Galaxy

Accessible and Reproducible Data Analysis for Bench Scientists

Institute of Ecology and Evolution
University of Oregon
October 21, 2014

Dave Clements
Johns Hopkins University

THE INSTITUTE OF ECOLOGY AND EVOLUTION

UNIVERSITY OF OREGON

Galaxy

Galaxy is an open-source, web-based, data integration and analysis platform for life science research. 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 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 enables bench scientists to create ... bioinformatic analyses ...

... 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

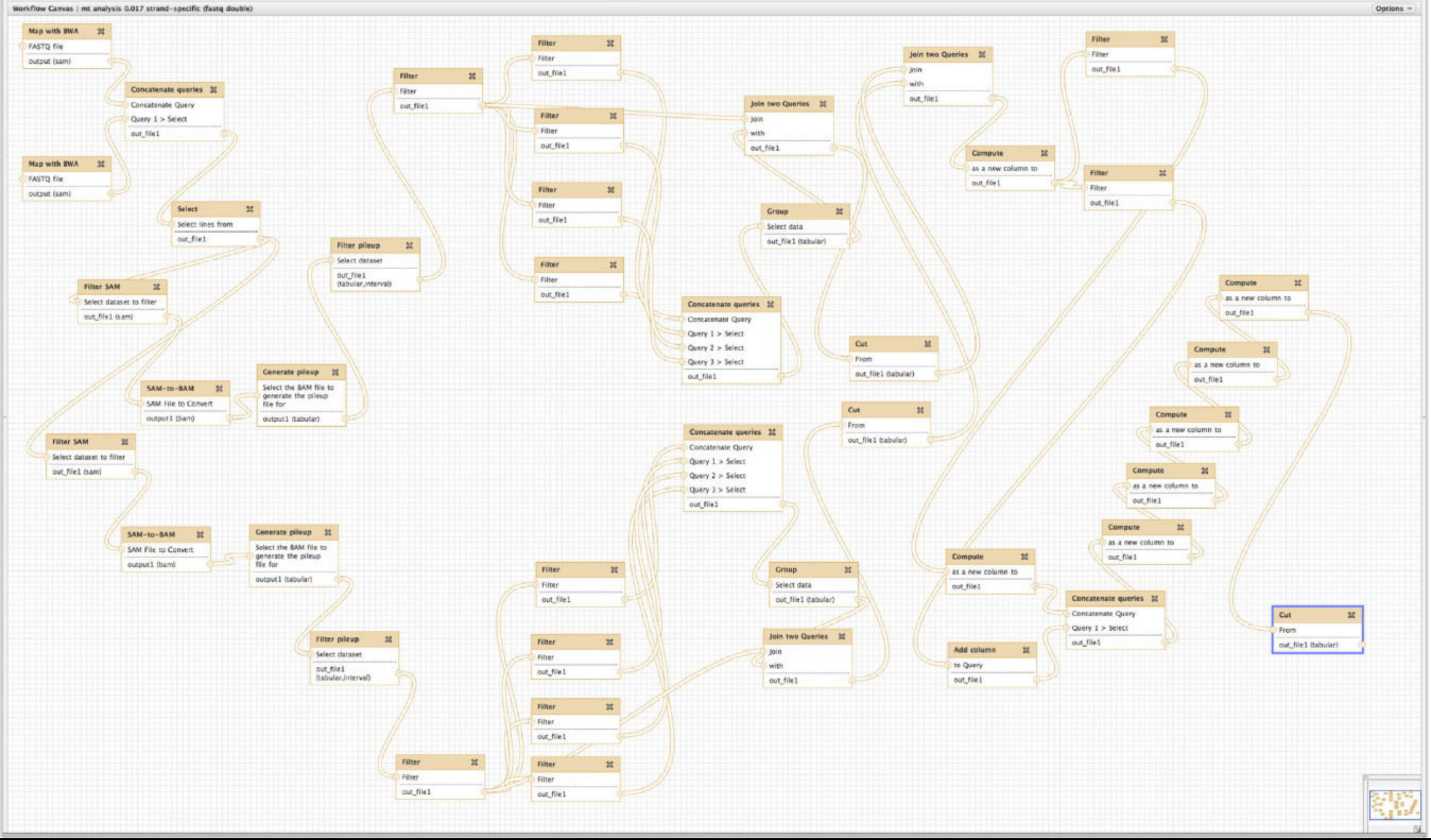
command line interfaces, or Unix system management skills. Galaxy can be accessed through the project's public server, or on one of the over 50 publicly accessible Galaxy servers.

