as 'Unrestricted'. The 'Contents' section shows a single file named 'repository_dependencies.xml'. Below the file listing is the XML code for the repository dependencies.

```
<?xml version="1.0"?>
<repositories description="Stacks tools for building loci from short-read sequences, such as RAD-seq"> <repository changeset_revision="ae705d244e28" name="stacks_stacks">
<repository changeset_revision="a62f7e799494" name="stacks_cstacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="998dfec32bde" name="stacks_denovomap" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="969a09779e3c" name="stacks_genotypes" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="dbb325dfe508" name="stacks_populations" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="c2733bcc51f7" name="stacks_procrad" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="77e29e646101" name="stacks_pstacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="827c8c6ec4c1" name="stacks_re fmap" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="fec69a2d92d7" name="stacks_rxstacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="679f97ca0dd5" name="stacks_sstacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
<repository changeset_revision="ff04b3748654" name="stacks_ustacks" owner="iuc" toolshed="https://toolshed.g2.bx.psu.edu"/>
</repositories>
```

The screenshot shows the Galaxy Tool Shed interface. The top navigation bar includes links for 'Repositories', 'Groups', 'Help', and 'User'. A 'Repository Actions' button is located in the top right corner. The main content area displays the 'stacks_denovomap' repository, which has been cloned from https://toolshed.g2.bx.psu.edu/repos/iuc/stacks_denovomap. The repository type is listed as 'Unrestricted'. The 'Contents' section shows a directory structure with files like 'macros.xml', 'stacks_denovomap.xml', and a 'test-data/' folder. Below the directory listing is the XML code for the pipeline definition.

```
<tool id="stacks_denovomap" name="Stacks: de novo map" version="@WRAPPER_VERSION@.0">
<description>the Stacks pipeline without a reference genome (denovo_map.pl)</description>
<macros>
<import>macros.xml</import>
</macros>
<expand macro="requirements"/>
<expand macro="stdio"/>
<command><![CDATA[
#from os.path import splitext
#import re
```

Companion tools: Dependencies & VM

- Conda / Docker
 - Github recipes

This screenshot shows the GitHub repository `bioconda/bioconda-recipes`. It displays the contents of the `stacks` directory. The directory structure includes a folder named `1.37` containing the file `new stacks version 1.40`, a folder named `1.40` containing the file `new stacks 1.42 versino`, and files `build.sh` and `meta.yaml`.

This screenshot shows the `bioconda/packages/stacks` page. It lists a table of conda packages for the `stacks` package. The columns include Type (conda), Size, Name, Uploaded, Uploader, Downloads, and Labels. The table shows several versions of the `stacks` package, such as `1.42-3`, `1.42-2`, `1.40-1`, `1.40-0`, and `1.37-0`, each with its corresponding details.

Type	Size	Name	Uploaded	Uploader	Downloads	Labels
conda	18.4 MB	linux-64/stacks-1.42-3.tar.bz2	2 months and 9 days ago	bioconda	45	main
conda	18.4 MB	linux-64/stacks-1.42-2.tar.bz2	2 months and 14 days ago	bioconda	9	main
conda	18.4 MB	linux-64/stacks-1.42-1.tar.bz2	2 months and 14 days ago	bioconda	2	main
conda	14.6 MB	linux-64/stacks-1.40-1.tar.bz2	5 months and 23 days ago	bioconda	281	main
conda	12.5 MB	linux-64/stacks-1.40-0.tar.bz2	5 months and 26 days ago	bioconda	77	main
conda	11.4 MB	linux-64/stacks-1.37-1.tar.bz2	8 months and 1 day ago	bioconda	160	main
conda	11.4 MB	linux-64/stacks-1.37-0.tar.bz2	8 months and 28 days ago	bioconda	37	main

This screenshot shows the Quay.io repository `biocontainers/stacks`. It displays a list of repository tags. The tags listed are `1.42-3`, `1.42-2`, `1.40-1`, `1.40-0`, and `1.37-0`. Each tag entry includes the last modified date (e.g., "2 months ago") and a status of "Unsupported".

This screenshot shows the Continuum Analytics `bioconda/packages/stacks 1.42` page. It provides information about the `stacks` package version 1.42. It includes a table of installers, a command to install the package using `conda install`, and a note about the package being GPL licensed with 611 total downloads.

Companion tools: Planemo

A Galaxy Philosophy

John Chilton

Galaxy Project, Penn State University

- The most important Galaxy user is the *bench scientist* using the GUI, they come first!
 - No one wants to inconvenience a bioinformatician or developer, but we will if absolutely required for the biologist.
- Galaxy workflows will *never require an SDK*.
 - Planemo, workflow formats, etc... are conveniences for people who prefer developer processes.

Companion tools: Planemo

Planemo's Success

It is *the way* to develop Galaxy tools in 2016! Why?

- Artifact-centric - not Galaxy-centric or registry-centric.
- Works with existing developer tools - CLI, Git(hub), CI (Travis).
- Very flexible, easily configurable.
- Well documented with focus on *usage examples*.

It is about **developer workflow**.



Companion tools: Planemo

Workflows are Programs

When I write programs...

- ... I write *tests* (and write them first)!
- ... I commit them to *Github*!
- ... use a text editor - *my* text editor!

CWL & Galaxy

Experimental tool support today using planemo.

```
$ planemo serve --cwl <tool.cwl>
$ planemo test <tool.cwl>
$ planemo run <tool.cwl> <job.json>
```

When *serve*, *test*, *run* encounter CWL tools they will use a Galaxy fork.

Work in progress at <https://github.com/common-workflow-language/galaxy>.

