

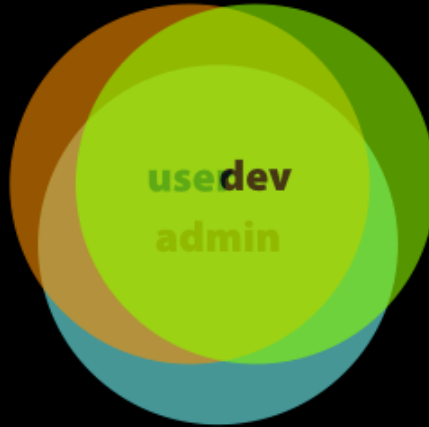
Galaxy Community Update

2015

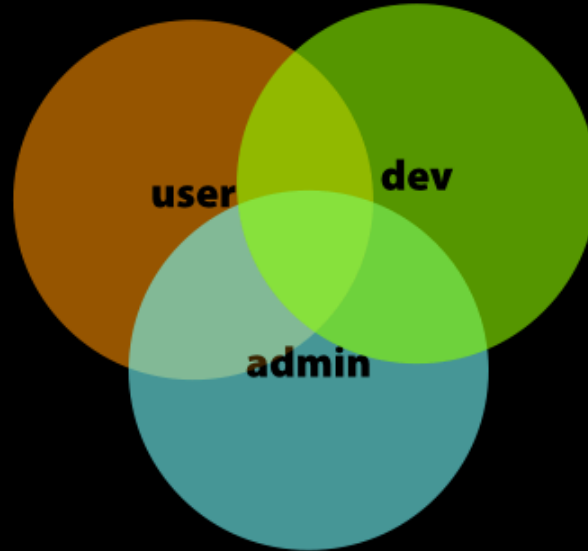


#usegalaxy

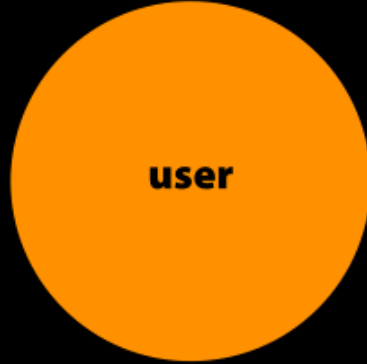
Community



Community



Community



The Galaxy Developer Community

All Galaxy development moved to **github**

github.com/galaxyproject

~30 repositories

June 6, 2015 – July 6, 2015

Period: 1 month ▾

Overview

100 Active Pull Requests

11 Active Issues

82

Merged Pull Requests

18

Proposed Pull Requests

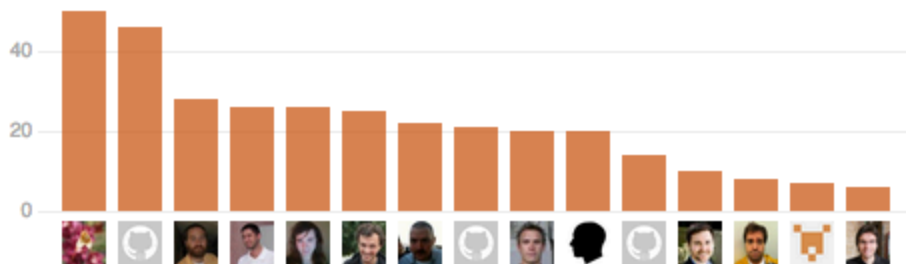
1

Closed Issue

10

New Issues

Excluding merges, **47 authors** have pushed **207 commits** to dev and **417 commits** to all branches. On dev, **462 files** have changed and there have been **13,731 additions** and **7,664 deletions**.



82 Pull requests merged by 22 people

Policy for code contribution to Galaxy core

<https://github.com/galaxyproject/galaxy/blob/dev/CONTRIBUTING.md>

Contribution Policy

(highlights)

1. All changes submitted as pull requests
2. A pull request must have two +1 votes by members of the committers group
3. ...and no -1 votes by members of the committers group

<https://github.com/galaxyproject/galaxy/blob/dev/CONTRIBUTING.md>

Organization

Committers: “trusted developers and advocates who manage the core Galaxy code base”

Committers can be added by existing committers following the contribution rules

All committers have equal power

<https://github.com/galaxyproject/galaxy/blob/dev/doc/source/project/organization.rst>

Introducing... the initial cohort of Galaxy committers

Enis Afgan (@afgane)

Dannon Baker (@dannon)

Daniel Blankenberg
(@blankenberg)

Dave Bouvier (@davebx)

Martin Čech (@martenson)

John Chilton (@jmchilton)

Dave Clements (@tnabtaf)

Nate Coraor (@natefoo)

Carl Eberhard (@carlfeberhard)

Jeremy Goecks (@jgoecks)

Björn Grüning (@bgruening)

Aysam Guerler (@guerler)

Jennifer Hillman Jackson (@jennaj)

Ross Lazarus (@fubar2)

Anton Nekrutenko (@nekrut)

Eric Rasche (@erasche)

Nicola Soranzo (@nsoranzo)

James Taylor (@jxtx)

Nitesh Turaga (@nitesh1989)

Introducing... the initial cohort of Galaxy committers

Enis Afgan (@afgane)

Dannon Baker (@dannon)

Daniel Blankenberg
(@blankenberg)

Dave Bouvier (@davebx)

Martin Čech (@martenson)

John Chilton (@jmchilton)

Dave Clements (@tnabtaf)

Nate Coraor (@natefoo)

Carl Eberhard (@carlfeberhard)

Jeremy Goecks (@jgoecks)

Björn Grüning (@bgruening)

Aysam Guerler (@guerler)

Jennifer Hillman Jackson (@jennaj)

Ross Lazarus (@fubar2)

Anton Nekrutenko (@nekrut)

Eric Rasche (@erasche)

Nicola Soranzo (@nsoranzo)

James Taylor (@jxtx)

Nitesh Turaga (@nitesh1989)

Supporting tool developer community

planemo

proliferation of tool development

future of toolshed

PLANEMO

docs

latest

pypi package

0.13.2

build

passing

coverage

83%

Command-line utilities to assist in building and publishing [Galaxy](#) tools.

