

Galaxy Update

GMOD Workshop PAG 2013



Jennifer Hillman-Jackson
Penn State University



Wednesday, January 16
Workshop 3:45-4:30 pm

galaxyproject.org
gmod.org

Agenda

Galaxy project mission

Who's on the team

Overview start of 2013

How to get involved

Agenda

Galaxy project mission

Who's on the team

Overview start of 2013

How to get involved

Galaxy Project Mission

Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research.

Accessible: Users without programming experience can easily specify parameters and run tools and workflows.

Reproducible: Galaxy captures information so that any user can repeat and understand a complete computational analysis.

Transparent: Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

Galaxy Project **Mission**

"Next-generation sequencing data interpretation: enhancing reproducibility and accessibility", by Nekrutenko & Taylor, *Nature Reviews Genetics*, 13, 667-672 (September 2012)

Agenda

Galaxy project mission

Who's on the team

Overview start of 2013

How to get involved



Enis Afgan



Guru Ananda



Dannon Baker



Dan Blankenberg



Dave Bouvier



Dave Clements



Nate Coraor



Greg von Kuster



Carl Eberhard



Jeremy Goecks



Jen Jackson

The Core Galaxy Team



Ross Lazarus



**Anton
Nekrutenko**

**James
Taylor**



<http://galaxyproject.org/wiki/GalaxyTeam>



Nuwan Goonasekera



Yousef Kowsar

Finn Bacall
Brad Chapman
John Chilton
Peter Cock
Kyle Ellrott

*** Plus many others via suggestions/help/code through Trello, Tool Shed, and galaxy-dev@bx.psu.edu mailing list. Galaxy is truly a community effort!**

Extended Team and Other Current Contributors

<http://galaxyproject.org/wiki/GalaxyTeam>

Agenda

Galaxy project mission

Who's on the team

Overview start of 2013

How to get involved

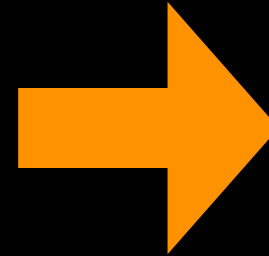
Agenda

Galaxy project mission

Who's on the team

Overview start of 2013

How to get involved



Using Galaxy
Web Objects/UI
Data & Tools
Visualizations
Publications
Tool Shed
Support

Using Galaxy - 4 ways

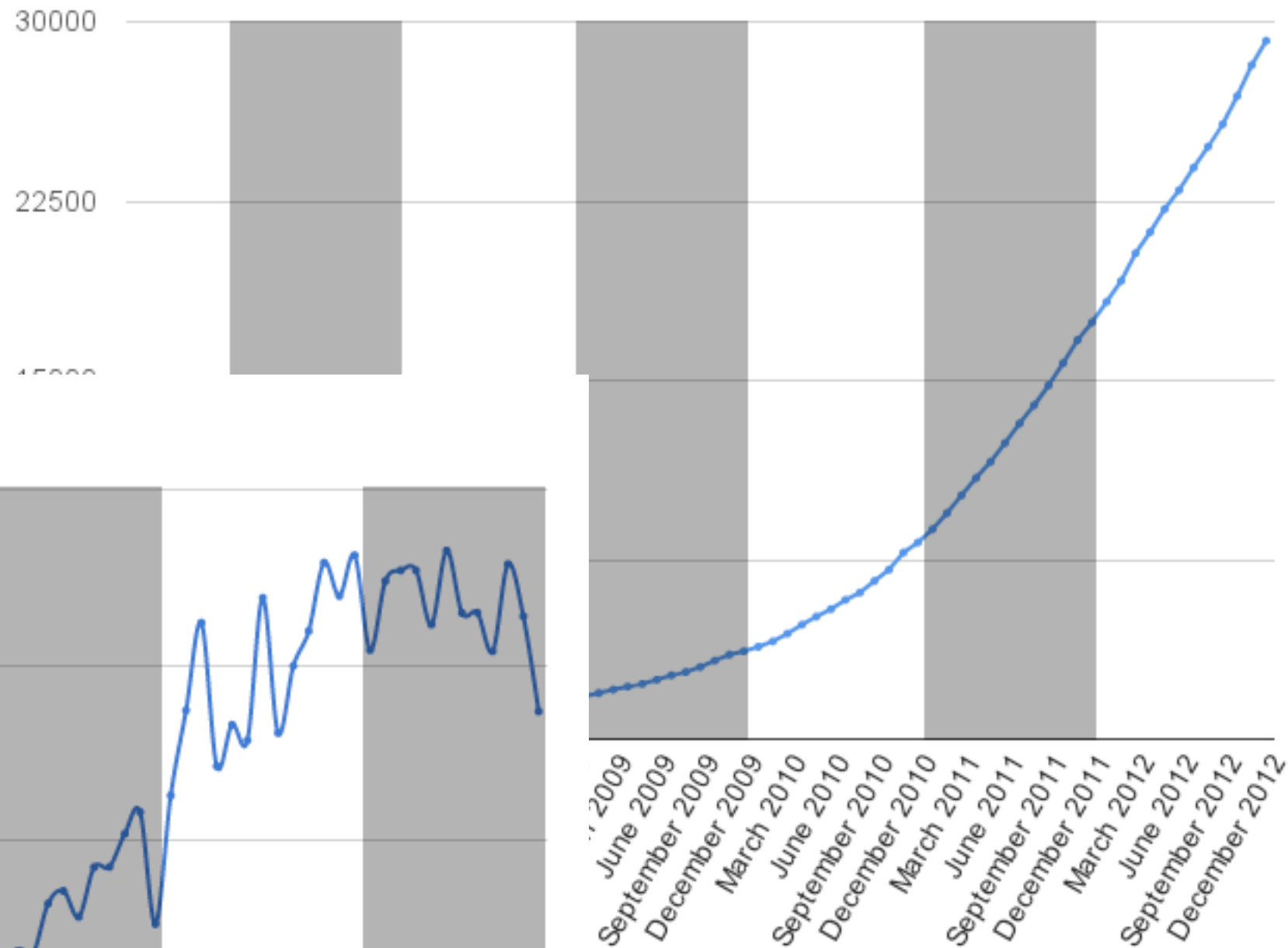
- **Public Main** Galaxy web instance: *usegalaxy.org*
- Local instance: *getgalaxy.org*
- Cloud instance: *usegalaxy.org/cloud*
- **Other Public** Galaxy web instances hosted by various groups:
wiki.galaxyproject.org/PublicGalaxyServers