... 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 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,

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

.... discuss the project's recent work and plans going forward.

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.

Semantic Scalability: Dataset Collections

Make Galaxy aware of how datasets are related.

Build workflows that can reason about paired datasets, technical replicates, multiple biological samples, ...

Run tools once on each dataset in the collection.
Run tools on the collection as a whole.

Support map/reduce paradigm.

Scaling for the Bioinformatician: Galaxy API

Scaling up also requires support for bioinformaticians and core staff.

Allows compute-savvy researchers to use scripting and still get the reproducibility, sharing, and publishing advantages of Galaxy.

galaxy-dist.readthedocs.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

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

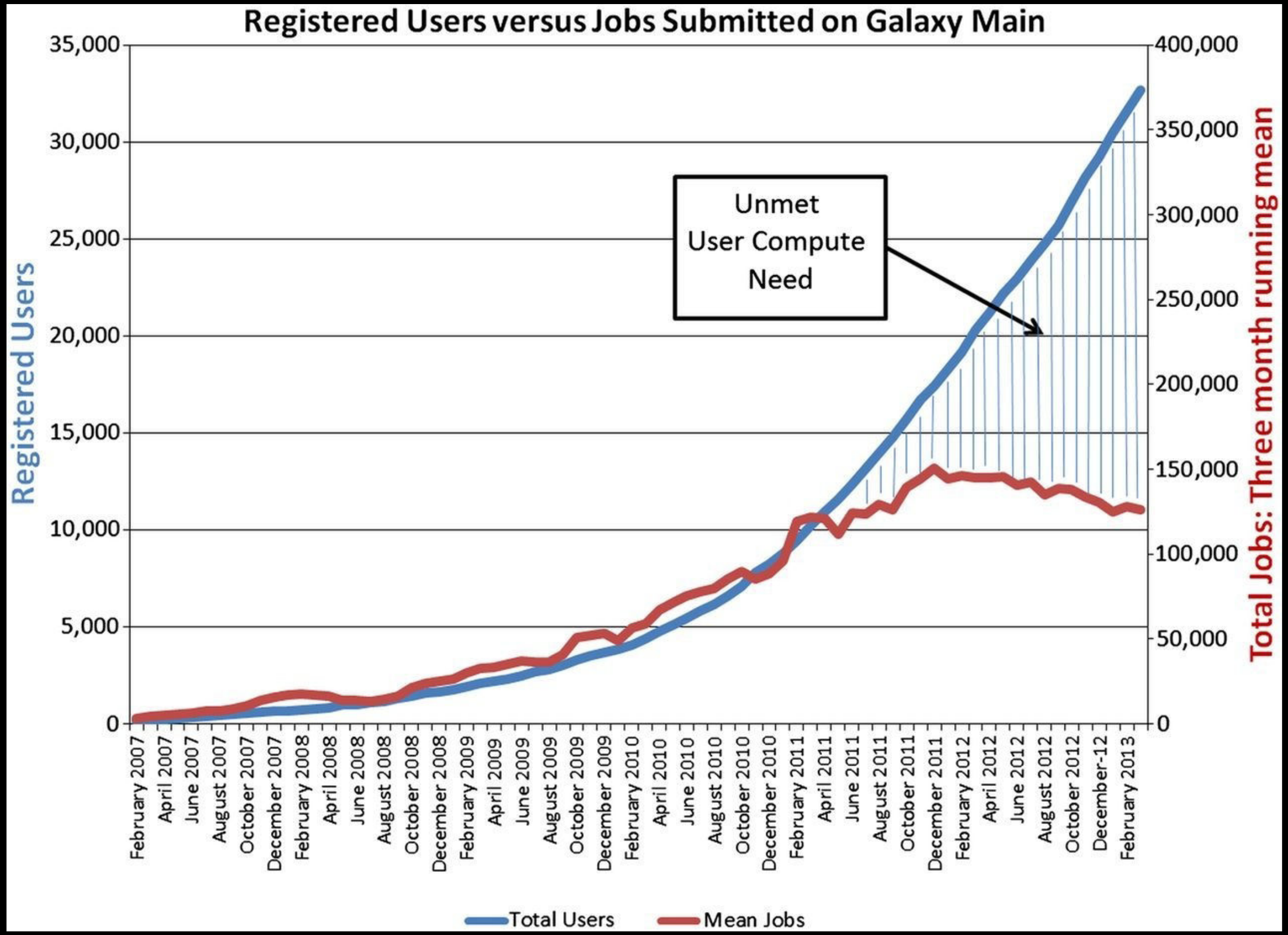
Galaxy can also be installed locally and on cloud infrastructures.

