

# Galaxy @ UCSC BME 110

Oct 4, 2013

---



Jennifer Hillman-Jackson  
Penn State University

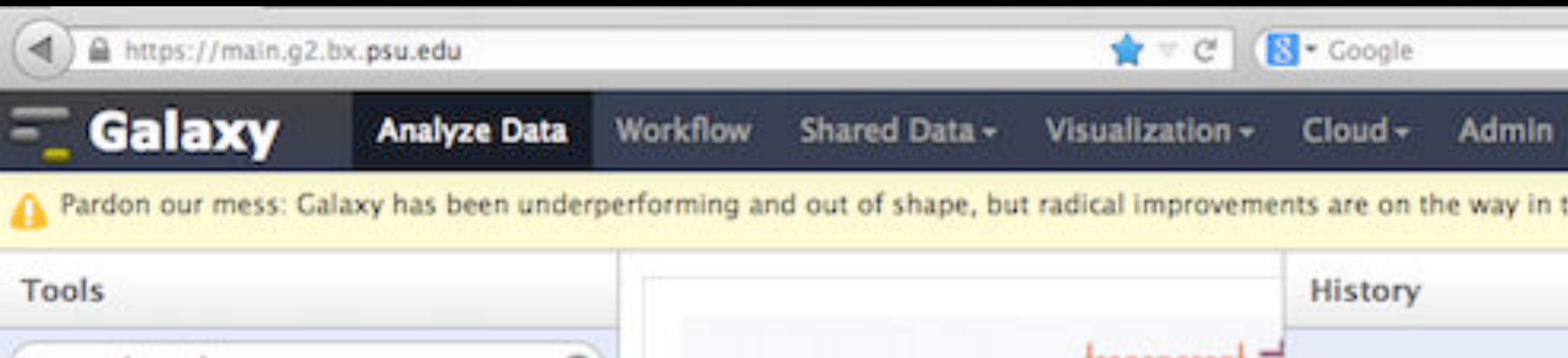
[galaxyproject.org](http://galaxyproject.org)  
[genome.ucsc.edu](http://genome.ucsc.edu)

Before we get started ...

## ***What's up with MAIN????***

major upgrades to infrastructure ... vast improvements

*.....coming soon*



*deep breath.... ohmm*



# 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)

# Galaxy as a Genomics WorkBench

## 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.

# More Galaxy Terminology

## Share:

Make something available to someone else

## Publish:

Make something available to everyone

## Galaxy Page:

Analysis documentation within Galaxy; easy to embed and link to any Galaxy object (histories, datasets, workflows, visualization) or external resource (video, graphics, publications).

The tutorial we will do today is in a **Galaxy Page**.

# What is Galaxy?

- **A free (for everyone) web service** integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage
- **Open source software** that makes integrating your own tools and data and customizing for your own site simple
- These options result in several **ways to use Galaxy**

*<http://galaxyproject.org>*

# Using Galaxy - 5 ways

- **Public Main** Galaxy web instance: [\*usegalaxy.org\*](http://usegalaxy.org)
- **Local** instance: [\*getgalaxy.org\*](http://getgalaxy.org)
- **Cloud** instance: [\*usegalaxy.org/cloud\*](http://usegalaxy.org/cloud)
- **SlipStream Appliance, Galaxy Edition:**  
[\*www.bioteam.net/slipstream/galaxy-edition\*](http://www.bioteam.net/slipstream/galaxy-edition)
- **Other Public** Galaxy web instances hosted by various groups:  
[\*wiki.galaxyproject.org/PublicGalaxyServers\*](http://wiki.galaxyproject.org/PublicGalaxyServers)  
[\*http://wiki.galaxyproject.org/Big%20Picture/Choices\*](http://wiki.galaxyproject.org/Big%20Picture/Choices)