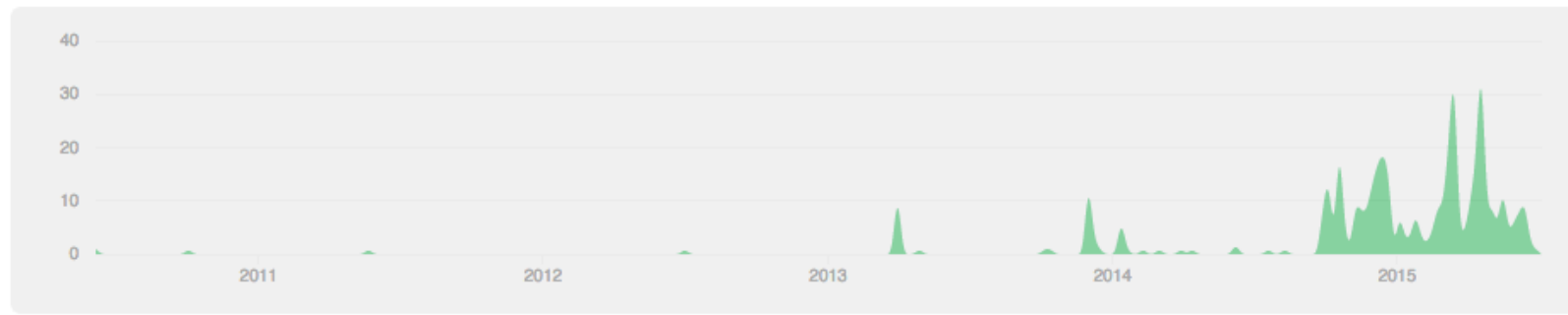
- Free software: Academic Free License version 3.0
- Documentation: <https://planemo.readthedocs.org>.
- Code: <https://github.com/galaxyproject/planemo>

dev

Jun 6, 2010 – Jul 6, 2015

Contributions: **Commits** ▾

Contributions to master, excluding merge commits

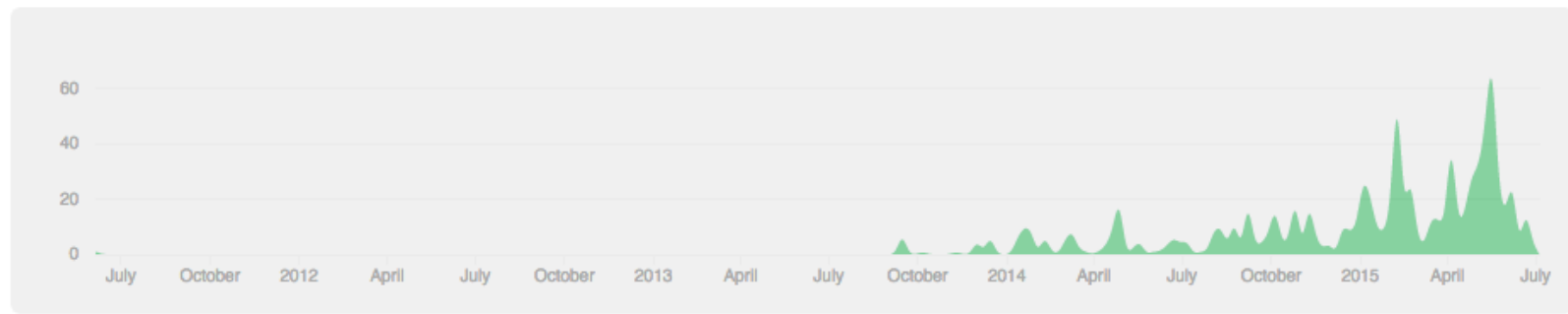


iuc

Jun 5, 2011 – Jul 6, 2015

Contributions: **Commits** ▾

Contributions to master, excluding merge commits



Supporting Galaxy Deployers

Unification and automation of all Galaxy
deployment

Ansible playbooks for deploying Galaxy from
base OS up – *same ones we use for main!*

As much as possible, reuse playbooks between
main, cloud, ...

Tools to build machine images, Docker containers

Galaxy UI Highlights

analyzing *many* datasets

Histories

https://usegalaxy.org/history/view_multiple

☆

ABP

☰

Galaxy

Analyze Data Workflow Shared Data Lab Visualization Cloud Help User

Using 31.5 GB

Done Include deleted histories Order histories by... search histories search all datasets Include deleted datasets Include hidden datasets

Create new Delete Copy Switch to Delete Copy Switch to Delete Copy Switch to

Coursera RNA-seq full

54 shown, 24 hidden

447.2 MB

search datasets

58: Cuffdiff on data 16, data 11, and others: transcript FPKM tracking

57: Cuffdiff on data 16, data 11, and others: transcript differential expression testing

56: Cuffdiff on data 16, data 11, and others: gene FPKM tracking

55: Cuffdiff on data 16, data 11, and others: gene differential expression testing

54: Cuffdiff on data 16, data 11, and others: TSS groups FPKM tracking

53: Cuffdiff on data 16, data 11, and others: TSS groups differential expression testing

52: Cuffdiff on data 16, data 11, and others: CDS FPKM tracking

51: Cuffdiff on data 16, data 11, and others: TSS groups FPKM tracking

Imported: Exons and Repeats

7 shown

3.5 MB

search datasets

7: UCSC Main on Human: cpGISland Ext (chr22:1-51304566)

6: Cut on data 5

5: Join two Datasets on data 4 and data 1

4: Group on data 3

3: Join on data 2 and data 1

2: UCSC Main on Human: rmsk (chr22:1-51304566)

1: UCSC Main on Human: knownGene (chr22:1-51304566)

Exons and CpG Islands

6 shown

1.1 MB

search datasets

6: Cut on data 5

5: Join two Datasets on data 4 and data 1

4: Group on data 3

3: Join on data 7 and data 1

2: UCSC Main on Human: cpGISland Ext (chr22:1-51304566)

1: UCSC Main on Human: knownGene (chr22:1-51304566)

Exons and Repeats

7 shown

3.5 MB

search datasets

7: UCSC Main on Human: cpGISland Ext (chr22:1-51304566)