<http://wiki.galaxyproject.org/Big%20Picture/Choices>

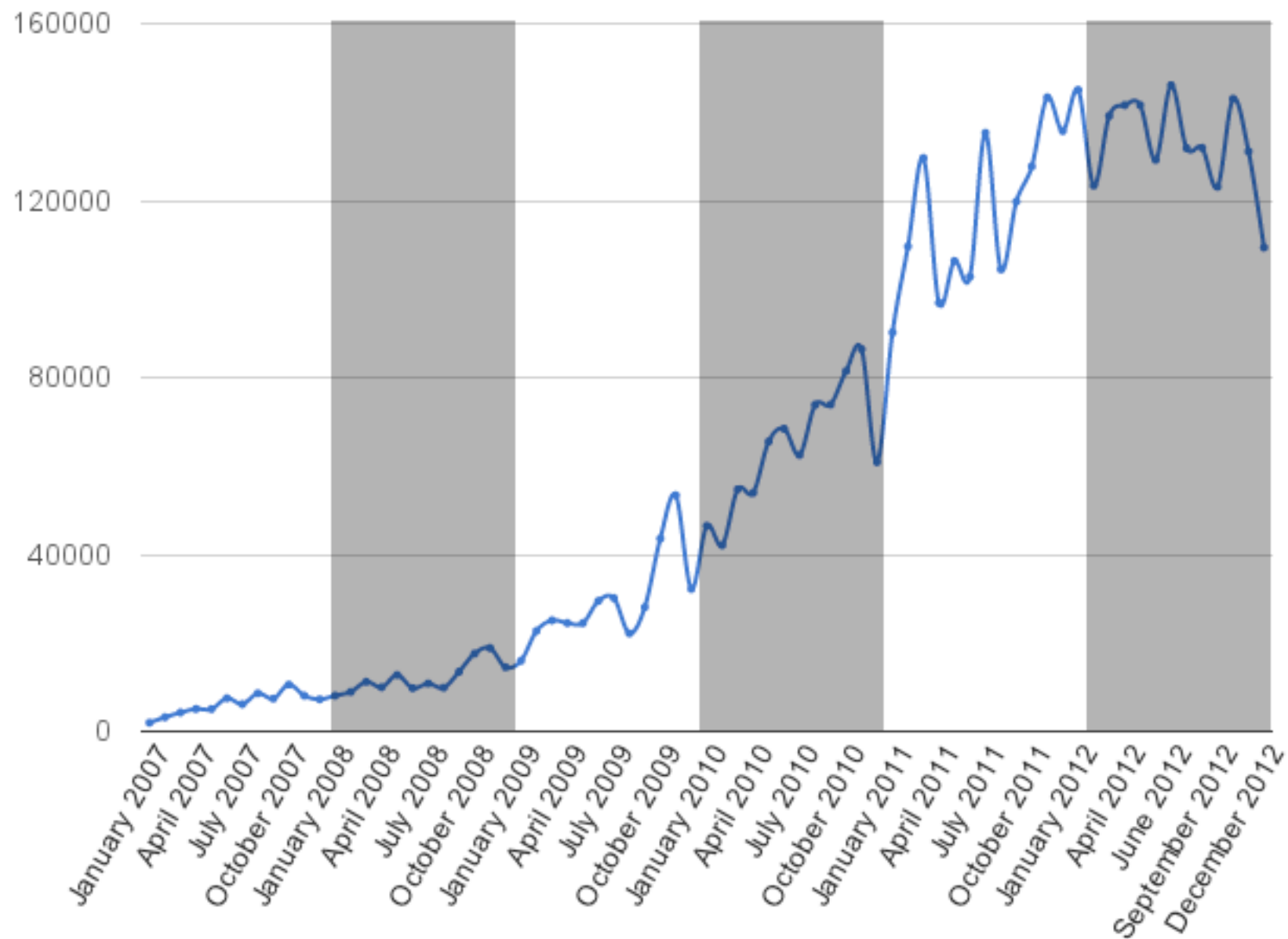
usegalaxy.org

"Main"

