Accessible and Reproducible Data Analysis for Bench Scientists

Meetups

Galaxy is an open-source, web-based, x. Galaxy enables bench scientists to create, share, and publish sophisticated, reproducible bioinformatic analyses without requiring researchers to learn command line interfaces, or Unix system management skills. Galaxy can be accessed through the project's public server, or on one of the over 60 publicly accessible Galaxy servers. Galaxy can also be installed locally, and on cloud infrastructures.

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. Time allowing, there will also be a brief demonstration.

Galaxy is ...

Galaxy is a ... web-based, data integration and analysis platform for life science research ...

... **platform for life science research** scientists to create, share, and publish sophisticated, reproducible bioinformatic analyses without requiring researchers to learn command line interfaces, or Unix system management skills. Galaxy can be accessed through the project's public server, or on one of the over 60 publicly accessible Galaxy servers. Galaxy can also be installed locally, and on cloud infrastructures.

This talk will introduce the Galaxy platform ...

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. **there will also be a brief demonstration.**

Galaxy is ...

...

platform for life science research.

scientists to create

reproducible

researchers to learn command line interfaces or Unix system management skills. Galaxy can be accessed through the project's public server, or on one of the over 90 publicly accessible Galaxy servers. Galaxy can also be installed locally, and on cloud infrastructures.

Galaxy enables bench scientists to create ... bioinformatic analyses ...

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward.

there will also be a brief demonstration.

Basic Analysis

Which exons have most overlapping repeats in 3 spine stickelback, chromosome XXI?

(~ http://usegalaxy.org/galaxy101)

Galaxy is
analysis platform for life science research. Galaxy enables bench scientists to create,
reproducible
researchers to learn command line interfaces, or Unix system management skills. Galaxy can be accessed through the project's public servers or one of the over 60 publicly accessible Galaxy servers. Galaxy can also be installed locally, and on cloud infrastructures.

... reproducible ...

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. ——————
there will also be a brief demonstration.

Galaxy is ...

...
platform for life science research. Galaxy enables bench scientists to create,
bioinformatic analyses
command line interfaces, or Unix system management skills.
Galaxy can be accessed through the project's public server, or
on one of the over 100 publicly accessible Galaxy servers.
Galaxy can also be installed locally, and on cloud infrastructures.

... share, and publish ...

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. ——————
there will also be a brief demonstration.

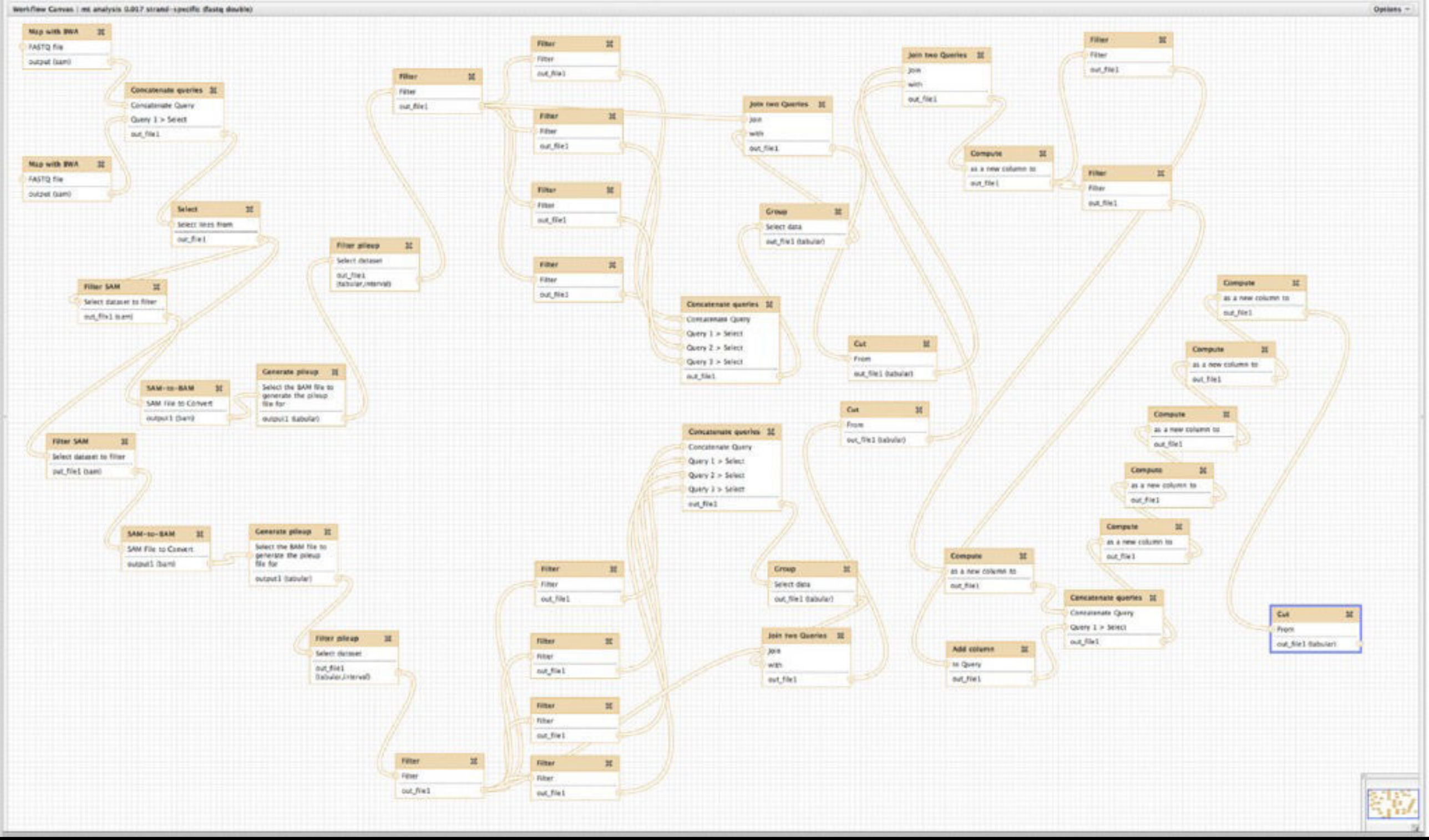
Galaxy is ...