6: Cut on data 5

5: Join two Datasets on data 4 and data 1

4: Group on data 3

3: Join on data 7 and data 1

2: UCSC Main on Human: cpGISland Ext (chr22:1-51304566)

1: UCSC Main on Human: knownGene (chr22:1-51304566)

https://usegalaxy.org/history/view_multiple


search datasets 

Isolates
8 shown




356 bytes   


- 8: [isolate-1141_2.fastq](#)   
- 7: [isolate-1141_1.fastq](#)   
- 6: [isolate-1140_2.fastq](#)   
- 5: [isolate-1140_1.fastq](#)   
- 4: [isolate-1139_2.fastq](#)   
- 3: [isolate-1139_1.fastq](#)   
- 2: [isolate-1138_2.fastq](#)   
- 1: [isolate-1138_1.fastq](#)   

Individual Datasets

search datasets 

Isolates
1 shown, 8 [hidden](#)

356 bytes   

9: Isolates-exp1 
a list of paired datasets

Collection

Dataset

9: Isolates-exp1
a list of paired datasets

isolate-1141
a pair of datasets

forward

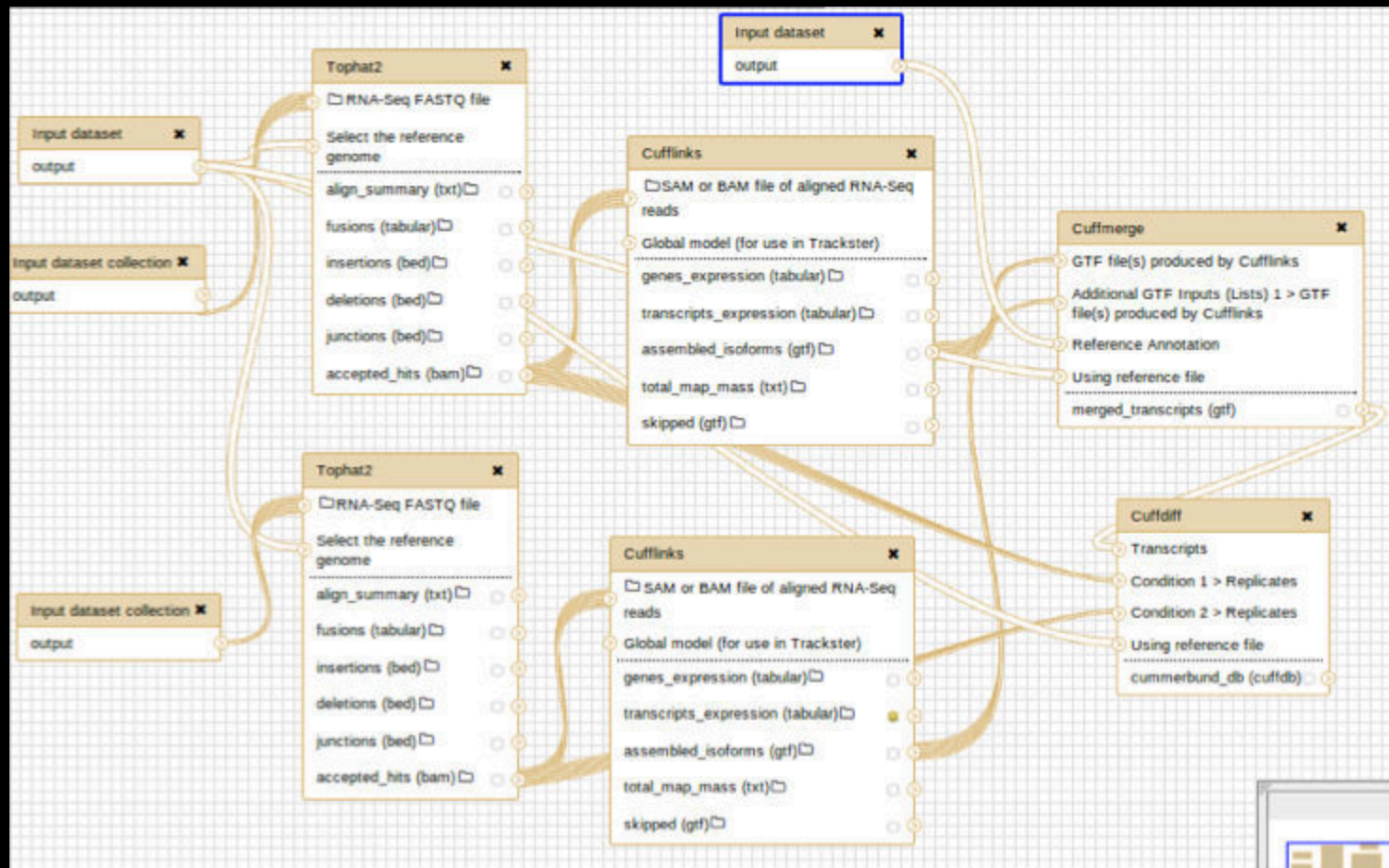
reverse

isolate-1140
a pair of datasets

isolate-1139
a pair of datasets

isolate-1138
a pair of datasets

Collection Contents



visualization plugin framework

- Area charts



🔍 Regular (NVD3)



🔍 Expanded (NVD3)



🔍 Stream (NVD3)



🔍 Pie chart (NVD3)

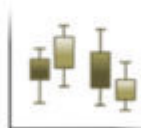
- Data processing (requires 'charts' tool from Toolshed)



Histogram (NVD3)



🔍 Discrete Histogram (jqPlot)



🔍 Box plot (jqPlot)



🔍 Clustered Heatmap (Custom)

• Area charts



Regular (NVD3)



Regular (NVD3)