Registered Users on Galaxy Main



User Jobs per month on usegalaxy.org



What's new?
More hardware
- soon.

getgalaxy.org
“Local”

wiki.galaxyproject.org/DevNewsBriefs
galaxy-dist.readthedocs.org
bitbucket.org/galaxy/galaxy-dist

Code Downloads -- how often??

Alas, this information does not appear to be available from Bitbucket. Therefore, *we don't know*.

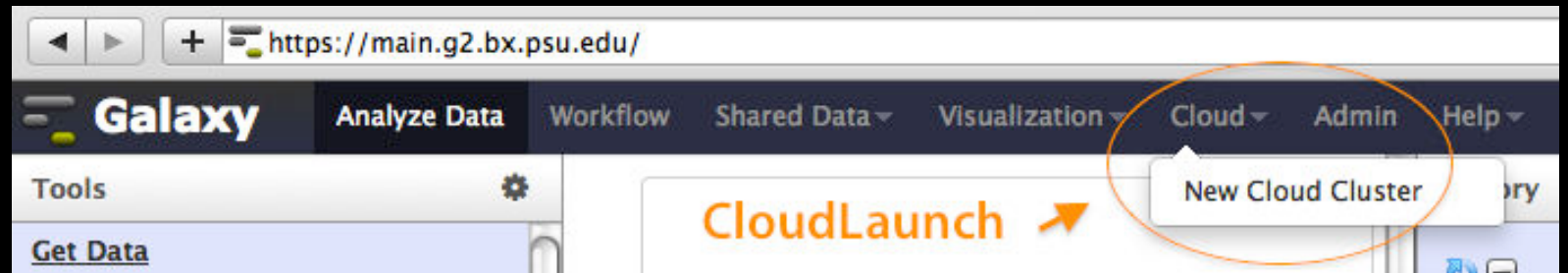


What's new?

- ~ 2 week release cycle
- “readthedocs” documentation

usegalaxy.org/cloud

“CloudMan”

A screenshot of the 'Launch a Galaxy Cloud Instance' page. The page title is 'Launch a Galaxy Cloud Instance'. Below the title, there is a paragraph of text explaining the process. There are two input fields: 'Key ID' and 'Secret Key'. Below each field is a descriptive sentence. The 'Key ID' field is followed by the text: 'This is the text string that uniquely identifies your account, found in the [Security Credentials](#) section of the [AWS Console](#).' The 'Secret Key' field is followed by the text: 'This is your AWS Secret Key, also found in the [Security Credentials](#) section of the [AWS Console](#).'

What's new?

- Educational grants for cloud time from Amazon
- CloudLanuch directly from within “Main”
- Publications integrating Galaxy cloud workflows

Galaxy CloudMan

<http://usegalaxy.org/cloud>

- Start with a **fully configured and populated** (tools and data) Galaxy instance.
- Allows you to scale up and down your compute assets as needed.
- Someone else manages the data center.



<http://aws.amazon.com/education>

wiki.galaxyproject.org/PublicGalaxyServers

“Known Publicly Accessible Servers”

This is not an absolute count, but it is a rough measure of the trend.

Date	# Servers
2011/07	15
2012/01	21
2012/07	20
2013/01	25

What's new?

- **“GalaxyAdmins” community group founded**

Public Galaxy Servers

Interested in:

ChIP-chip and ChIP-seq?

✓ Cistrome

Statistical Analysis?

✓ Genomic Hyperbrowser

Sequence and tiling arrays?

✓ Oqtans

Text Mining?

✓ DBCLS Galaxy

Reasoning with ontologies?

✓ GO Galaxy

Internally symmetric protein structures?

✓ SymD

<http://galaxyproject.org/wiki/PublicGalaxyServers>

Common to all Development contributors and general users, the new Trello Issue Board replaced bitbucket in 2012:

<http://wiki.galaxyproject.org/Issues>

The screenshot shows a Trello web interface in a browser. The address bar displays the URL <https://trello.com/board/galaxy-development-inbox/50686d0302dfa79d13d90c45>. The Trello logo and navigation links (Help, Notifications, Boards) are at the top. The board is titled "Galaxy: Development Inbox" and is associated with the "Galaxy Project" workspace. The board is organized into four columns: "Inbox", "Developer ideas", "Bug Reports", and "Issues from Bitbucket".