...
platform for life science research. Galaxy enables bench scientists to create, share, and publish
bioinformatic analyses
command line interfaces, or Unix system management skills.
Galaxy can be accessed through the project's public server, or on one of the over 50 publicly accessible Galaxy servers.

... sophisticated ...

Galaxy can also be installed locally, and on cloud infrastructures.

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. ——————
there will also be a brief demonstration.



Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study, Goto *et al. Genome Biology* 2011, 12:R59
http://genomebiology.com/2011/12/6/R59

Galaxy is ...

...

platform for life science research. Galaxy enables bench scientists to create, share, and publish sophisticated, reproducible bioinformatics analyses.

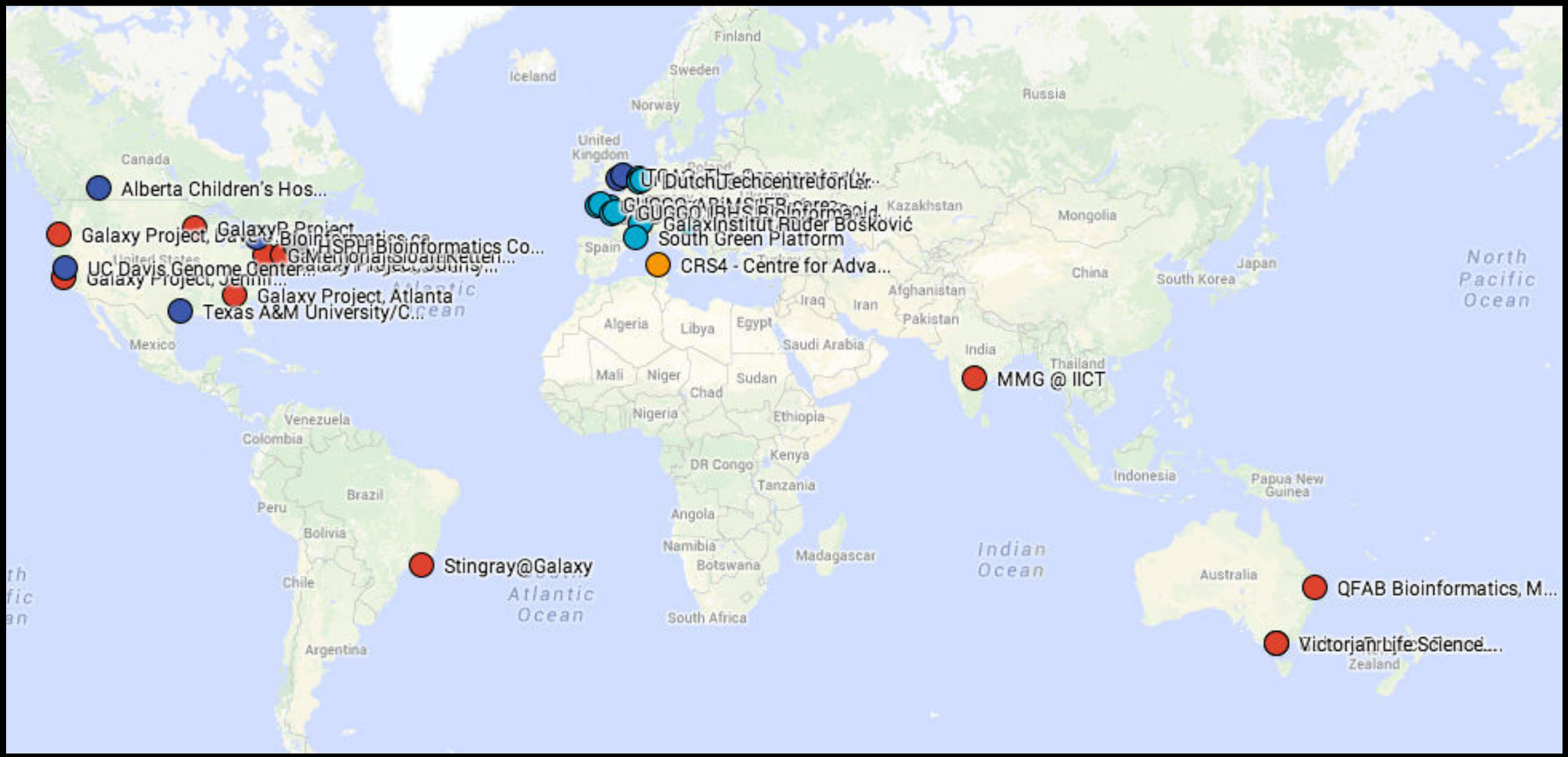
... without requiring researchers to learn command line interfaces, or Unix system management skills.

Galaxy can be accessed through the main public server, or on one of the over 60 publicly accessible Galaxy servers. Galaxy can also be installed locally and on cloud infrastructures.

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. ——————

there will also be a brief demonstration.

Galaxy Training Network



bit.ly/gxygtn

Galaxy is ...

...

platform for life science research. Galaxy enables bench scientists to create, share, and publish sophisticated, reproducible bioinformatic analyses

command line interfaces, or Unix system management skills.

Galaxy can be accessed through the project's public server

or on one of the over 60 publicly accessible Galaxy servers.

Galaxy can also be installed locally and on cloud infrastructures.

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. ——————

there will also be a brief demonstration.

Galaxy can be accessed through the project's public server ...

usegalaxy.org

Galaxy is ...

...

platform for life science research. Galaxy enables bench scientists to create, share, and publish sophisticated, reproducible bioinformatic analyses.

... open-source ...

command line interfaces, or Unix system management skills. Galaxy can be accessed through the project's public server, one of the over 50 publicly accessible Galaxy servers.

Galaxy can also be installed locally ...

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. ——————— there will also be a brief demonstration.

getgalaxy.org

Galaxy is ...

...

platform for life science research. Galaxy enables bench scientists to create, share, and publish sophisticated, reproducible bioinformatic analyses

command line interfaces, or Unix system management skills.

Galaxy can be accessed through the project's public server,

one of the over 60 publicly accessible Galaxy servers.

can also be installed locally.

... or on one of the over 60 publicly accessible Galaxy servers.

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward.

there will also be a brief demonstration.