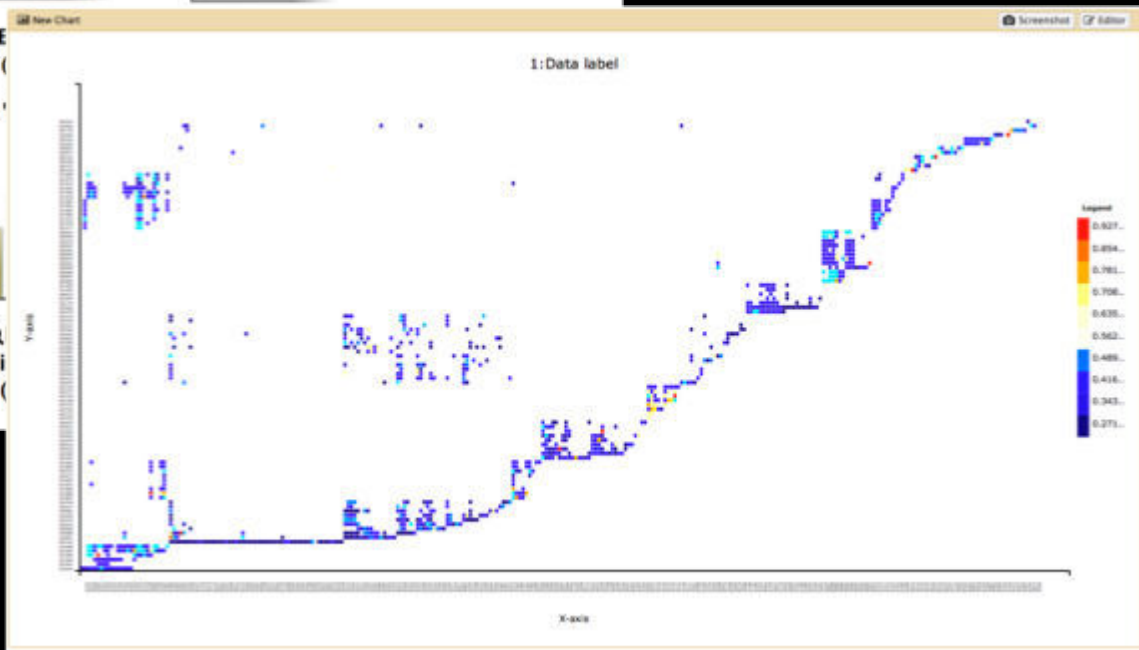
• Data processing (requires)



Histogram
(NVD3)



Histogram
(NVD3)



• Area charts



Regular (NVD3)



Regular (NVD3)



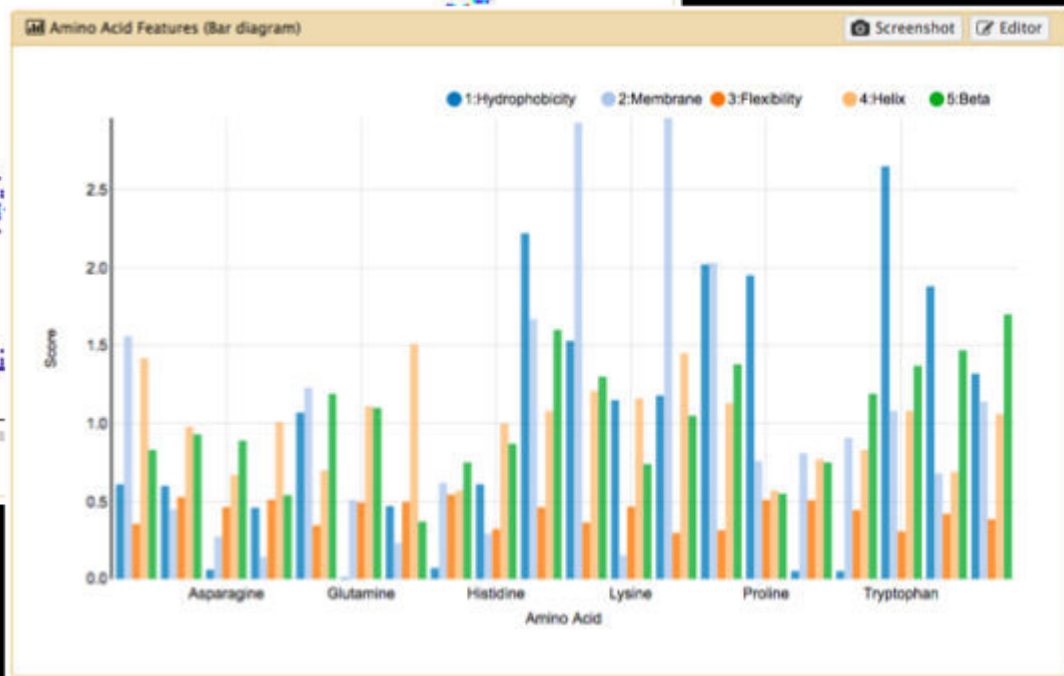
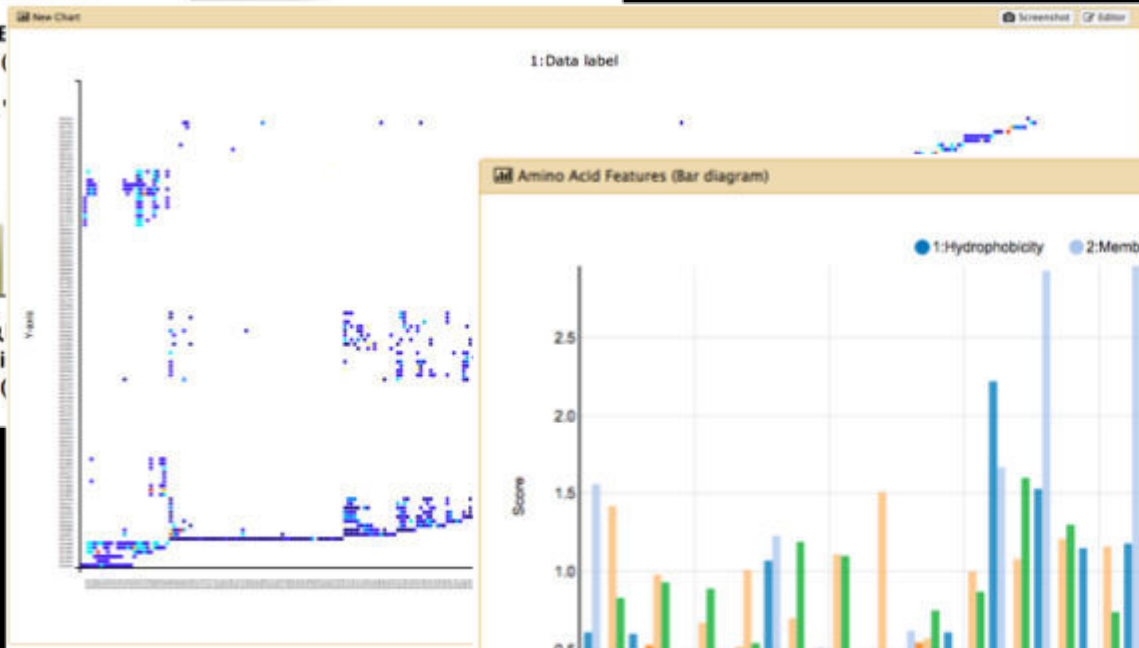
• Data processing (requires)