- Inbox:** Contains one card titled "Add a card...".
- Developer ideas:** Contains three cards: "Google Drive / Dropbox / Box / ... integration", "Standalone web application(s) for visualizations", and "Assistive UI" (with a "0/2" label). A fourth card "Use unicode type for all text columns in database" is partially visible.
- Bug Reports:** Contains three cards: "823: picard index indicates failure, but it is successful", "822: cannot run updatencbi.sh", and "Change in # of parameters for a tool in a workflow causes an unhandled exception".
- Issues from Bitbucket:** Contains four cards: "5: Option to disable automatic history creation", "6: Option to require that histories have names", "8: More flexible output handlers", and "10: Allow overriding parameters when running a workflow". A fifth card "20: Suggestion: new tag" is partially visible.

On the right side of the board, there is a "Members" section showing a grid of member avatars, including "CE" and "DB". Below this is an "Add Members..." button. The "Board" section includes "Options", "Add List", and "Search and Filter Cards". At the bottom right, an "Activity" section is highlighted with an orange border, showing a list of recent actions:

- Activity View all...**
- Dannon Baker enabled self join on this board. yesterday at 8:35am
- Dannon Baker moved Change in # of

Galaxy UI Terminology

Dataset:

Any input, output or intermediate set of data + metadata.
A record of a specific data or analysis step.

History:

A series of inputs, analysis steps, intermediate datasets, and outputs. A record of a group of data and analysis steps.

Tool:

An operation within Galaxy that acts upon dataset(s) as an analysis step. May be developed by Galaxy team or a 3rd party program that has been “wrapped” for Galaxy.

Workflow:

A series of analysis steps executed in a sequential stream

What's **new** in the UI?

Dataset:

More “i” info page details. Metadata includes underlying tool version for most tools in the “info” attribute. Adding to all is a goal.

History:

Re-run warns if original input datasets or tool versions are not available. Finds inputs even if “hidden”.

Tool:

Wrapper versions displayed in UI for all.

Workflow:

Enhancements to control via API. Warning similar to history re-run actions to support reproducibility.

General:

Multiple display improvements and tunings.

Data and Tools

Reference Genomes:

Dozens of full genomes added and over a hundred genomes had some content (index, liftOver) added in 2012. **New data early 2013** to include **Bowtie2** indexes both on **Main** and **rsync** download area.

- The **rsync area was new** in 2012, too:

<http://wiki.galaxyproject.org/Admin/Data%20Integration>

Key Tools Included:

GATK (beta); Updates to the RNA-seq tool set **Bowtie2/****Tophat2**, **Cufflinks/merge/diff**; **FreeBayes**; **Trinity** (Tool Shed); **Wormbase 2**; **IGB**; **GenomeSpace**; **Megablast** to use **BLAST+**; **MPileup**, and the **Tool Factory** (Tool Shed):

"Creating re-usable tools from scripts: The Galaxy Tool Factory," Ross Lazarus, Antony Kaspi, Mark Ziemann, The Galaxy Team, Bioinformatics (28 September 2012)

Visualizations

Trackster had significant number of new refinements in 2012 leading to a publication.



Jeremy Goecks, Nate Coraor, The Galaxy Team, Anton Nekrutenko & James Taylor, "NGS analyses by visualization with Trackster." *Nature Biotechnology* 30, 1036–1039 (2012)

And Circster (not shown) and Scatterplot were introduced:

[illegible]

Publications - CiteULike

A **Galaxy CiteULike** group was started in late 2011. It lists all the pubs that are about, reference, or mention Galaxy that we know about. We started keeping track of this partway through 2011, so it is an undercount for that year (and previous years are almost entirely absent). For years after 2011, it is likely to be more accurate, but still approximate, and to still be an undercount.

As of January 2013, there are this many papers in the Galaxy CiteULike Group:

Publication Year	# Papers in CiteULike Group
2005	2
2006	3
2007	8
2008	22
2009	42
2010	76
2011	183
2012	398
2013	15
Total	759

Date	Papers in CiteULike Group
2012/01	174
2012/07	361
2013/01	759

Publications - Example 1

The screenshot shows the top of a web page for 'GENOME RESEARCH' by CSH PRESS. The header includes navigation links (HOME, ABOUT, ARCHIVE, SUBMIT, SUBSCRIBE, ADVERTISE, AUTHOR INFO, CONTACT, HELP) and a search bar. The article title is 'Windshield splatter analysis with the Galaxy metagenomic pipeline' by Sergei Kosakovsky Pond and Samir Wadhawan. A highlighted 'Footnotes' section contains the following text: [Supplemental material is available online at <http://www.genome.org>. All data and tools described in this manuscript can be downloaded or used directly at <http://galaxyproject.org>. Exact analyses and workflows used in this paper are available at <http://usegalaxy.org/u/aun1/p/windshield-splatter>.]

CSH PRESS GENOME RESEARCH

EXPRESS ION ANALYSIS illumina Apply today for the Cancer GWAS Grant.

HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | CONTACT | HELP

Institution: PENN STATE UNIV Sign In via User Name/Password

Search for Keyword: Go
Advanced Search

Windshield splatter analysis with the Galaxy metagenomic pipeline

Sergei Kosakovsky Pond^{1,2,6,9}, Samir Wadhawan^{3,6,7},
Fran James

Footnotes

[Supplemental material is available online at <http://www.genome.org>. All data and tools described in this manuscript can be downloaded or used directly at <http://galaxyproject.org>. Exact analyses and workflows used in this paper are available at <http://usegalaxy.org/u/aun1/p/windshield-splatter>.]