bit.ly/gxyServers

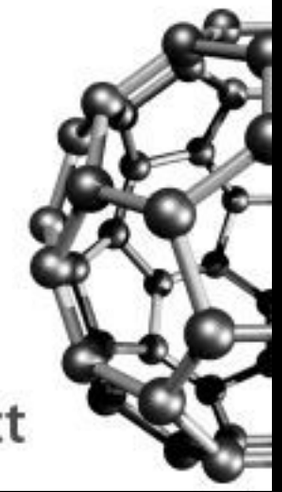
Cistrome

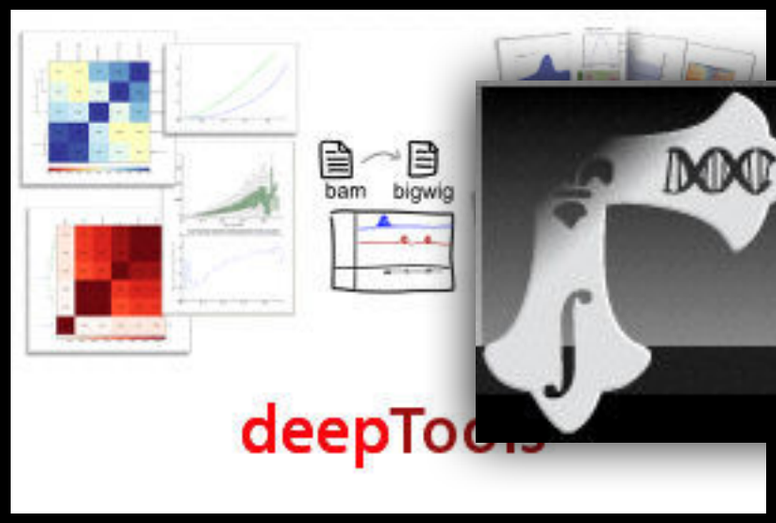
A Galaxy Server dedicated to ChIP-* analysis