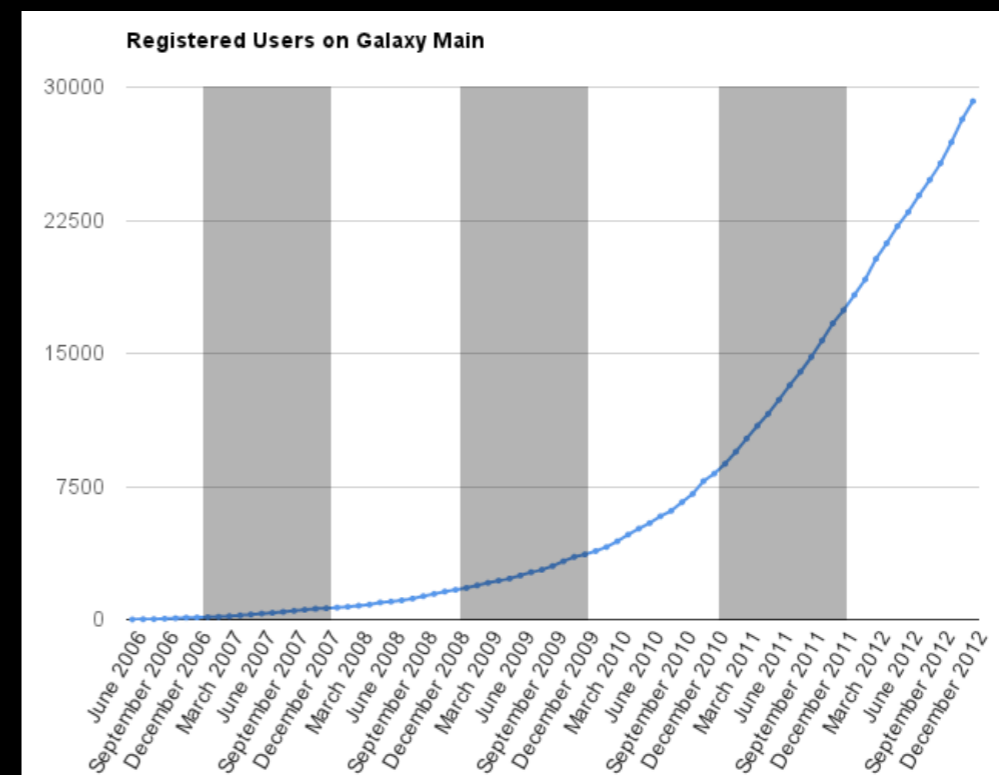
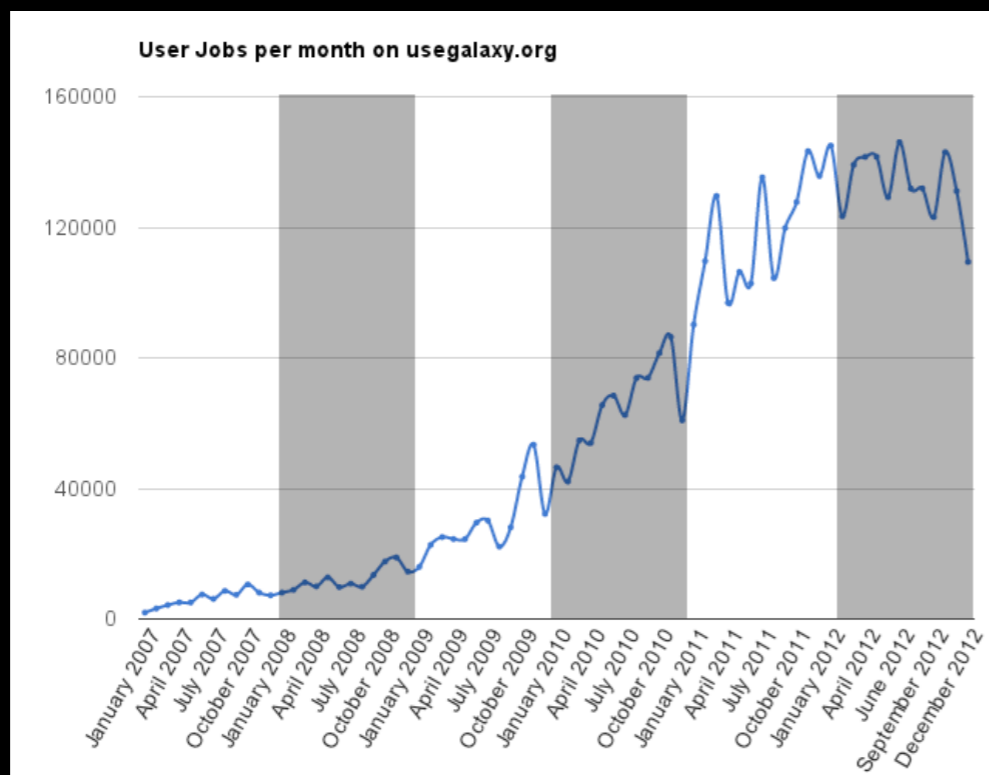
# Galaxy is available ...