OPEN ACCESS ARTICLE

This Article

Published in Advance October 9, 2009, doi: 10.1101/gr.094508.109
Copyright © 2009 by Cold

Current Issue
October 2010, 20 (10)

GENOME RESEARCH

Histories, workflows, visualizations and *pages* can be shared with others or published to the world.

<http://usegalaxy.org/u/aun1/p/windshield-splatter>

Publications - Example 2

Galaxy Analyze Data Workflow **Shared Data** Visualization Cloud Admin Help U

Published Pages | galaxyproject | Using Galaxy 2012

June 2012

Using Galaxy to Perform Large-Scale Interactive Data Analysis: A live supplement

Jennifer Hillman-Jackson,¹ Dave Clements,² Daniel Blankenberg,¹ James Taylor,² Anton Nekrutenko,¹ and the Galaxy Team^{1,2}

¹Penn State University, University Park, Pennsylvania
²Emory University, Atlanta, Georgia

Correspondence should be addressed to [Jennifer Hillman-Jackson](#)

How to use this document

This document is an interactive supplement to "Using Galaxy: Finding Human Coding Exons with Highest SNP Density". Every protocol, dataset, and supplementary items at Galaxy can be examined, copied, and migrated to a local or cloud Galaxy instance ([getgalaxy.org](#)), IGV, Ensembl Browser or other tool of interest. All external datasets are public; please review each dataset's license. Citations should reference this publication, the core Galaxy tools used as appropriate.

For each Protocol, the following is provided:

- Input datasets
- Complete history

History

A complete history for **Basic Protocol 1**, showing all input, intermediate, and output datasets, and a description of each step in the analysis.

[+ Galaxy History | CPB2012 - BasicProtocol1 - Finding Human Coding Exons with the Highest SNP Density](#)

Screencast Video Tutorial

["Using Galaxy: Finding Human Coding Exons with Highest SNP Density"](#)

Protocol 1 step-by-step video tutorial that includes a supplemental Trackster walk-through for visualizing input and result datasets.

Using Galaxy
protocol 1

Finding Human
Coding Exons with
Highest SNP Density

Shared Data:
Published Pages

Get all of the data and follow a tutorial start to finish using the supplemental methods, workflows, and screencasts.

<http://main.g2.bx.psu.edu/u/galaxyproject/p/using-galaxy-2012>

Galaxy Tool Shed

- **Allow users to share “containers”** of tools, datatypes, workflows, sample data, READMEs, and automated installation scripts for tool dependencies.
- **Integration with Galaxy instances** to automate tool installation and updates.
- Is currently **undergoing an audit** to identify “valid tools” with community collaborators.
- In process of supporting **improved dependency documentation** and upgrading installation processes.

Date	# Repositories	# Tools
2011/10	100	
2012/04	~160	1244
2012/07	~230	1967
2013/01	464	2414