ballaxy

Powered by the

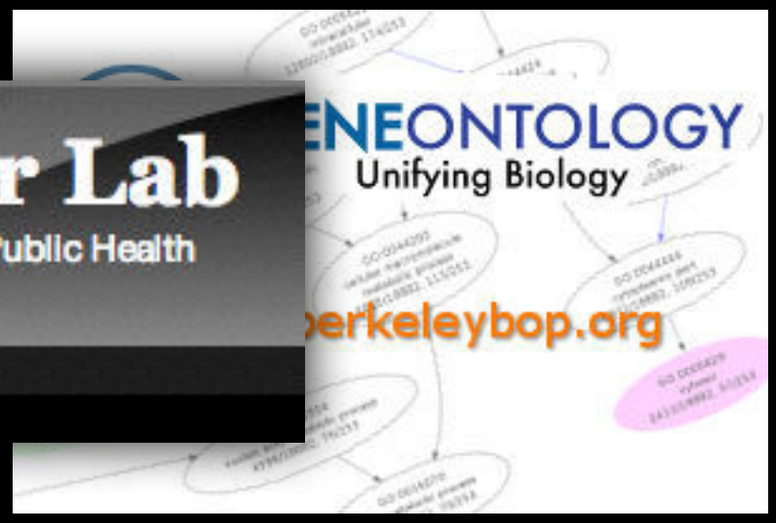
Biochemical Algorithms Library Project

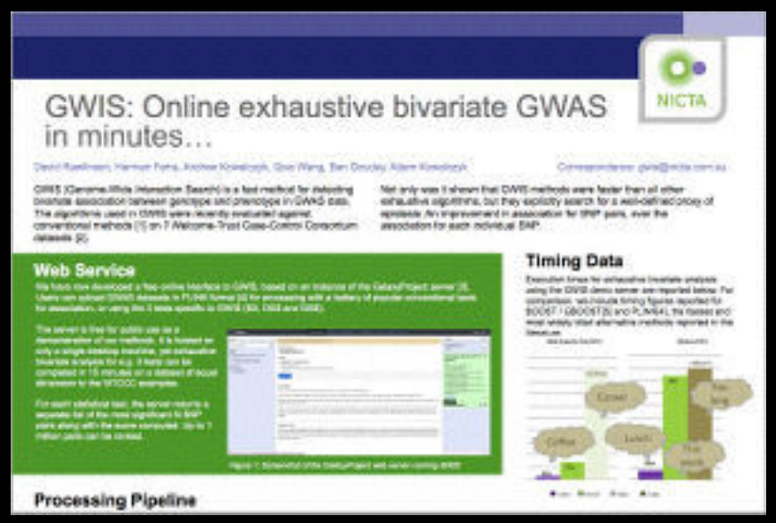


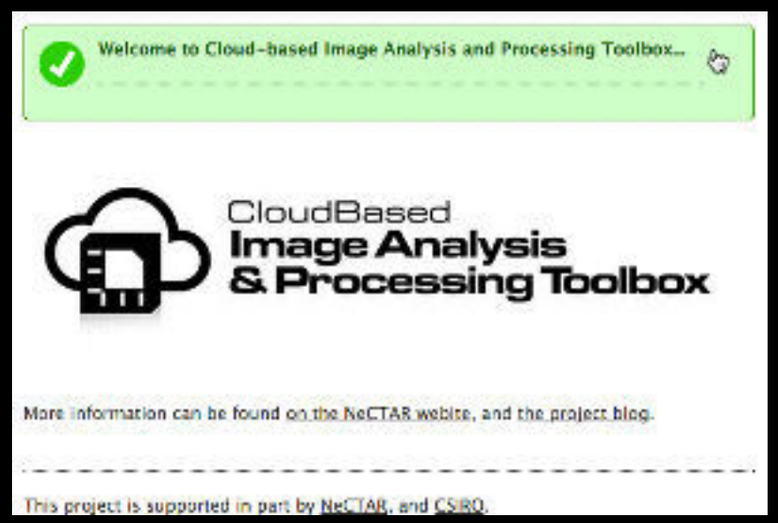


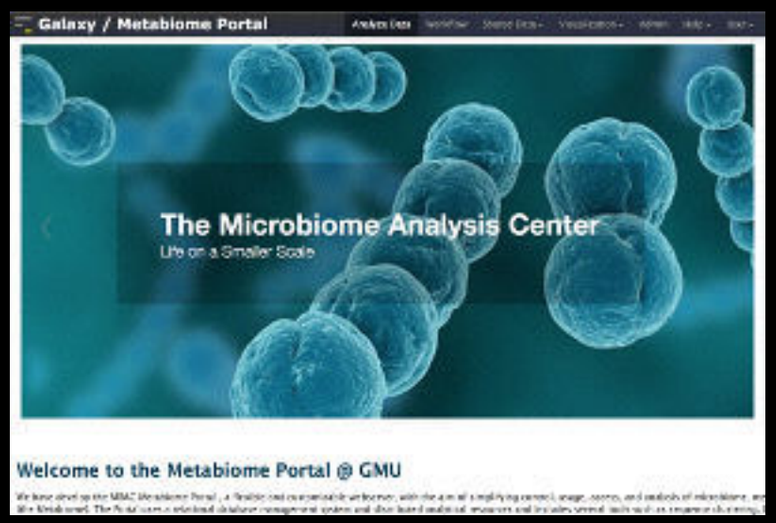
The Huttenhower Lab

Department of Biostatistics, Harvard School of Public Health









bit.ly/gxyServers

Galaxy is ...

...
platform for life science research. Galaxy enables bench scientists to create, share, and publish sophisticated, reproducible bioinformatic analyses
command line interfaces, or Unix system management skills.
Galaxy can be accessed through the browser on one of the over 60 publicly accessible Galaxy servers. Galaxy can also be installed locally,

... and on cloud infrastructures.

This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. ——————
there will also be a brief demonstration.

wiki.galaxyproject.org/Cloud

openstack™
CLOUD SOFTWARE

amazon web services™

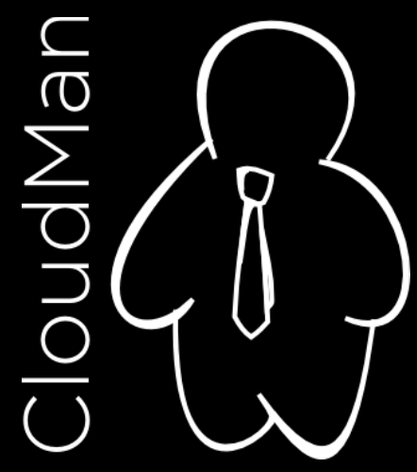
OpenNebula.org
The Open Source Toolkit for Cloud Computing

globus genomics

http://aws.amazon.com/education

http://globus.org/

http://wiki.galaxyproject.org/Cloud



Scalability ...