As a free (for everyone) web service

## Galaxy “Main”

<http://usegalaxy.org>

However, *a centralized solution cannot scale to meet the analysis needs of the entire world.*



# Galaxy is available ...

- As a free (for everyone) web service

<http://usegalaxy.org>

- As open source software

<http://getgalaxy.org>

Galaxy “Local”

*[wiki.galaxyproject.org/DevNewsBriefs](http://wiki.galaxyproject.org/DevNewsBriefs)*

*[galaxy-dist.readthedocs.org](http://galaxy-dist.readthedocs.org)*

*[bitbucket.org/galaxy/galaxy-dist](http://bitbucket.org/galaxy/galaxy-dist)*

# As Open Source Software: Local Galaxy Instances

- Galaxy is designed for local installation and customization
- Easily integrate new tools
- Easy to deploy and manage on nearly any (unix) system
- Run jobs on existing compute clusters
- Requires a computational resource on which to be deployed

<http://getgalaxy.org>

# Encourage **Local** Galaxy Instances

- Encourage and support Local Galaxy Instances
- Support **increasingly decentralized model** and improve access to existing resources
- Focus on building **infrastructure to enable the community to integrate and share** tools, workflows, and best practices

**Galaxy Tool Shed**

**<http://toolshed.g2.bx.psu.edu>**

The screenshot shows the Galaxy Tool Shed interface for the 'clustalomega' repository. The left sidebar contains links for 'Repositories', 'Browse by category', 'Browse all repositories', and 'Login to create a repository'. The main content area displays the repository revision '2:bb1847435ec1' and provides a link to clone the repository. Below this, there is a 'Synopsis' and a 'Detailed description' for the 'clustalomega' tool. At the bottom, a table lists the tools available in the repository, including 'Clustal Omega'.

name	description	version	requirements
Clustal Omega	multiple sequence alignment program for proteins	1.0.2	none

The screenshot shows the Galaxy Tool Shed interface with a list of repositories. The left sidebar is the same as in the previous screenshot. The main content area displays a table of repositories with columns for Name, Synopsis, Revision, Category, and Owner. The table lists several repositories, including 'abyss\_toolsuite', 'agile\_wrapper', 'asdf', 'assemblystats', and 'bam\_to\_bigwig'.

Name	Synopsis	Revision	Category	Owner
abyss_toolsuite	This suite contains Abyss and Abyss-PE config files and wrappers for Galaxy	0:92636934a189	Assembly	edward-kirton
agile_wrapper	Quickly match reads to a reference genome or sequence file	0:d6a426afaa46	Next Gen Mappers, Sequence Analysis	simonl
asdf	asdf	-1:000000000000	Statistics, Text Manipulation	vivek
assemblystats	Summarise an assembly (e.g. N50 metrics)	0:6544228ea290	Next Gen Mappers, Sequence Analysis	konradpaszkiewicz
bam_to_bigwig	Generate BigWig coverage files from BAM files. Allows gapped reads to be split (useful for RNA-Seq).	5:5b40b93ebae3	Convert Formats, SAM, Visualization	lparsons

# As Open Source Software: Local Galaxy Instances

- Galaxy is designed for local installation and customization
- Easily integrate new tools
- Easy to deploy and manage on nearly any (unix) system
- Run jobs on existing compute clusters
- Requires a **computational resource** on which to be deployed