Levels of reproducibility

- Repeat: allowing redoing the exact same experiment
- Reuse: searching for workflows in ToolShed, MyExperiment, Github
- Replicate: Reanalyse data on another system with different config
- Reproduce: Obtain the same conclusions with different Workflows

Levels of reproducibility

- Repeat: allowing redoing the exact same experiment

The screenshot shows the Galaxy bioinformatics platform interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', 'User', and a session status 'Using 52%'. The left sidebar, titled 'Tools', lists categories like BIOSTATISTICS, METABOLOMICS, and OTHERS, each with a list of specific tools. The central workspace displays a workflow titled 'Running workflow "Workflow constructed from history 'human exon'''. The workflow consists of six steps: Step 1: Input dataset (Exons), Step 2: Input dataset (SNPs), Step 3: Join (version 1.0.0), Step 4: Count (version 1.0.2), Step 5: Select first (version 1.0.0), and Step 6: Compare two Datasets (version 1.0.2). Each step has a dropdown menu and a 'type to filter' input field. At the bottom of the workspace, there is a checkbox for 'Send results to a new history' and a blue 'Run workflow' button. The right panel, titled 'History', shows a list of datasets and workflows, with the 'human exon' dataset highlighted. The history list includes: 1: UCSC Main on Human: refGene (chr22:1-50818468), 2: UCSC Main on Human:.snp147Common (chr22:1-5081841), 3: Join on data 2 and data 1, 4: Count on data 3, 5: Selection des 100 meilleurs SNPs, 6: Compare two Datasets on data 5 and data 1, and the 'human exon' dataset.

Levels of reproducibility

- Reuse: searching for workflows in ToolShed, MyExperiment, Github

The screenshot shows the Galaxy web interface with a dark header bar. The header includes the Galaxy logo, the text "Galaxy / GenOuest", a "Data" button, a "Workflow" button (which is currently selected), "Shared Data", "Visualization", "Help", "User", and a grid icon. A green button on the right says "Using 52%". Below the header is a yellow navigation bar with the text "Import Galaxy workflow". The main content area has three sections:

- Galaxy workflow URL:** A text input field with placeholder text: "If the workflow is accessible via a URL, enter the URL above and click Import." Below the input field is a note: "Aucun fichier sélectionné."
- Galaxy workflow file:** A file selection input field with placeholder text: "If the workflow is in a file on your computer, choose it and then click Import." Below the input field is a large "Import" button.
- Import a Galaxy workflow from myExperiment:** A section containing a link "Visit myExperiment" and a note: "Click the link above to visit myExperiment and browse for Galaxy workflows."

Levels of reproducibility

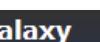
- Reuse: searching for workflows in ToolShed, MyExperiment, Github

The screenshot shows the Galaxy web interface with the following sections:

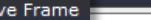
- Import Galaxy workflow:** Includes a "Galaxy workflow URL:" input field and a note: "If the workflow is accessible via a URL, enter the URL above and click Import." A "Import" button is present.
- Galaxy workflow file:** Includes a "Parcourir..." button and a note: "If the workflow is in a file on your computer, choose it and then click Import."
- Import a Galaxy workflow from myExperiment:** Includes a "Visit myExperiment" link and a note: "Click the link above to visit myExperiment and browse for Galaxy workflows."
- Galaxy Workflows:** A search interface with "Search filter terms" and "Filter by type" (checkboxes for Taverna 2, Taverna 1, RapidMiner, Galaxy, KNIME, Kepler, Bioclipse, LONI Pipe, GWorkflow, BioExtract, More...). It also includes a "Type: Galaxy" filter, a search bar, and pagination (1-8).
- Results:** Three workflow cards are displayed:
 - Detrprok (4)** by Clairetn: "In the case study of stranded and prokaryotic RNAseq data, the Det'rprok workflow detects candidates of 3 kinds of non coding RNA: 5'UTRs, antisense RNAs, and small RNAs. Inputs: i) an mapping file (bam format) containing one valid alignment by read, ii) a feature file (gff format) annotating the genomic sequences used for the mapping. Dependencies (from the Galaxy toolshed): "s_mart", "detrprok_scripts""
Created: 2013-05-20 | Last updated: 2015-11-03
 - Basic RNA-Seq Analysis - Differential Expr... (1)** by David De Roure: "From the RNA-Seq analysis tutorial during the Functional Genomics Workshop 2012 https://caps.osu.edu/pfg-workshop Workflow published by mejia-guerra on Galaxy Jun 22, 2012 imported to myExperiment Jul16, 2012 during demonstration of Galaxy-myExperiment integration"
Created: 2012-07-16 | Last updated: 2012-07-16
 - NGS : RNA-seq differential expression anal... (1)** by Ylebras: "Analyse paired-end RNA-seq reads of 2 different individuals with FastQC, Tophat2, Cufflinks and Cuffdiff using a reference annotation in gtf format"
Created: 2014-02-27 | Last updated: 2014-06-10
Credits: Ylebras

Levels of reproducibility

- Reuse: searching for workflows in ToolShed, MyExperiment, Github

Galaxy / GenQuest  **Galaxy**

Return to Galaxy Galaxy Workflows Help

Return to Galaxy Galaxy Workflows Help 

Import Galaxy workflow

Galaxy workflow URL:

If the workflow is accessible via a URL, enter it here.

Galaxy workflow file:
Parcourir... Aucun fichier sélectionné.
If the workflow is in a file on your computer, select it here.

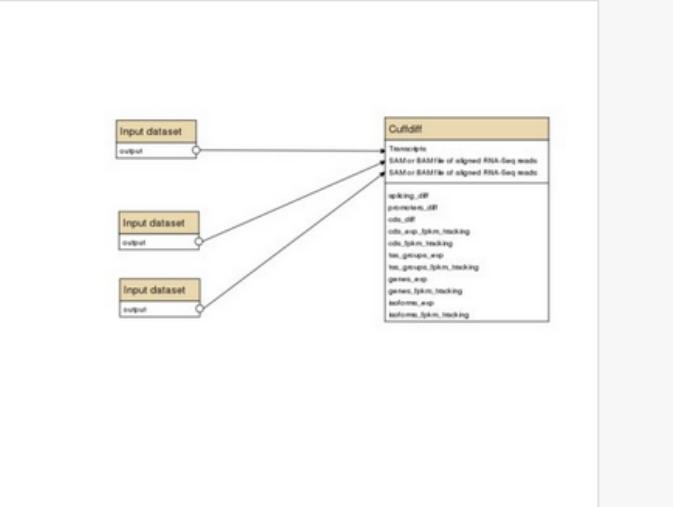
Import

Import a Galaxy workflow from myExperiment
[Visit myExperiment](#)
Click the link above to visit myExperiment and import a workflow.