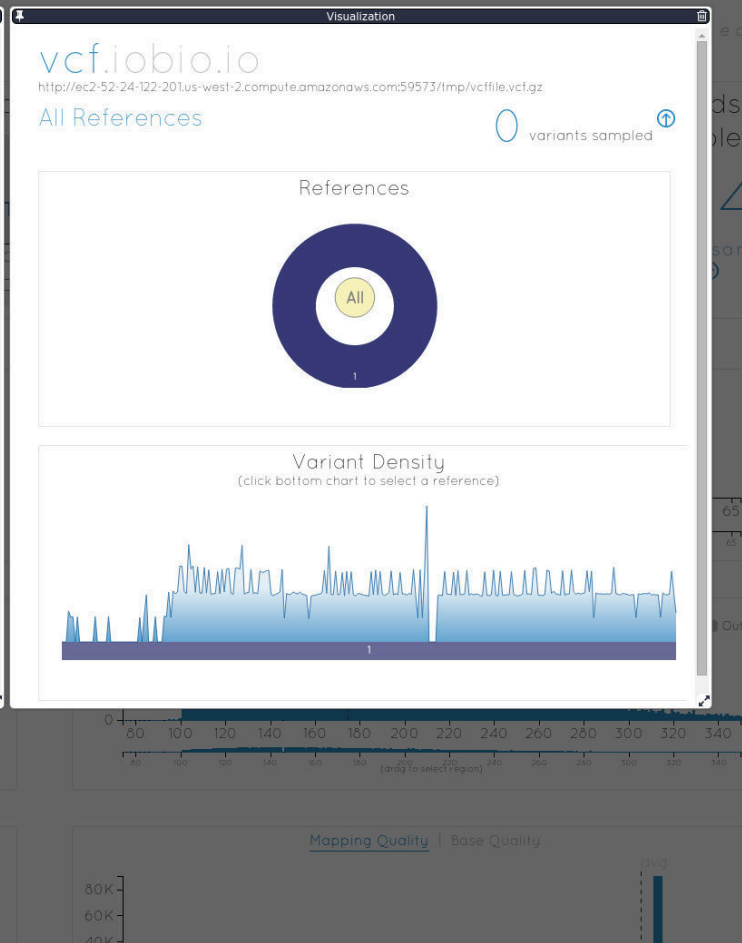
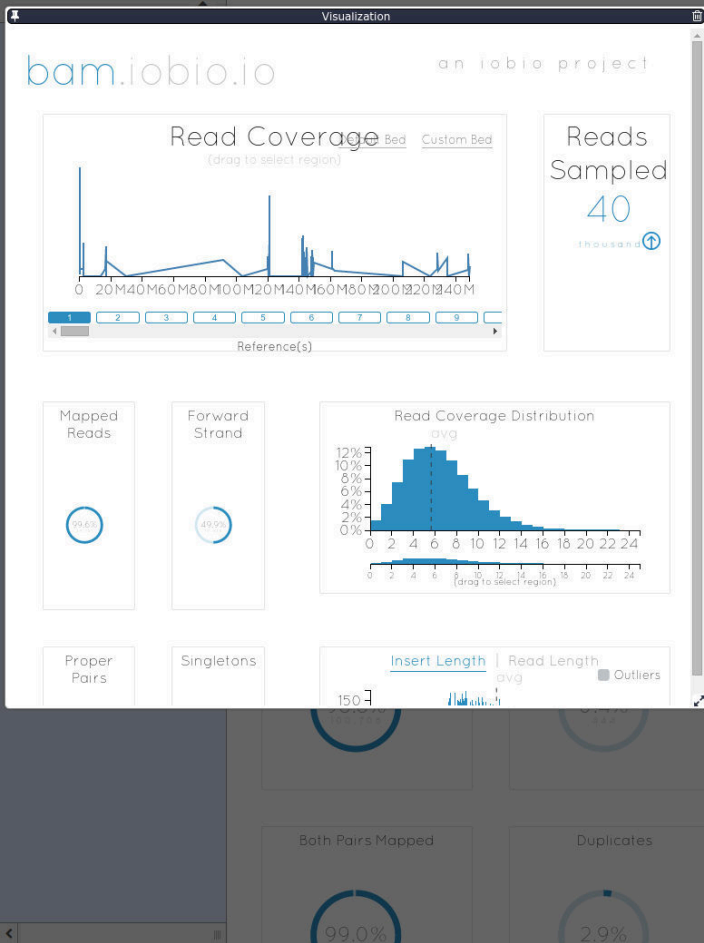
Histogram
(NVD3)



Histogram
(NVD3)



interactive environments



```
1.Chrom 2.Pos 3.ID 4.Ref 5.Alt 6.Qual
##fileformat=VCFv4.0
##fileDate=20150416
##source=dbSNP
##dbSNP_BUILD_ID=142
##reference=GRCh38
##phasing=partial
```



IP[y]: Notebook ipython_galaxy_notebook (autosaved)

File

Edit

View

Insert

Cell

Kernel

Help



Heading 1 ▾

Cell Toolbar: None ▾



Welcome to the interactive Galaxy IPython Notebook.

You can access your data via the dataset number. For example, `handle = get(42)`. To save data, write your data to a file, and then call `put('filename.txt')`. The dataset will then be available in your galaxy history. Notebooks can be saved to Galaxy by clicking the large green button at the top right of the IPython interface.

More help and informations can be found on the project [website](#).