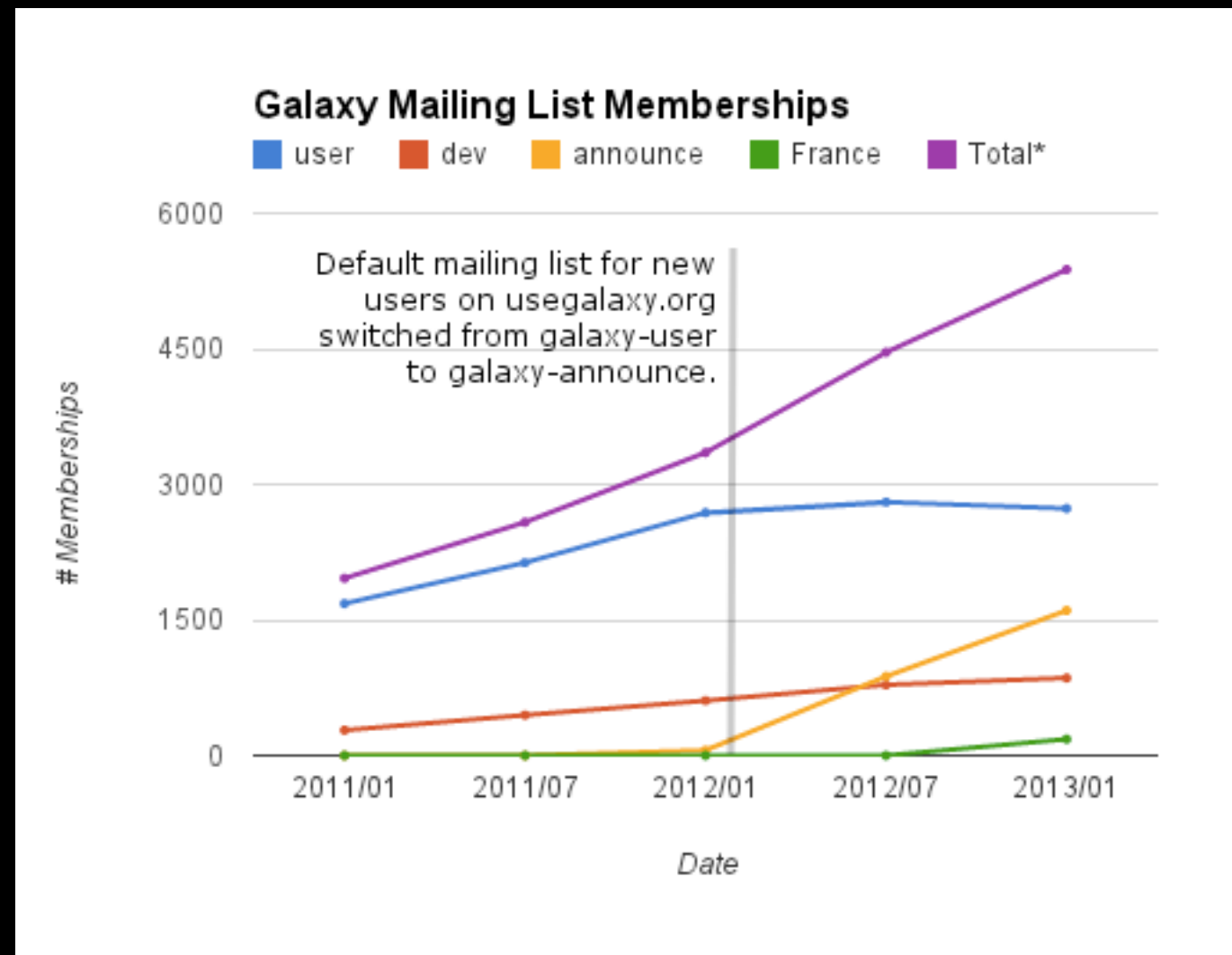
toolshed.g2.bx.psu.edu

Galaxy Tool Shed

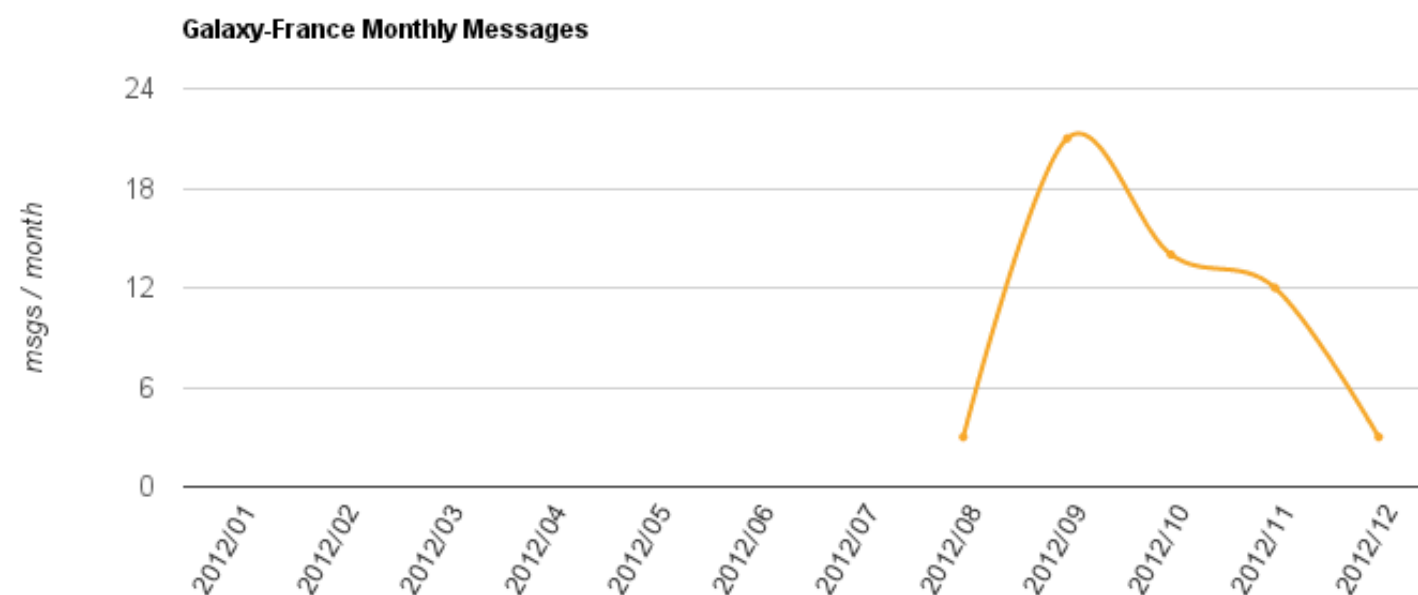
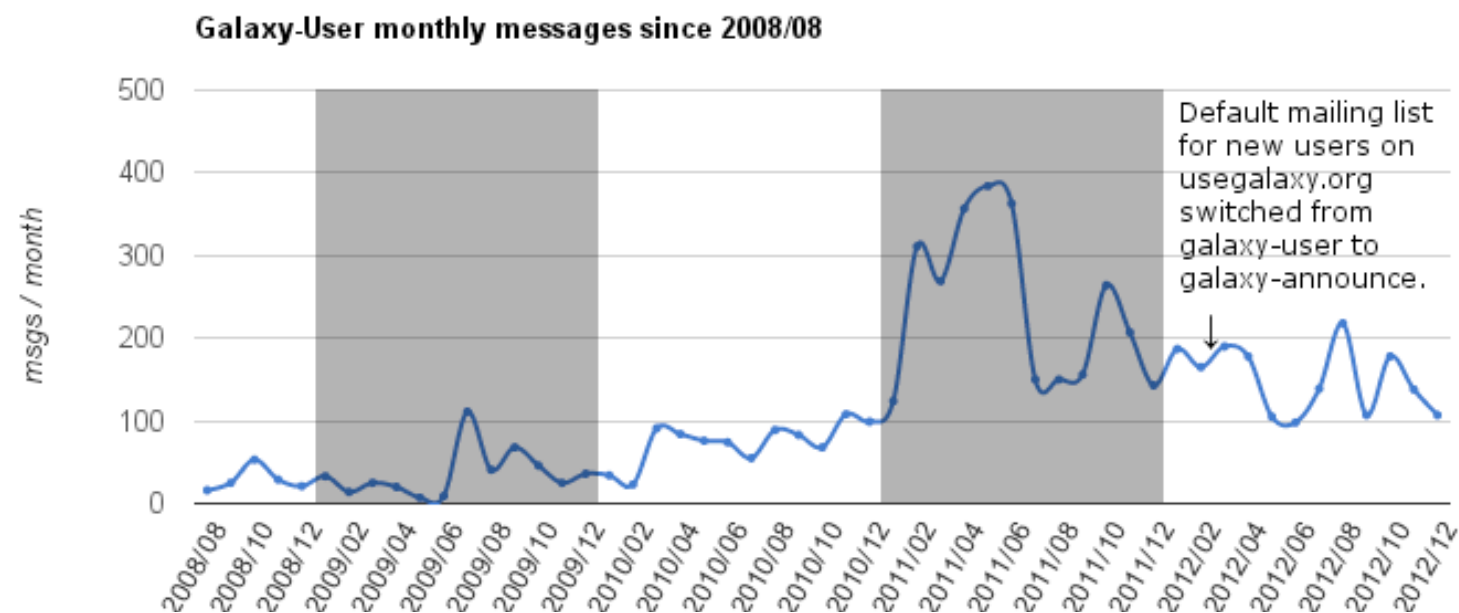
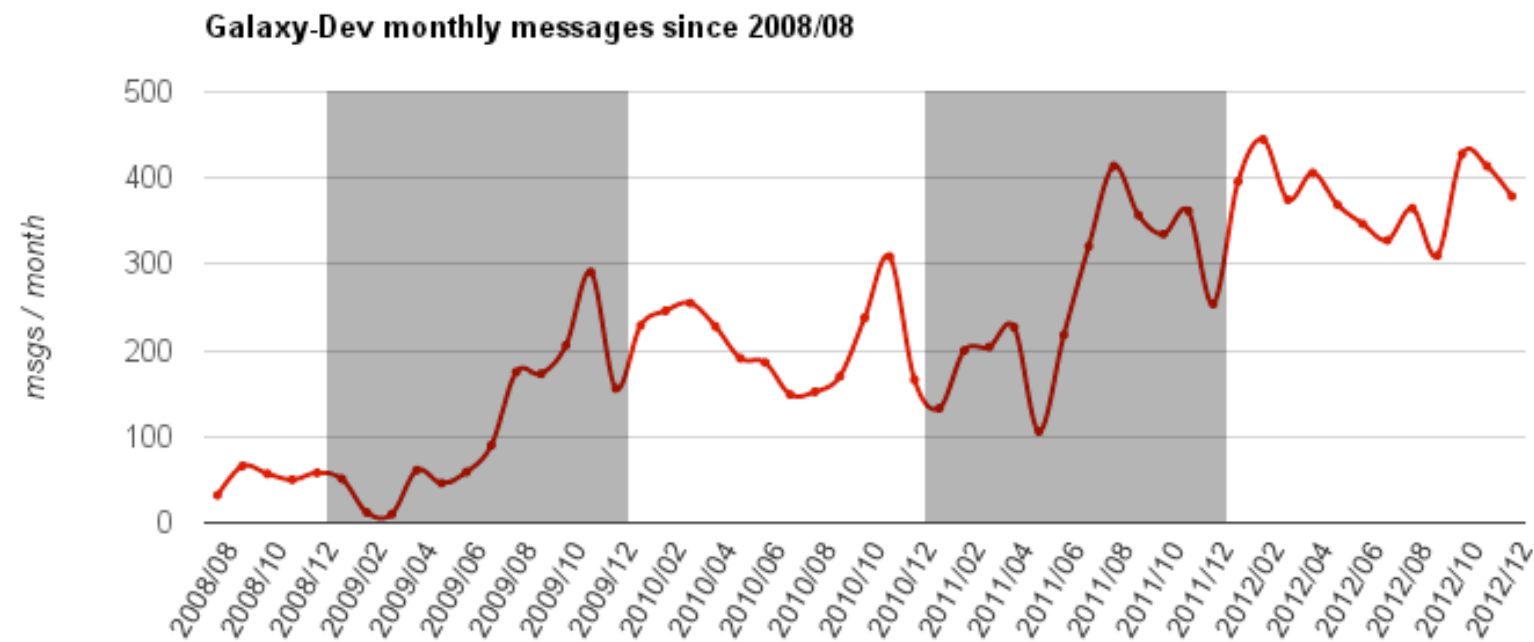
- There is a “**Main**” tool shed hosted by the core Galaxy team, but satellite tools sheds are encouraged. We’d like to learn about them and list on our wiki, as public Galaxies are.
- This “**Main**” tool shed currently tracks the Galaxy distribution as many enhancements are linked/dependent.
- The complete list of updates, usage examples, features, etc. (and there have been many this year!), are in the News Brief Archives at: wiki.galaxyproject.org/DevNewsBriefs
- The tool shed documentation covers a LOT of material relevant to the Galaxy framework as a whole:
wiki.galaxyproject.org/Tool%20Shed