From the RNA-Seq analysis tutorial during the Functional Genomics Workshop 2012 <https://caps.osu.edu/pfg-workshop>

Workflow published by **mejia-guerra** on Galaxy Jun 22, 2012 imported to **myExperiment** Jul16, 2012 during demonstration of Galaxy-myExperiment integration

Preview



The diagram illustrates a Galaxy workflow. Three "Input dataset" nodes are connected to a single "Cuffdiff" tool node. The "Cuffdiff" node has several output options listed in its configuration panel, including "SAM or BAM file of aligned RNA-seq reads" and "splicing.gff".

[Download as scalable diagram \(SVG\)](#)

Import

Import this workflow into Galaxy

Import

Workflow Type
Galaxy

Uploader

David De Roure

License
All versions of this Workflow are not licensed.

Version 1 (of 1)

Credits (0)
(People/Groups)
None

Attributions (0)
(Workflows/Files)
None

Tags (0)
None
[Log in to add Tags](#)

Shared with Groups (0)
None

Featured In Packs (0)
None
[Log in to add to one of your Packs](#)

Sort by: **Most downloaded** Results per page: **10**

Levels of reproducibility

- Reuse: searching for workflows in ToolShed, MyExperiment, Github

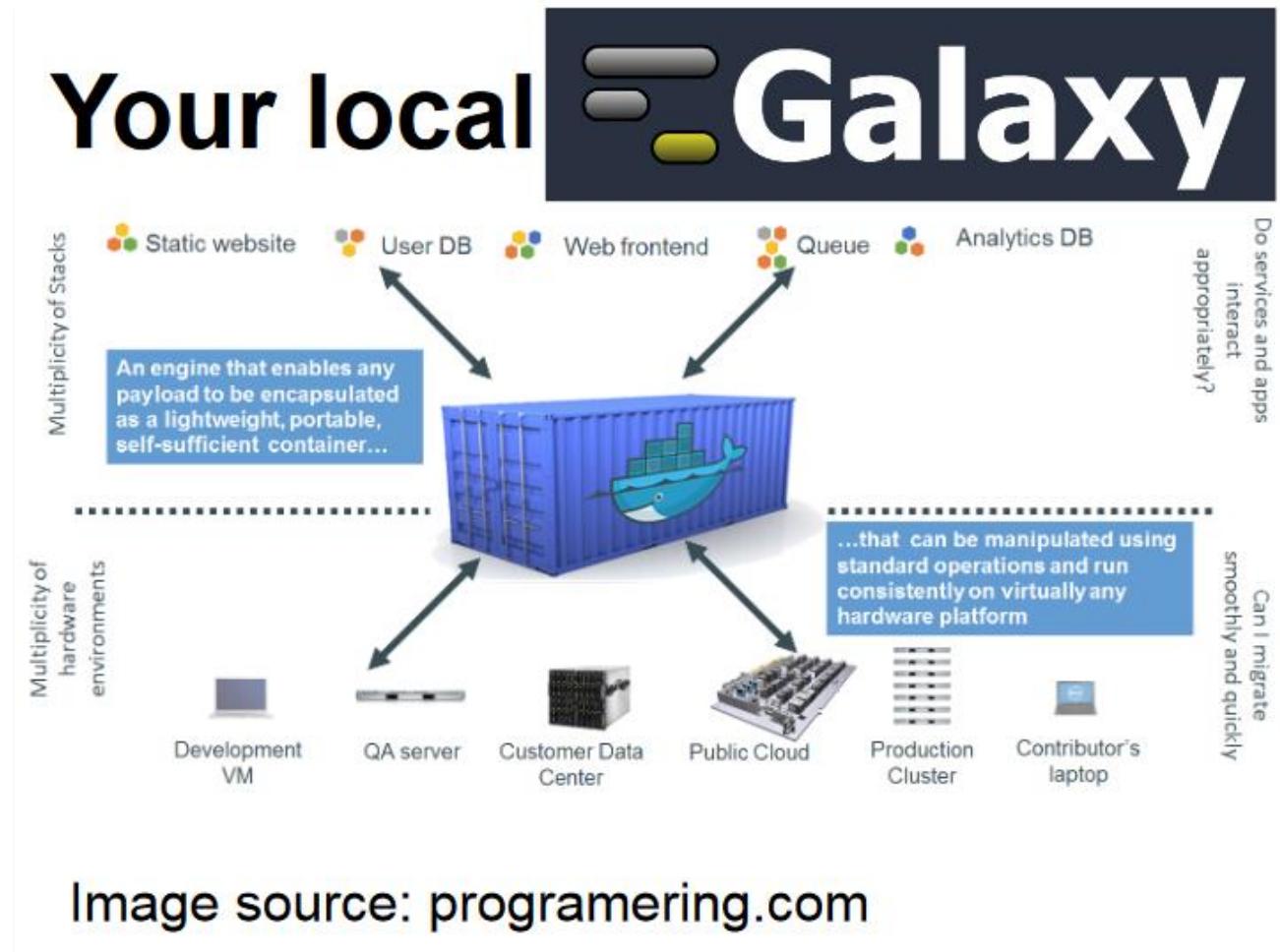
The screenshot shows the Galaxy web interface with the title "Galaxy / GenOuest" and a sub-header "Galaxy / GenOuest Analyze Data". The main navigation bar includes "Workflow", "Shared Data", "Visualization", "Help", and "User". A progress bar at the top right indicates "Using 52%". On the left, there are sections for "Import Galaxy workflow" (with a "Galaxy workflow URL" input field), "Galaxy workflow file" (with a "Parcourir..." button and "Aucun fichier sélectionné" message), and "Import" (with a "Import" button). Below these, under "Import a Galaxy workflow from myExperiment", there is a link "Visit myExperiment" and a note "Click the link above to visit myExperiment". The main content area is titled "Your workflows" and contains a table with the following data:

Name	# of Steps
Basic RNA-Seq Analysis - Differential Expression (Functional Genomics Workshop 2012) (imported from myExperiment)	4
Workflow constructed from history 'human exon'	6
Workflow constructed from history 'Tigriopus californicus transcriptome'	53
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 4: population genomics'	7
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 3: Create mini contig from PE sequences test'	12
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 1 : SNP calling.'	4
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 2: Genetic map'	102
Workflow constructed from history 'STACKS 1.40 GCC Training RAD 5: population genomics with reference genome'	10
bwa	3
Workflow constructed from history 'Tuto Blast'	4
Workflow constructed from history 'RNAseq poulet'	8

On the right side of the interface, there is a sidebar with various search and filter options, including "downloaded", "Results per page: 10", and a "Remove Frame" button.

Levels of reproducibility

- Replicate: Reanalyse data on another system with different config



Galaxy Flavours
Björn Grünig
University of Freiburg



Levels of reproducibility

- Replicate: Reanalyse data on another system with different config

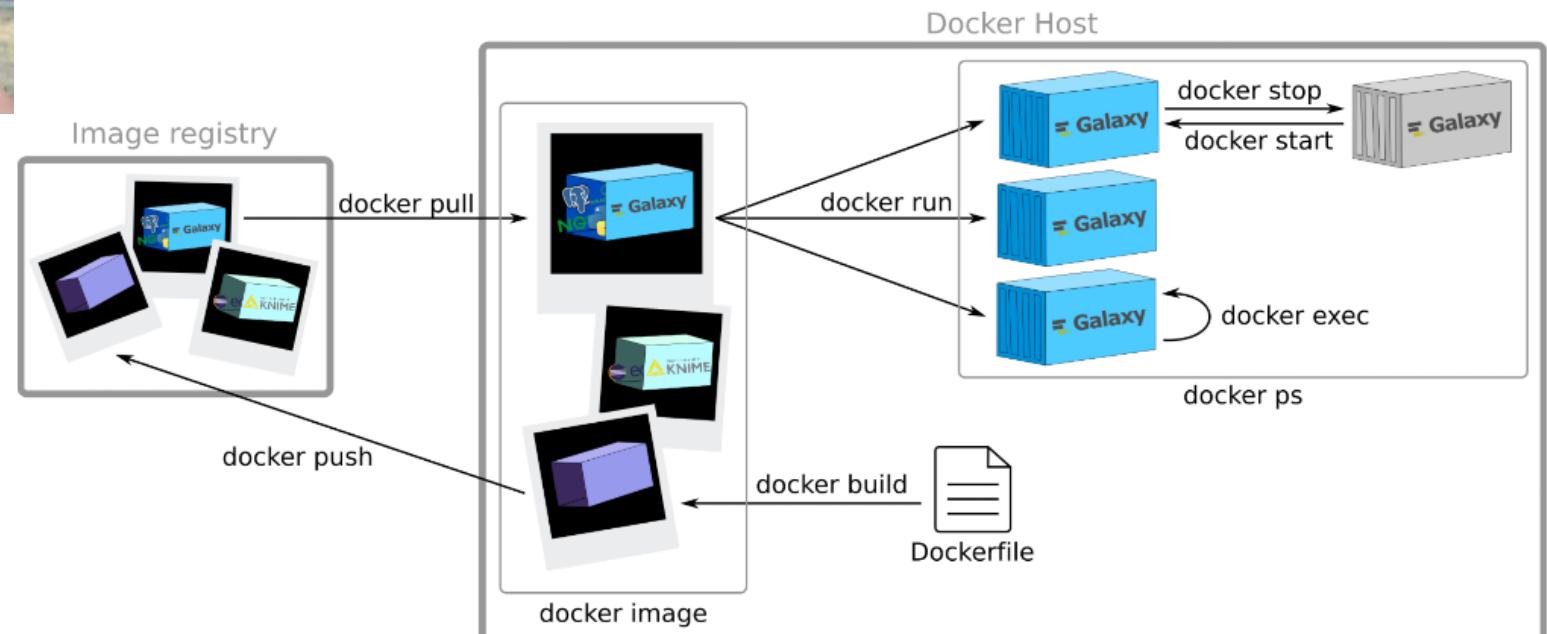
Bérénice Batut
University of Freiburg



Galaxy Flavours
Björn Grünig
University of Freiburg

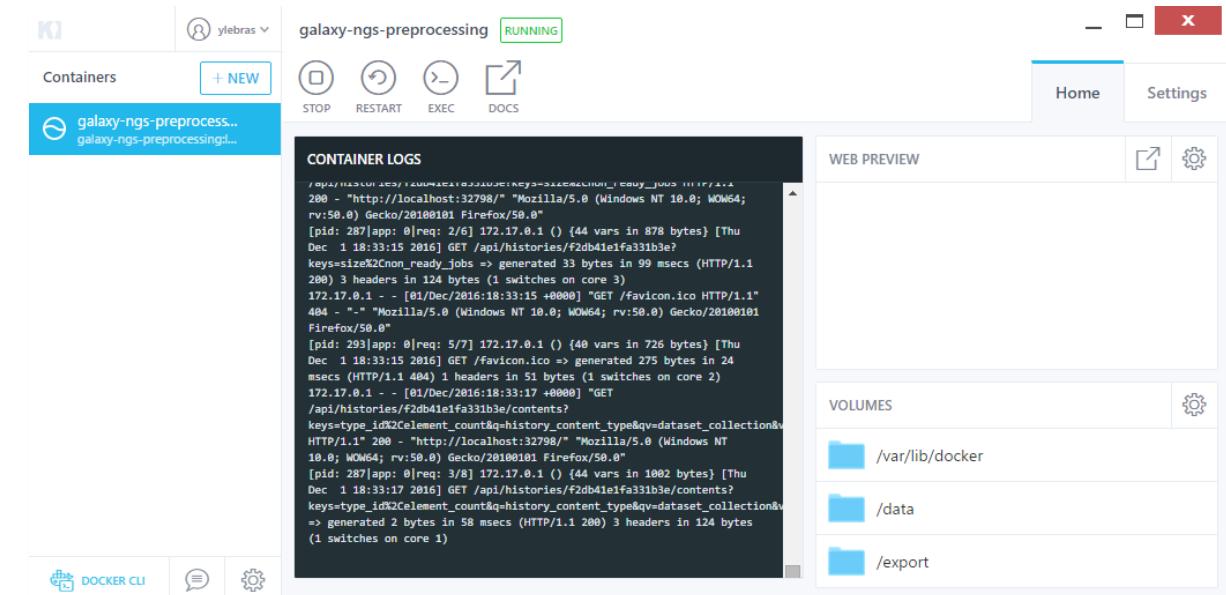
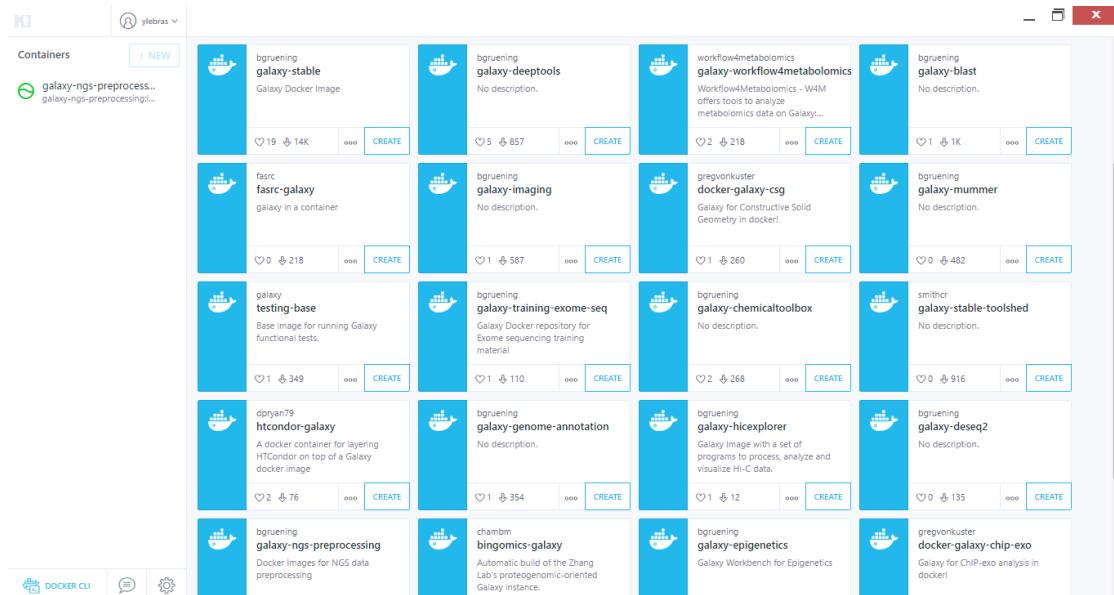