In [1]:

History



search datasets



RNAseqTry2

61 shown, 11 [deleted](#)

857.9 MB



72: cummeRbund database for data 60, data 1, and others



897,201 lines

format: **cuffdb**, database: ?

[23:00:48] Loading reference annotation and sequence.

[23:00:52] Inspecting maps and determining fragment length distributions.

[23:01:39] Modeling fragment count overdispersion.

[23:03:12] Modeling fragment count overdispersion.

> Map Properties:

> Norm



bin

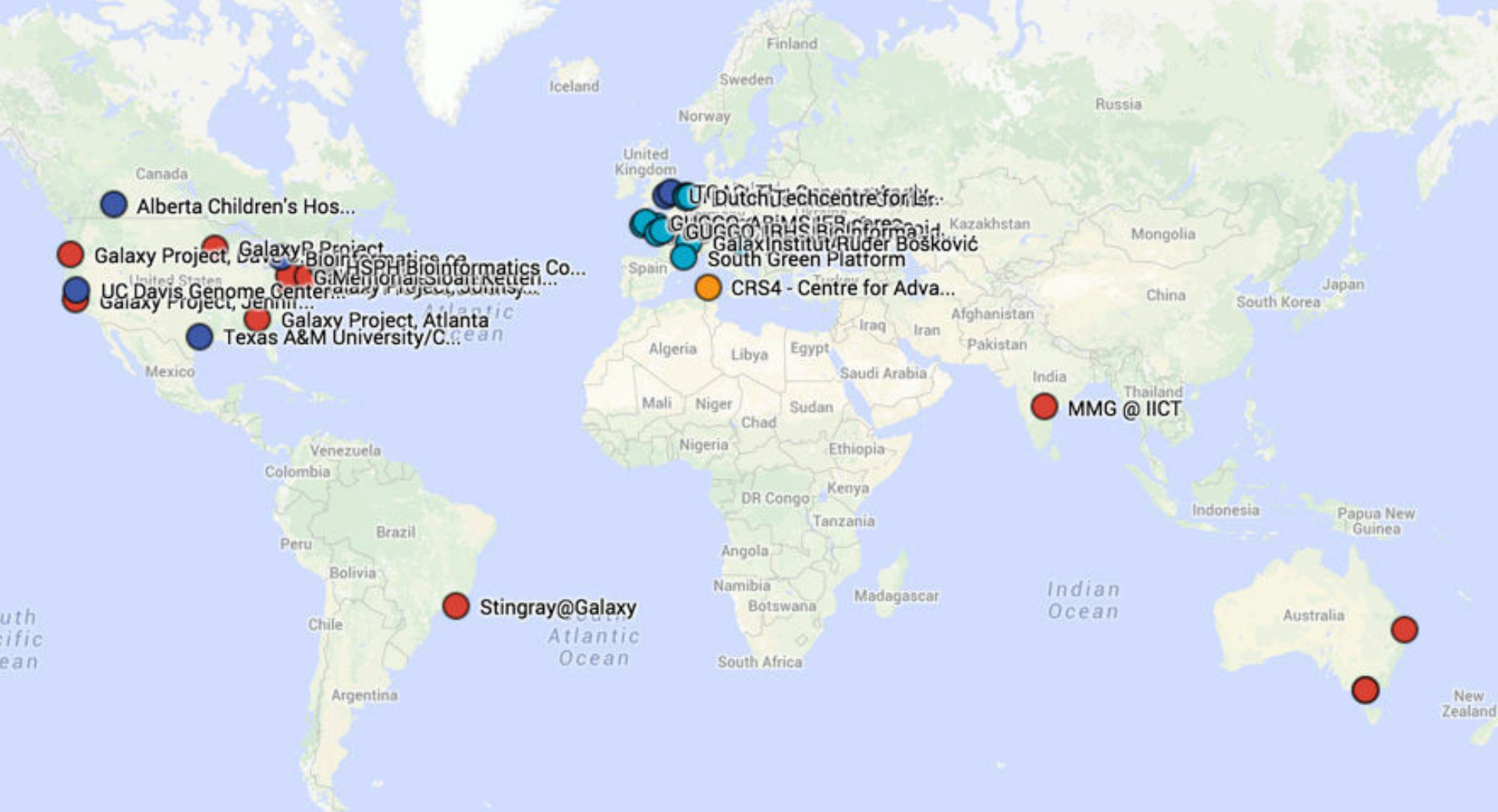
Visualize in IPython

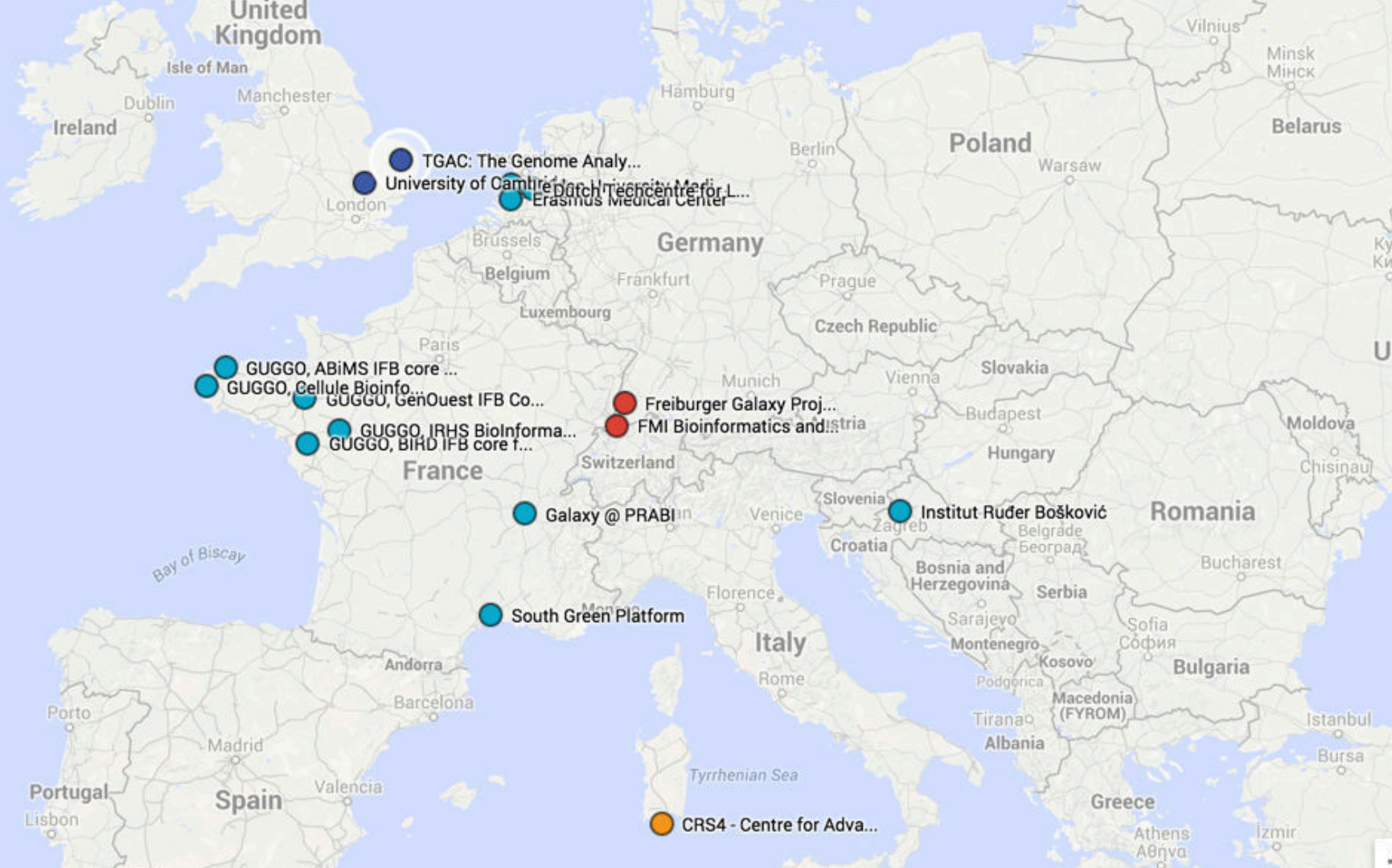
60: Cuffmerge on data 1, data 28, and others:



Training and Outreach







Publishing biology

Group: Galaxy - library 2337 articles 

Tags

 Search  Copy  Export  Sort  Hide Details

✓ Deciphering the Epigenetic Code of Cardiac Myocyte Transcription

Circulation Research (23 June 2015), doi:10.1161/circresaha.115.306337

by [Sebastian Preissl](#), [Martin Schwaderer](#), [Alexandra Raulf](#), et al.

posted to [no-tag](#) by [bgruening](#) to the group [Galaxy](#) on 2015-06-24 21:16:48 ★★

■ [Abstract](#)