**<http://getgalaxy.org>**

# Galaxy is available ...

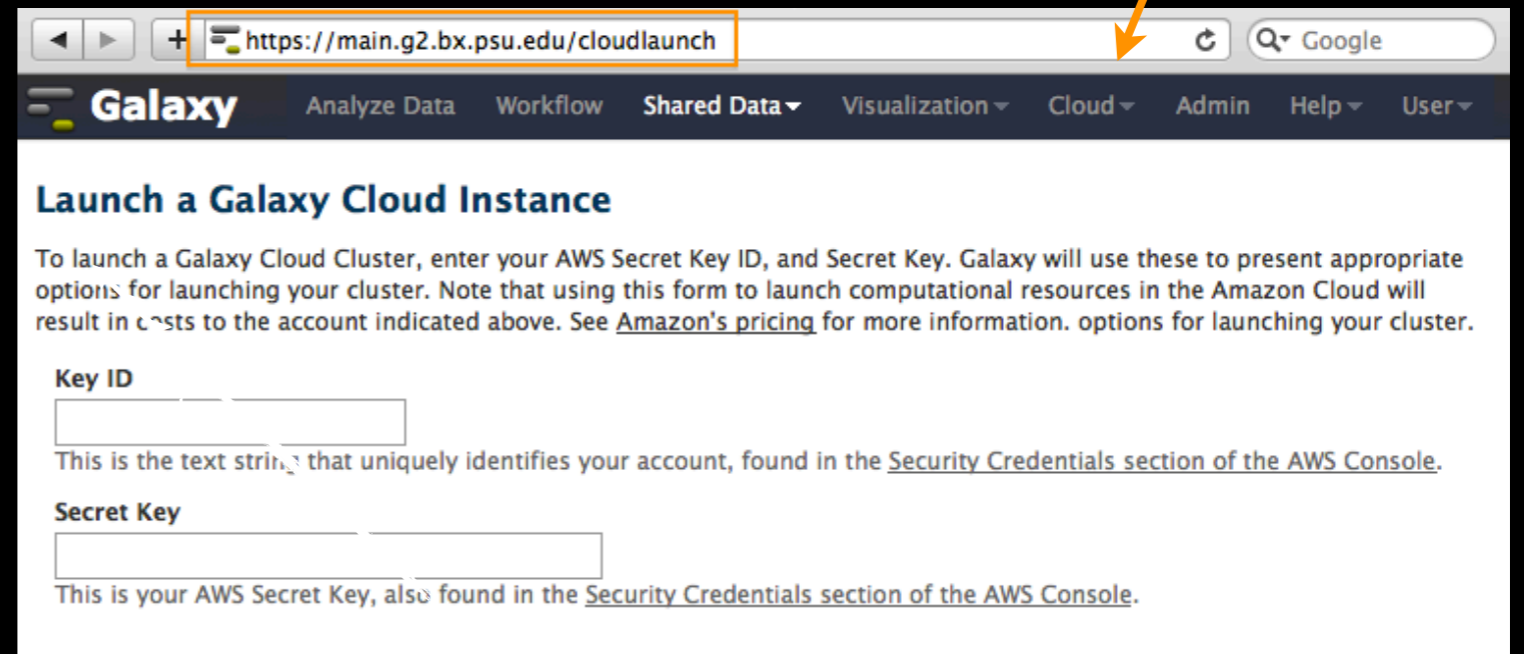
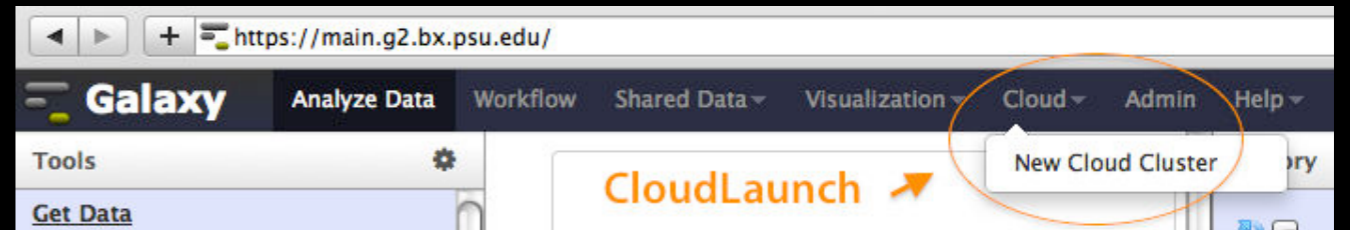
1. Start with a **fully configured and populated** (tools and data) Galaxy instance.
2. Allows you to scale up and down your compute assets as needed.
3. Someone else manages the data center.

- **On the Cloud**

<http://usegalaxy.org/cloud>

Galaxy “CloudMan”

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



# Galaxy “Public” Instances

<http://bit.ly/gxyServers>

Interested in:

ChIP-chip and ChIP-seq?