Push your image on a registry



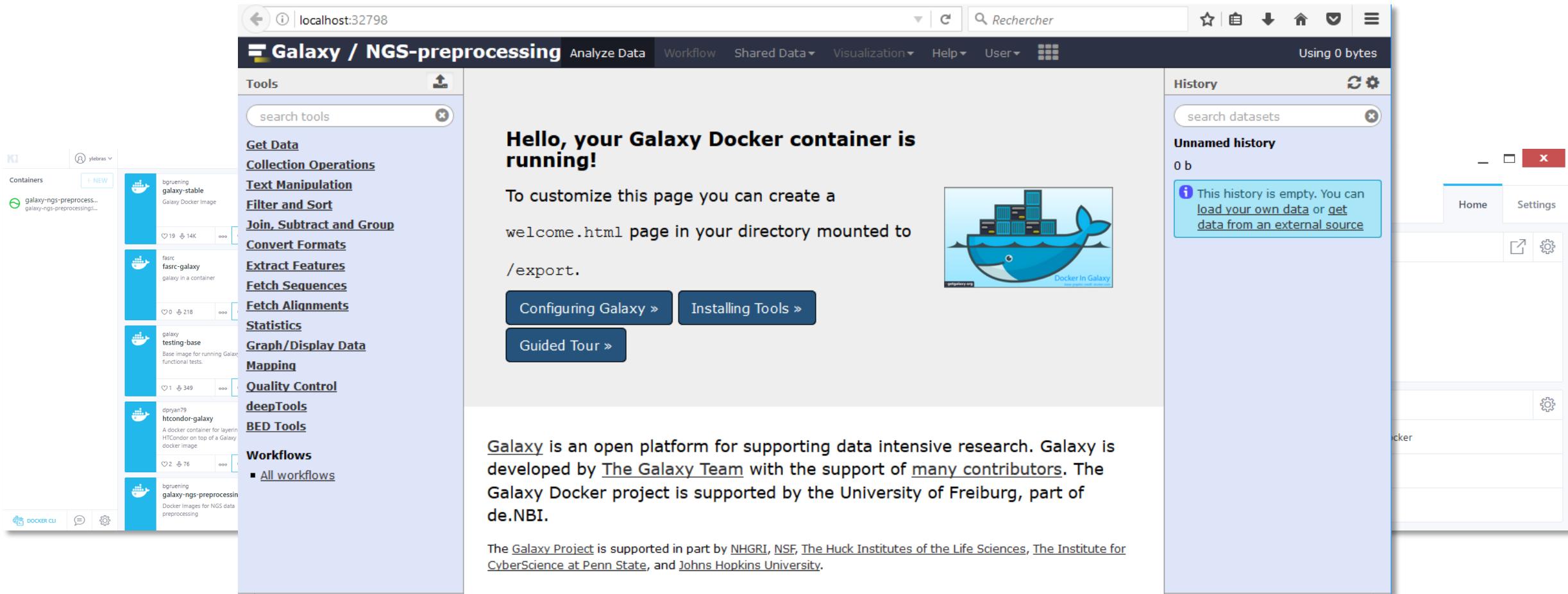
Levels of reproducibility

- Replicate: Reanalyse data on another system with different config



Levels of reproducibility

- Replicate: Reanalyse data on another system with different config



Levels of reproducibility

- Reproduce: Obtain the same conclusions with different Workflows

The screenshot shows the Galaxy web interface with a workflow titled "Workflow constructed from history 'human exon'". The workflow consists of six steps:

- Step 1: Input dataset**: Exons, 6: Compare two Datasets on data 5 and data 1.
- Step 2: Input dataset**: SNPs, 6: Compare two Datasets on data 5 and data 1.
- Step 3: Join** (version 1.0.0).
- Step 4: Count** (version 1.0.2).
- Step 5: Select first** (version 1.0.0).
- Step 6: Compare two Datasets** (version 1.0.2).

The "History" panel on the right shows the following items:

- 5: Selection des 100 meilleures SNPs
- 4: Count on data 3
- 3: Join on data 2 and data 1
- 2: UCSC Main on Human: snp147Common (chr22:1-50818468)
- 1: UCSC Main on Human: refGene (chr22:1-50818468)

The Galaxy interface includes a sidebar with various tools and a main menu at the top.

Levels of reproducibility

- Reproduce: Obtain the same conclusions with different Workflows

The image displays two separate Galaxy web interfaces, each showing the same workflow configuration. Both interfaces have a dark blue header with the Galaxy logo and navigation links: Analyze Data, Workflow, Shared Data, Visualization, Help, User, and a grid icon.

Left Interface:

- Tools:** A sidebar on the left lists various bioinformatics tools categorized under DATA TRANSFER, SYMBIOSE, Metagenomics benchmark, Text Manipulation, Fasta to OTUs, NGS: 454 data Manipulation, Manipulation, Motif, NGS: Alignment, NGS: Assembly, NGS: Genome manipulation, Systems biology, Data management, GenOuest: Mothur toolkit, and KLAST.
- Workflow Details:** The main area shows a workflow titled "Running workflow ""Workflow constructed from history 'human exon'" alternative". It consists of seven steps:
 - Step 1: Input dataset (Exons)
 - Step 2: Input dataset (SNPs)
 - Step 3: Join (version 1.0.0)
 - Step 4: Count (version 1.0.2)
 - Step 5: Sort (version 1.0.3)
 - Step 6: Select first (version 1.0.0)
 - Step 7: Compare two Datasets (version 1.0.2)
- Buttons:** A checkbox for "Send results to a new history" and a "Run workflow" button.

Right Interface:

- Tools:** A sidebar on the left lists the same set of tools as the left interface.
- Workflow Details:** The main area shows the same workflow titled "Running workflow ""Workflow constructed from history 'human exon'" alternative". It consists of seven steps:
 - Step 1: Input dataset (Exons)
 - Step 2: Input dataset (SNPs)
 - Step 3: Join (version 1.0.0)
 - Step 4: Count (version 1.0.2)
 - Step 5: Select first (version 1.0.0)
 - Step 6: Compare two Datasets (version 1.0.2)
- Buttons:** A checkbox for "Send results to a new history" and a "Run workflow" button.
- History:** A panel on the right shows the history of the workflow runs. It includes:
 - 5: Selection des 100 meilleures SNPs (2,425 lines, tabular format, hg38 database)
 - 4: Count on data 3 (Count of unique values in c4)
 - 3: Join on data 2 and data 1 (UCSC Main on Human:.snp147Comm on (chr22:1-50818468))
 - 2: UCSC Main on Human:.snp147Comm on (chr22:1-50818468)
 - 1: UCSC Main on Human:refGene (chr2:1-50818468)

Take home message

- For developpers:
 - Planemo: write tests / tool xml & publish tools & some features:
 - Conda to manage package
 - Travis for continuous integration
 - CWL tools
 - Homebrew for tool requirements
 - Dependencies managed through Conda &/or Docker
- The Galaxy Philosophy: the “biologist” first!
- Galaxy = the best way to reproduce Life sciences **data analysis** processes
- Extremely powerful when integrated with:
 - Data management system
 - Scientific collaboration platform (ie CNRS Core / HUBzero)

The most important Galaxy user is the *bench scientist* using the GUI, they come first!

- No one wants to inconvenience a bioinformatician or developer, but we will if absolutely required for the biologist.