✓ Adaptation of the targeted capture Methyl-Seq platform for the mouse genome identifies novel tissue-specific DNA methylation patterns of genes involved in neurodevelopment

Epigenetics (18 May 2015), pp. 00-00, doi:10.1080/15592294.2015.1045179

by [Benjamin Hing](#), [Enrique Ramos](#), [Patricia Braun](#), et al.

posted to [methods](#) by [galaxyproject](#) to the group [Galaxy](#) on 2015-05-28 21:46:38 ★★

■ [Abstract](#)

✓ Genomic and experimental evidence for multiple metabolic functions in the RidA/YjgF/YER057c/UK114 (Rid) protein family

BMC Genomics, Vol. 16, No. 1. (15 May 2015), 382, doi:10.1186/s12864-015-1584-3

by [Thomas D. Niehaus](#), [Svetlana Gerdes](#), [Kelsey Hodge-Hanson](#), et al.

posted to [methods](#) [usemain](#) by [galaxyproject](#) to the group [Galaxy](#) on 2015-05-28 21:41:14 ★★

■ [Abstract](#)

✓ NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data

Nat. Protocols, Vol. 10, No. 6. (07 June 2015), pp. 823-844, doi:10.1038/nprot.2015.052

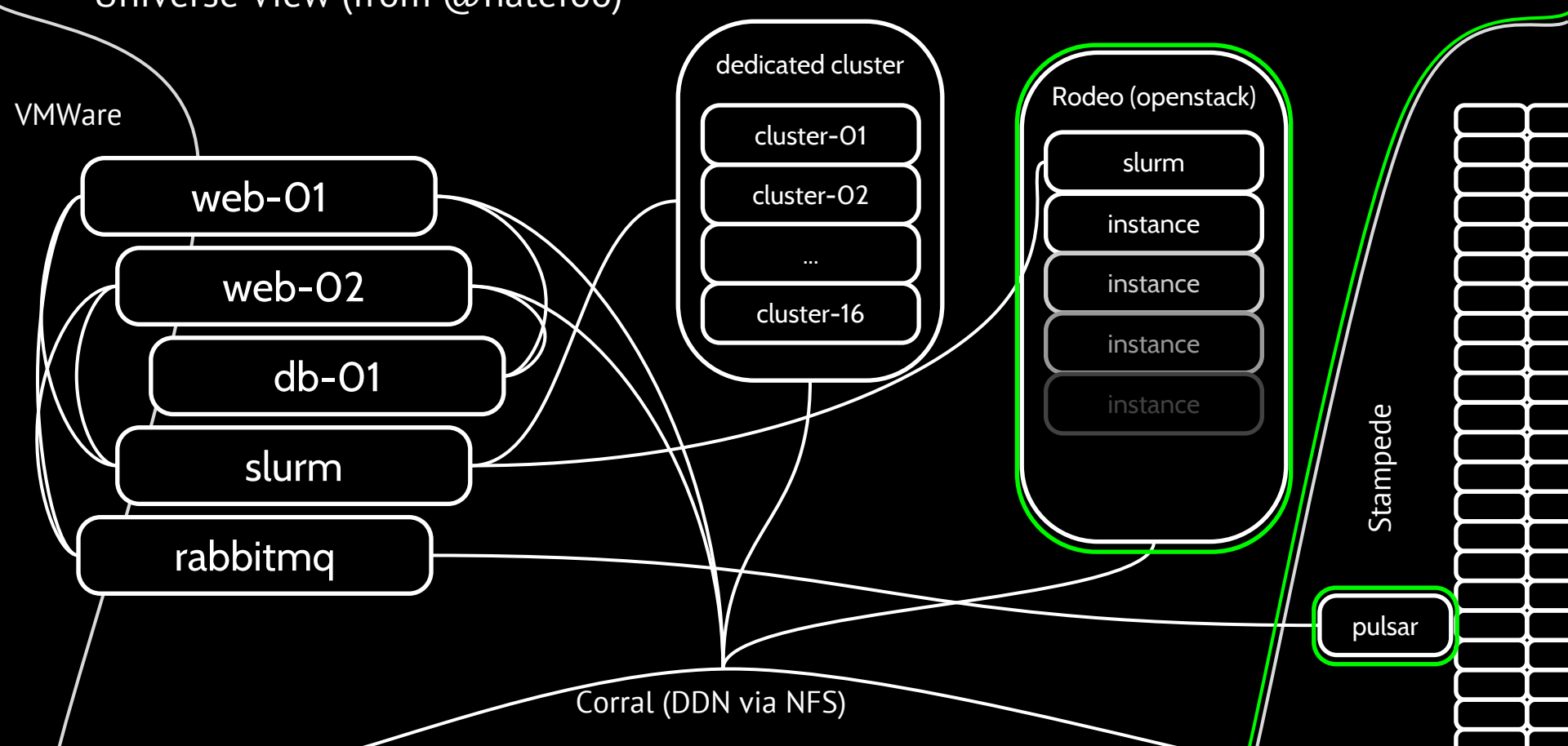
by [Jianguo Xia](#), [Erin E. Gill](#), [Robert E. W. Hancock](#)