✓ Cistrome

Statistical Analysis?

✓ Genomic Hyperbrowser

Protein synthesis?

✓ GWIPS-viz

*de novo* assembly?

✓ CBIIT Galaxy

Reasoning with ontologies?

✓ OPPL Galaxy

Repeats!

✓ RepeatExplorer

Everything?

✓ Andromeda

Plus many more

# Galaxy Resources and Community

Mailing Lists (very active)

Unified Search

Issues Board

Events Calendar, News Feed

Community Wiki

GalaxyAdmins

Screencasts

Tool Shed

Public Installs

CiteULike group, Mendeley mirror

Annual Community Meeting

<http://wiki.galaxyproject.org>

# Galaxy Resources and Community: Mailing Lists

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

## Galaxy-Announce

Project announcements, low volume, moderated

Low volume ( 42 posts in 2012, 2100+ members)

## Galaxy-User

Questions about using Galaxy and usegalaxy.org


High volume (2900 posts in 2012, 2700+ members)

## Galaxy-Dev


Questions about developing for and deploying Galaxy

High volume (4500 posts in 2012, 900+ members)

# Unified Search: <http://galaxyproject.org/search>

 **Galaxy Web Search**

Google™ Custom Search

Search 

Search the entire set of Galaxy web sites and mailing lists using Google.

[Run this search at Google.com \(useful for bookmarking\)](#)

Want a [different search](#)?

[Project home](#)

**Find**

Everything on ...

Tools for ...

Email about ...


Source code for ...

Published Histories, Pages, Workflows, about ...

Documentation on ...

Papers using Galaxy for ...

Related feature requests

 **Galaxy Web Search**

chip-seq

All Tools Email Source code Shared Documentation Abstracts Requests

About 444 results (0.06 seconds)

[Galaxy | Accessible Page | ChIP-seq exercise](#)

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

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

The screenshot displays the Trello web interface for the 'Galaxy: Development Inbox' board. The board is structured into four columns: 'Inbox', 'Developer ideas', 'Bug Reports', and 'Issues from Bitbucket'. The 'Inbox' column is currently empty. The 'Developer ideas' column contains three cards: 'Google Drive / Dropbox / Box / ... integration', 'Standalone web application(s) for visualizations', and 'Assistive UI'. The 'Bug Reports' column 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'. The 'Issues from Bitbucket' column 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'. The right sidebar shows 'Members' (a grid of 12 user avatars), 'Board' (options like 'Options', 'Add List', 'Search and Filter Cards'), and 'Activity' (a list of recent actions, including 'Dannon Baker enabled self join on this board' and 'Dannon Baker moved Change in # of'). The URL bar at the top shows the Trello board link: https://trello.com/board/galaxy-development-inbox/50686d0302dfa79d13d90c45.

# http://wiki.galaxyproject.org

Galaxy Wiki

FrontPage

LockedHistoryActions

Galaxy

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.

This is the Galaxy Community Wiki. It describes all things Galaxy.

Use Galaxy

Galaxy's [public service web site](#) makes analysis tools, genomic data, tutorial demonstrations, persistent workspaces, and publication services available to any scientist. Extensive [user documentation](#) (applicable to any [public](#) or local Galaxy instance) is available on [this wiki](#) and elsewhere.

usegalaxy.org

Deploy Galaxy

Galaxy is open source for all organizations. Local Galaxy servers can be set up by [downloading and customizing](#) the Galaxy application.

- [Admin](#)
- [Cloud](#)

getgalaxy.org

Community & Project

Galaxy has a large and active user community and many ways to [Get Involved](#).

- [Community](#)
- [News](#)
- [Events](#)
- [Support](#)
- [Galaxy Project](#)

Contribute

- **Users:** [Share](#) your histories, workflows, visualizations, data libraries, and [Galaxy Pages](#), enabling others to use and learn from them.
- **Deployers and Developers:** Contribute tool definitions to the Galaxy [Tool Shed](#) (making it easy for others to use those tools on their installations), and code to the core release.
- **Everyone:** [Get Involved!](#)

Galaxy Community Conference

OSLO

30 June - 2 July 2013


University of Oslo

Poster abstracts due 3 May

Use Galaxy

[Use Main \(about\)](#)  
[Use Others!](#) • [Learn](#)  
[Share](#) • [Search](#)