} Virtual Research Environments

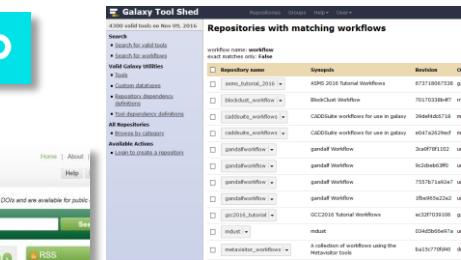
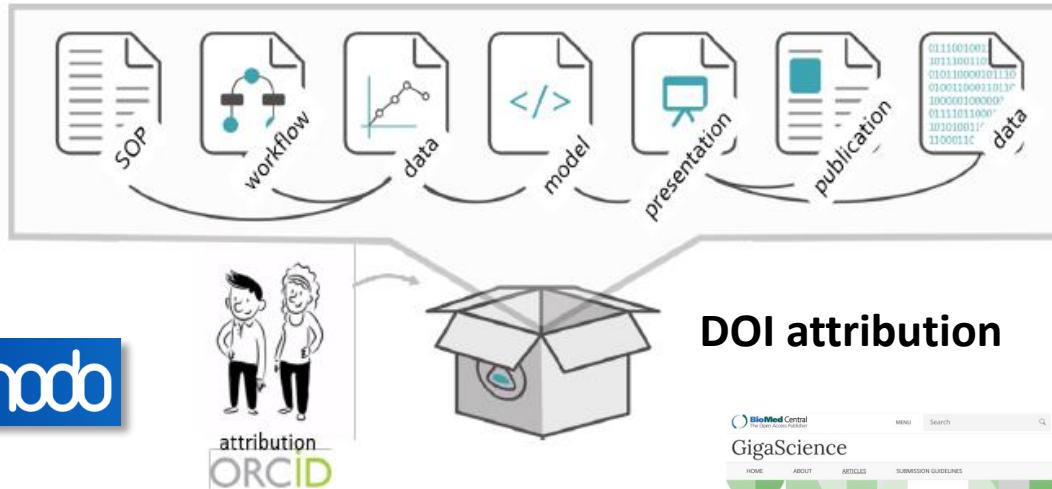
Take home message

Publish your research objects!!!

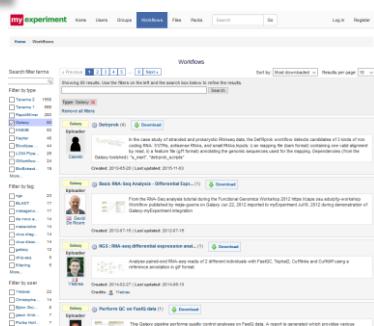
10 “Simple” Rules for Reproducible Computational Research: RACE

1. For Every Result, Keep Track of How It Was Produced
2. Avoid Manual Data Manipulation Steps
3. Archive the Exact Versions of All External Programs Used
4. Version Control All Custom Scripts
5. Record All Intermediate Results, When Possible in Standardized Formats
6. For Analyses That Include Randomness, Note Underlying Random Seeds
7. Always Store Raw Data behind Plots
8. Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected
9. Connect Textual Statements to Underlying Results
10. Provide Public Access to Scripts, Runs, and Results

Record Everything
Automate Everything
Contain Everything
Expose Everything



DOI attribution



Initiative: GigaScience & related GigaDB

Aims and scope

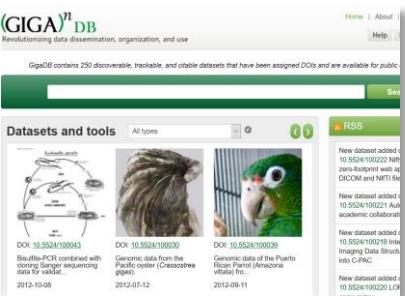
“*GigaScience* aims to revolutionize reproducibility of analyses, data dissemination, organization, understanding, and use. As an **open access** and **open-data** journal, we publish all research objects (data, software tools and workflows) from 'big data' studies across the entire spectrum of life and biomedical sciences...”



Open Science

“...Our scope covers **not just 'omic' type data** and the fields of high-throughput biology currently serviced by large public repositories, **but also the growing range of more difficult-to-access data**, such as imaging, neuroscience, ecology, cohort data, systems biology and other new types of large-scale sharable data.”

Initiative: GigaScience & related GigaDB



GigaDB: Latest Datasets

GigaScience can **host all relevant data and tools** from articles published in the journal in its affiliated database, [GigaDB*](#) under a public domain [CC0 waiver](#). All datasets hosted in [GigaDB](#) are assigned a **DOI** that allows data citation**. [GigaDB](#) can also provide hosting and DOIs for **Software**, **Containers** and **Workflows** integrated into our [GigaGalaxy server](#), making research more reproducible and easier to build on.

Data and Tools through DOI

No unique entry point

*Submission of data to *GigaScience* for hosting by GigaDB **does not serve as a replacement for community-approved public repositories**. Where possible, all supporting data and code should also be made available in a suitable public repository.

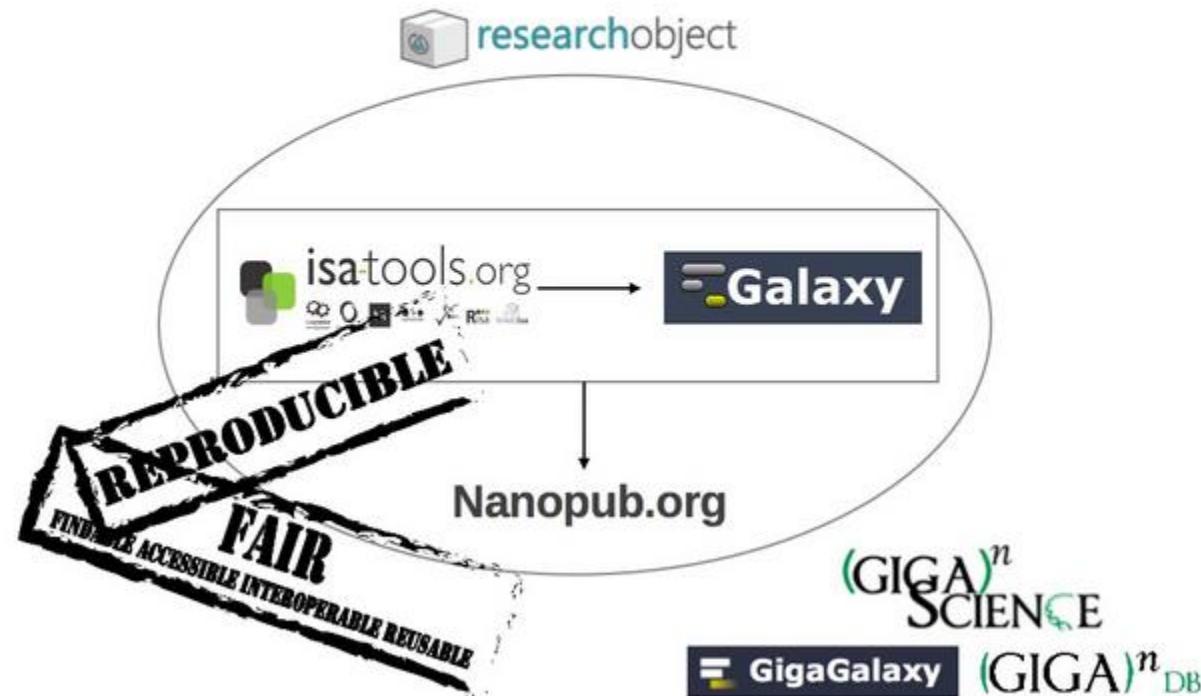
GigaDB is tracked by the [Data Citation Index](#). “The **Data Citation Index on the Web of Science provides a single point of access to quality research data from repositories across disciplines and around the world.”