Support

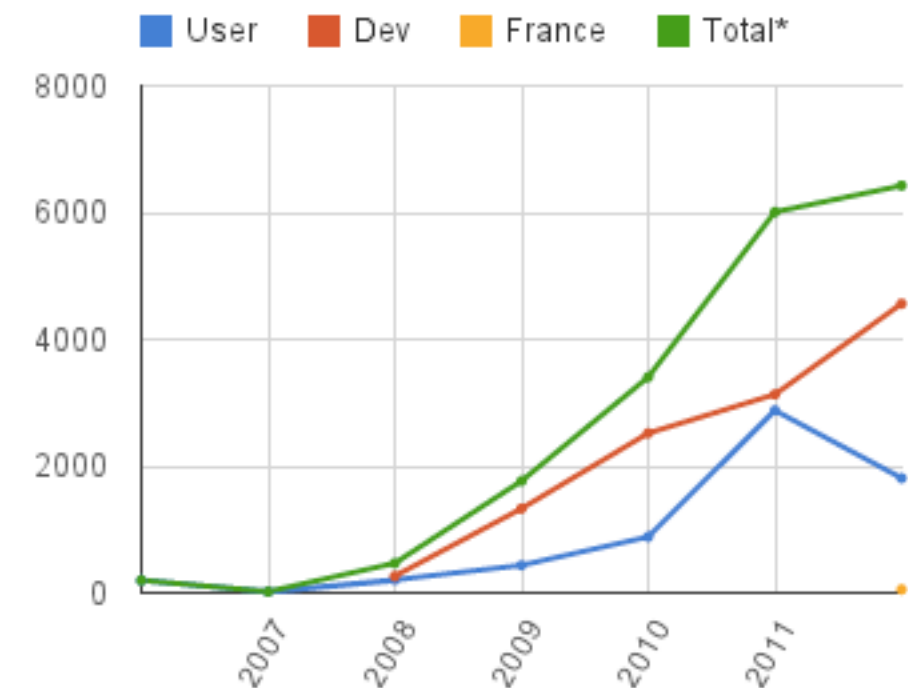
Mailing lists now include: galaxy-user, galaxy-dev, galaxy-announce (**new**), galaxy-france (**new**).



Support



Yearly Message Counts



“Public” mailing list traffic (does not include direct help through “bug reports”, which are most often usage help questions.)

Agenda

Galaxy project mission

Who's on the team

Overview start of 2013

How to get involved

Join the Galaxy Community

Tool Shed

Mailing Lists (Sci User, Development, Private Data help)

Screencasts (Community and Galaxy Team)

Events Calendar, News Feed, Twitter, Monthly Updates

Distributions & Release Notes/Feature Descriptions

Community Wiki

Local/Public Installs, GalaxyAdmins Monthly Mtg

CiteULike group, Mendeley mirror

Annual **Galaxy Community** Meeting - **GCC2013 is next!**

<http://galaxyproject.org/wiki>

Event/Year	Location	# Registered
GDC2010	Cold Spring Harbor, New York, United States	69
GCC2011	Lunteren, The Netherlands	148
GCC2012	Chicago, Illinois, United States	203
GCC2013	Oslo, Norway	

Galaxy

Community Conference

30 June
- 2 July

2013



UiO : University of Oslo

<http://galaxyproject.org/GCC2013>

**If you walk
away with one thing,
the rest can be found later,
including these slides ...**

<http://galaxypproject.org>

*Thanks for using Galaxy !!
(done)*

Acknowledgments

Scott Cain & GMOD Community

PAG 2013 Organizers

Dave Clements, Anton Nekrutenko,
James Taylor, & Galaxy Team

Galaxy Community & You!!

Hope to see
you in Oslo!

Galaxy Community Conference

30 June
- 2 July

2013



UiO : University of Oslo

<http://galaxypproject.org/GCC2013>