Communication

[Support](#) • [News](#)   
[Events](#) • [Twitter](#)  
[Mailing Lists \(search\)](#)

Deploy Galaxy

[Get Galaxy](#) • [Cloud](#)  
[Admin](#) • [Tool Config](#)  
[Tool Shed](#) • [Search](#)

Contribute

[Tool Shed](#) • [Share](#)  
[Issues & Requests](#)  
[Support](#)

Galaxy Project

[Home](#) • [About](#)  
[Community](#)  
[Big Picture](#)

Wiki

[Help](#) • [All Pages](#)

# Events

# News

Galaxy Wiki

Login | Search:



Titles

Events

Locked History






## Galaxy Event Horizon

Events with Galaxy-related content are listed here.

 Also see the [Galaxy Events Google Calendar](#) for a listing of events and deadlines that are relevant to the Galaxy Community. This is also available as an [RSS feed](#) .

If you know of any event that should be added to this page and/or to the Galaxy Event Calendar, please add it here or send it to [outreach@galaxyproject.org](mailto:outreach@galaxyproject.org).

### Upcoming Events



Date	Topic/Event	Venue/Location	Contact
April 29 - May 2	<a href="#">Introduction to Galaxy Workshops</a> 2013 Galaxy Workshop Tour	Washington University in St. Louis	Dave Clements
		Saint Louis University	
		University of Missouri Columbia	
May 14-16	Tutorial: <i>Exploring and Enabling Biomedical Data Analysis with Galaxy</i>	Great Lakes Bioinformatics Conference (GLBIO) 2013, Pittsburgh, Pennsylvania, United States	Anton Nekrutenko
May 15	<a href="#">GalaxyAdmins May 2013 Meetup</a>	GalaxyAdmins web meetup	Srinivas Maddhi, Dave Clements
May 16-17	<a href="#">Galaxy Workflows for Bioinformatics Analysis, and Workshop 1A - Galaxy Workflows for Bioinformatics Analysis</a>	Workshop in Next-Generation Sequence Analysis and Metabolomics (WINGS), UNC-Charlotte, North Carolina, United States	James Taylor
May 21 May 29	<a href="#">Initiation à l'utilisation de Galaxy</a> Les deux ateliers sont maintenant complets	Cycle "Bioinformatique par la pratique" 2013, INRA Jouy-en-Josas, France	Sandra Dérozier, Valentin Loux, Véronique Martin <veronique.martin AT jouy DOT inra DOT fr>
May 22 May 30	Analyse de données issues de séquenceurs nouvelle génération sous Galaxy Les deux ateliers sont maintenant complets		Jean-François Gibrat, Valentin Loux, Véronique Martin <veronique.martin AT jouy DOT inra DOT fr>
May 24 June 19	<a href="#">Introduction to Galaxy</a>		Nikhil Joshi <najoshi AT ucdavis DOT edu>
	<a href="#">A Genomics Virtual Lab for Cancer Research</a>	UC Davis Bioinformatics Core Davis, California, United States	Dominique Gorse
	<a href="#">Next-Gen Sequencing</a>		

Galaxy Wiki

Login | Search:

Titles

Text

News

Locked History Actions

## News

Announcements of interest to the Galaxy Community. These can include items from the Galaxy Team or the Galaxy community and can address anything that is of wide interest to the community.

The Galaxy News is also available as an [RSS feed](#) .

See [Add a News Item](#) below for how to get an item on this page, and the [RSS feed](#). Older news items are available in the [Galaxy News Archive](#).

### See also

- Galaxy News Briefs
- Galaxy Updates
- Galaxy on Twitter
- Events
- Learn
- Support
- About the Galaxy Project

### News Items

Environmental Metabolomics + Galaxy

A new UK-China collaboration in environmental metabolomics between the University of Birmingham, BGI and *GigaScience* has received funding from the UK's Natural Environment Research Council (NERC).

The first metabolomics project will send a developer from the University of Birmingham's School of Biosciences, to Hong Kong to work with *GigaScience* personnel on extending Galaxy for use in metabolomics data analyses.