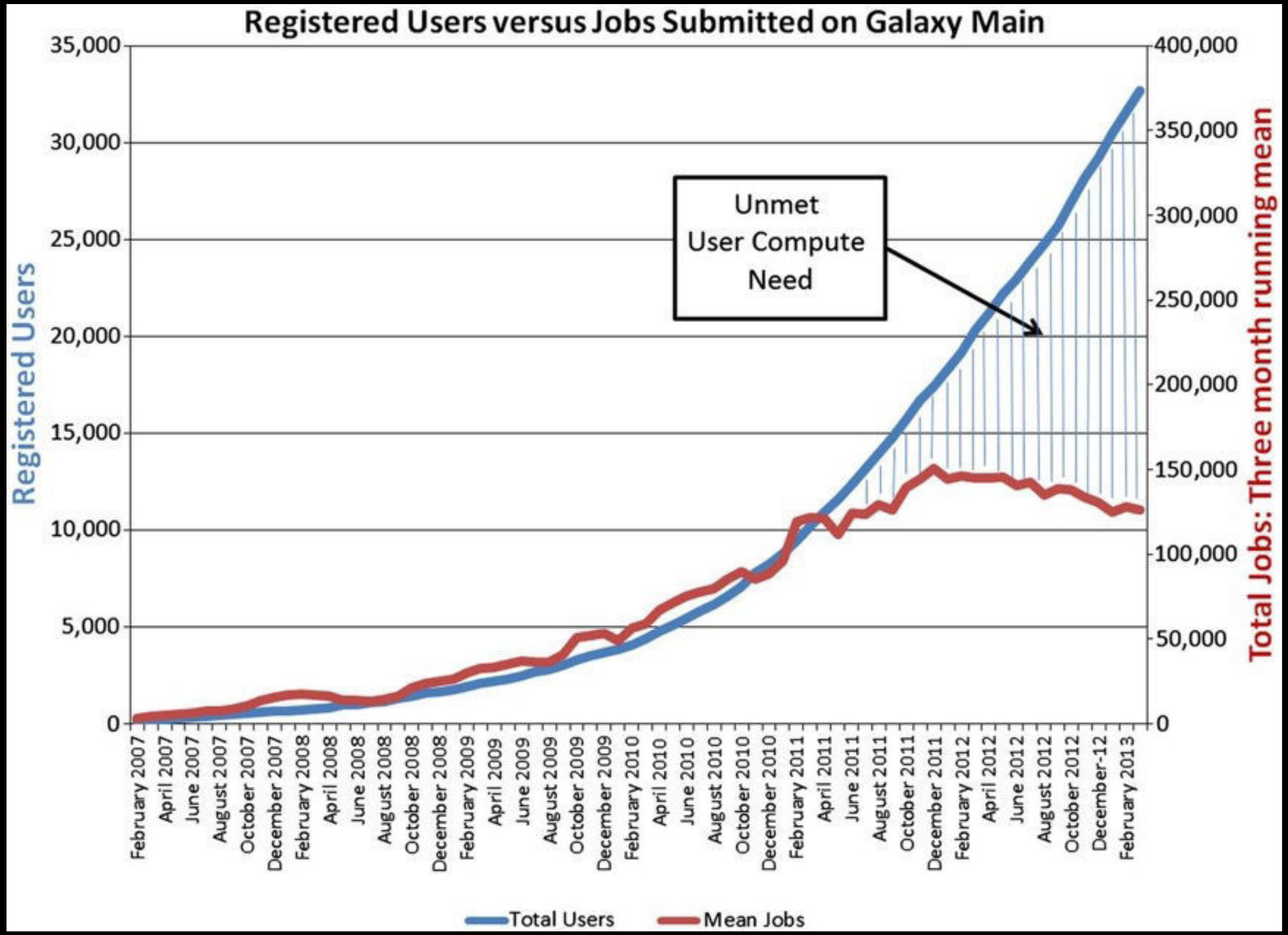
Scalability

Data generation is cheap and will stay cheap.

Scale & complexity of analysis will continue to grow.

More researchers are running bioinformatics analyses of all scales and complexities.

Galaxy needs to scale to the next few orders of magnitude.



Leveraging the national cyberinfrastructure for biomedical research

LeDuc, *et al. J Am Med Inform Assoc doi:10.1136/amiajnl-2013-002059*

- Dedicated resources
- Shared resources

ICS, Penn State

PSC, Pittsburgh

Blacklight
- 4,096 cores
- 32 TB memory

Galaxy Cluster
- 256 cores
- 2 TB memory

Rodeo
- 128 cores
- 1 TB memory

Corral/Stockyard
- 20 PB disk

SDSC, San Diego

Trestles
- 10,368 cores
- 20.7 TB memory

TACC, Austin

Stampede
- 462,462 cores
- 205 TB memory

The Galaxy Team

Enis Afgan

Dannon Baker

Dan Blankenberg

Dave Bouvier

Marten Cech

John Chilton

Dave Clements

Nate Coraor

Carl Eberhard

Jeremy Goecks

Sam Guerler

Jen Jackson

Ross Lazarus

Anton Nekrutenko

Nick Stoler

James Taylor

Nitesh Turaga

http://wiki.galaxyproject.org/GalaxyTeam

Scaling the Project: Community Gatherings