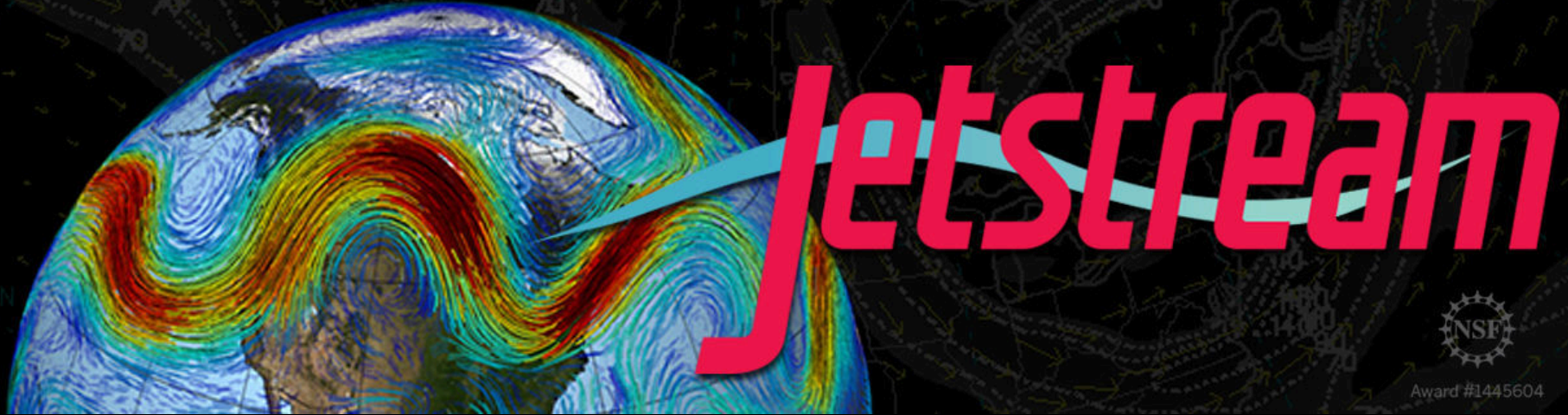
posted to [visualization](#) by [galaxyproject](#) to the group [Galaxy](#) on 2015-05-28 21:37:43 ★★ [along with 2 people and 1 group](#)

Let's not forget Galaxy Main...

This is your... *usegalaxy.org*

Universe View (from @natefoo)





A national science and engineering cloud

- Support by NSF through XSEDE
- Burst capacity for Galaxy main
- Private Galaxy cloud instances through XSEDE

Looking forward

Analysis at scale

- Continued evolution of dataset collections, expand to other data types, complex structures, make pervasive in Galaxy
- Workflow improvements in scheduling and representation, comparison, sweeps...
 - “Activities”, a new Galaxy entry point for simplifying interfaces to complex analyses

Streamlining tool development & distribution

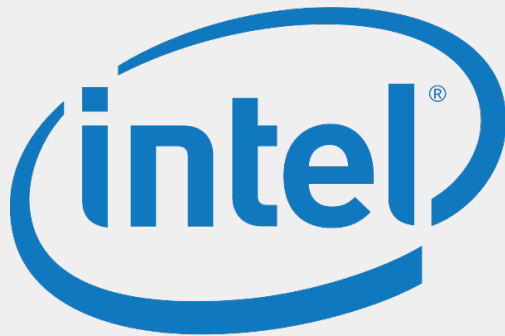
- Continue building tooling for git centric tool development
- Reposition ToolShed for tool discovery
 - Improve dependency management

Maximizing use of compute infrastructure

- Continue improving Pulsar for integrating heterogeneous resources
 - Improvements to cloud bursting, data federation, identity federation across resources
 - Fully unify and automate different types of Galaxy deployment (physical, virtual, docker, ...)

Framework support for increased data interactivity

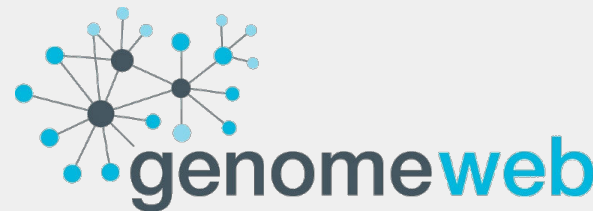
- Deeper integration for interactive environments
 - Expose Galaxy's provenance capabilities to Viz/IE plugins
 - Capture full provenance graph for all data produced from IEs
 - Allow *reusable parameterized tools* to be developed on the fly using IEs



kelway
beyond technology



BIOTeam



Genomic Data Science
Specialization