Initiative: GigaScience & related GigaGalaxy

The screenshot shows the GigaGalaxy web interface. The top navigation bar includes tabs for "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "User". A search bar labeled "Rechercher" is also present. The main content area displays a message: "GigaGalaxy hosts tools and data analyses from papers published in *GigaScience*". Below this message is a banner for "(GIGA)ⁿ DB" with the text: "Promoting reproducible research" and "GigaScience's integrated 'big-data' repository containing citable data available for public download and use". On the left side, there is a sidebar titled "Tools" with a search bar and a list of links: "Get Data", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Genome Diversity", "BGISOAP", "SOAPdenovo2", "SOAPdenovo1", and "Smilefinder". At the bottom of the sidebar, there is a section titled "(GIGA)ⁿ DB" with a sub-section titled "Datasets and tools". The right side of the interface features a "History" panel with a search bar and a message: "This history is empty. You can load your own data or get data from an external source". A green "DOI" button is visible on the far right.

Initiative: ISA, GigaScience & related GigaGalaxy

Can data models and computational workflows help in capturing the experimental processes and reproduce findings?
How?





Thanks to Galaxy

....to support effectively Accessibility, Reproducibility and Transparency!



Anton Nekrutenko
Galaxy Project, Penn State University



Daniel Blankenberg
Galaxy Project, Penn State University



Björn Grüning
University of Freiburg



Dannon Baker
Galaxy Project, Johns Hopkins University



Dave Bouvier
Galaxy Project, Penn State University



Dave Clements
Galaxy Project, Johns Hopkins University
Training and Outreach Coordinator



Enis Afgan
Galaxy Project, Johns Hopkins University



James Taylor
Johns Hopkins University



Jeremy Goecks
Galaxy Project, George Washington University



John Chilton
Galaxy Project, Penn State University



Martin Čech
Galaxy Project, Penn State University



Nitesh Turaga
Galaxy Project, Johns Hopkins University
Software Engineer



Nate Coraor
Galaxy Project, Penn State University

& so many others.....

<http://cesco.mnhn.fr/>

PIA 65 Millions d'observateurs



<https://www.france-bioinformatique.fr/>



contact@engineson.fr





Thanks for your Attention

Thank to Galaxy to support effectively Accessibility, Reproducibility and Transparency!



[The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update.](#) *Nucleic Acids Research* (2016) doi: 10.1093/nar/gkw343

Citing Specific Galaxy Components / Features

Cite these papers if you want to cite a particular aspect of Galaxy.

Application Programming Interface (API)

Clare Sloggett, Nuwan Goonasekera and [Enis Afgan](#), "BioBlend: automating pipeline analyses within Galaxy and CloudMan", *BMC Bioinformatics* 2013

Cloud

[Enis Afgan](#), Brad Chapman and [James Taylor](#), "CloudMan as a platform for tool, data, and analysis distribution", *BMC Bioinformatics* 2012, 13:315

Data Managers

[Daniel Blankenberg](#), James E. Johnson, The [Galaxy Team](#), [James Taylor](#) and [Anton Nekrutenko](#), "Wrangling Galaxy's Reference Data," in *Bioinformatics* (2014) doi: 10.1093/bioinformatics/btu119

Interactive Environments

[Björn Grüning](#), [Eric Rasche](#), Boris Rebolledo-Jaramillo, [Carl Eberhard](#), Torsten Houwaart, [John Chilton](#), [Nate Coroar](#), Rolf Backofen, [James Taylor](#), [Anton Nekrutenko](#). [Enhancing pre-defined workflows with ad hoc analytics using Galaxy, Docker and Jupyter](#), *bioRxiv*, 2016 DOI: 10.1101/075457

Reproducibility

Geir Kjetil Sandve, [Anton Nekrutenko](#), [James Taylor](#), Eivind Hovig, "Ten Simple Rules for Reproducible Computational Research," in *PLoS Computational Biology* 9(10): e1003285. doi:10.1371/journal.pcbi.1003285

ToolShed

[Daniel Blankenberg](#), [Gregory Von Kuster](#), [Emil Bouvier](#), Dannon Baker, [Enis Afgan](#), [Nicholas Stoler](#), the [Galaxy Team](#), [James Taylor](#) and [Anton Nekrutenko](#), "Dissemination of scientific software with Galaxy ToolShed," in *Genome Biology* 2014, 15:403, doi:10.1186/gb4161

Visualization

[Jeremy Goecks](#), [Carl Eberhard](#), Tomithy Too, [Anton Nekrutenko](#) and [James Taylor](#), "Web-based visual analysis for high-throughput genomics," *BMC Genomics* 2013, 14:397

<http://cesco.mnhn.fr/>

PIA 65 Millions d'observateurs



<https://www.france-bioinformatique.fr/>



contact@engineson.fr