"Metabolomics involves the detection and quantification of small molecules (metabolites) in living organisms and can provide an indication of their cellular condition and health. The toxicological responses of organisms to pollutants can be studied using environmental metabolomics, enabling researchers to discover diagnostic markers for monitoring and risk assessment of our environment. Research at Birmingham focuses extensively on the metabolic responses of the freshwater model organism, *Daphnia*, to both pollutants and engineered nanomaterials."

See the [official announcement](#) for more details.

Peter Li  
[GigaScience](#)

Posted to the [Galaxy News](#) on 2013-04-22

Galaxy @ ASMS 2013

Galaxy will have a significant presence at the 61st ASMS Conference on Mass Spectrometry and Allied Topics being held in Minneapolis, Minnesota, June 9-13. Galaxy related content includes the [Galaxy Framework as a Solution for MS-based Informatics](#) workshop and at least 9 posters either directly about or using Galaxy.

If you do research in proteomics than please consider attending.

Dave Clements

Posted to the [Galaxy News](#) on 2013-04-19

April 8, 2013 Galaxy Security Release

### Galaxy Community Conference 2013

30 June - 2 July  
OSLO University of Oslo

**Poster abstracts due 3 May**

### Use Galaxy

[Use Main \(about\)](#)  
[Use Others!](#) • [Learn](#)  
[Share](#) • [Search](#)

### Communication

[Support](#) • [News](#)   
[Events](#) • [Twitter](#)  
[Mailing Lists \(search\)](#)

### Deploy Galaxy

[Get Galaxy](#) • [Cloud Admin](#) • [Tool Config](#)  
[Tool Shed](#) • [Search](#)

### Contribute

[Tool Shed](#) • [Share Issues & Requests](#)  
[Support](#)

### Galaxy Project

[Home](#) • [About](#)  
[Community](#)  
[Big Picture](#)

### Wiki

[Help](#) • [All Pages](#)  
[Recent Changes](#)   
[Search](#) • [Create Page](#)



# The Galaxy Team



Enis Afgan



Dannon Baker



Dan Blankenberg



Dave Bouvier



Marten Cech



Dave Clements



Nate Coraor



Carl Eberhard



Dorine Francheteau



Jeremy Goecks



Sam Guerler



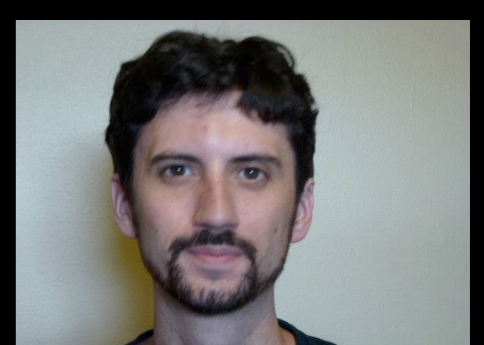
Jen Jackson



Greg von Kuster



Ross Lazarus



Nick Stoler



Anton Nekrutenko

James Taylor



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

# Acknowledgements

Dannon Baker  
Dave Clements  
Galaxy Team

UCSC CBSE  
Karen Miga  
David Bernick

## Resources

NIH NSF Huck Institute  
Penn State University  
Emory University  
TCAA

Forget everything else

usegalaxy.org

galaxyproject.org

*now let's try beta main!*

# Demo @ Beta Main

Orientation: Galaxy Masthead menus, tool pane (left), active pane (middle), history pane (right)
Review how to find and manage account: preferences stats, 1 account per user, History -> Saved
Do some more with current class exercise, using as baseline (CPB pub 'Using Galaxy, Protocol 1') Ref: <a href="http://usegalaxy.org/u/galaxyproject/p/using-galaxy-2012">http://usegalaxy.org/u/galaxyproject/p/using-galaxy-2012</a> <a href="http://vimeo.com/galaxyproject/cbp-usinggalaxy1">http://vimeo.com/galaxyproject/cbp-usinggalaxy1</a>
Workflow: extract, edit, save, add to tool menu, execute sending results to a new history
History/Datasets: delete/undelete, permanently delete, hidden, copy and related menu options
Share & Publish: user, via link, publishing, examples. Why Publish materials & methods (somewhere!)
Tool Groups: common used and how related to similar line command/SQL bioinfo methods
Wiki: Mailing lists, Support, Learning resources: screencasts/tutorials, Get Galaxy, Other Public Ref: <a href="http://galaxyproject.org">http://galaxyproject.org</a>
Various Q & A related to slides or the above