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



Leveraging the national cyberinfrastructure for biomedical research

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

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,

Galaxy can also be installed

one of the over 50 publicly accessible Galaxy servers.

can also be installed locally.

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.

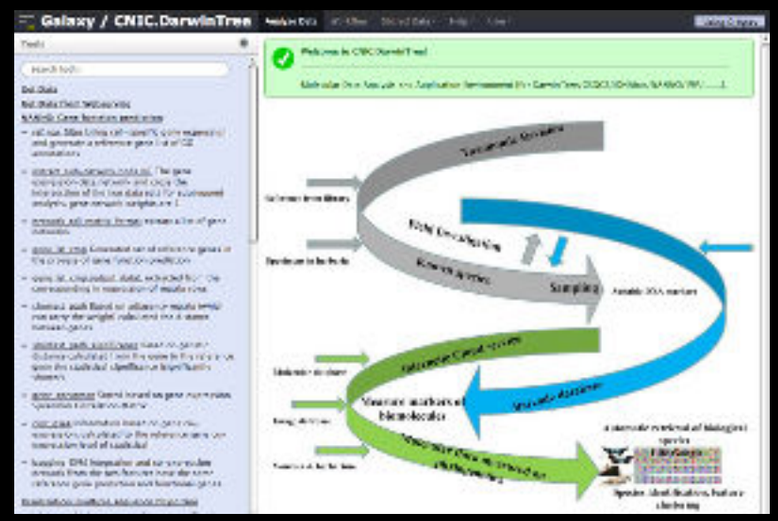
This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward.

there will also be a brief demonstration.

bit.ly/gxyServers

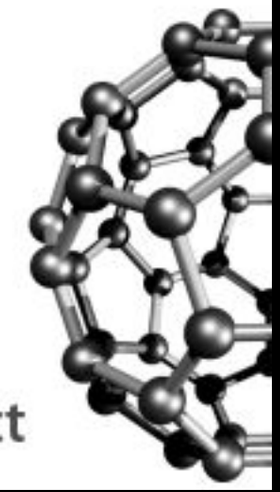
Cistrome

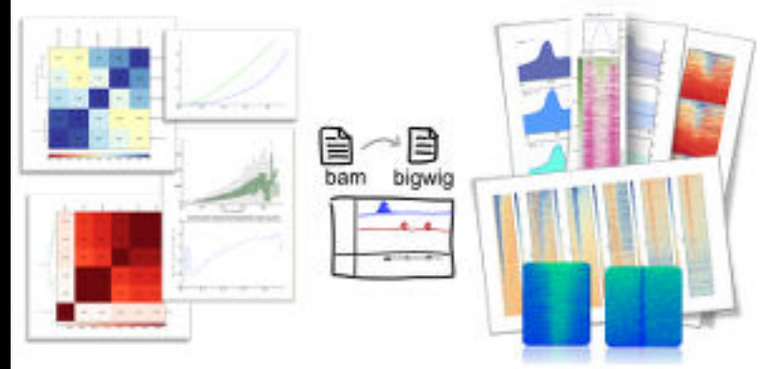
A Galaxy Server dedicated to ChIP-* analysis



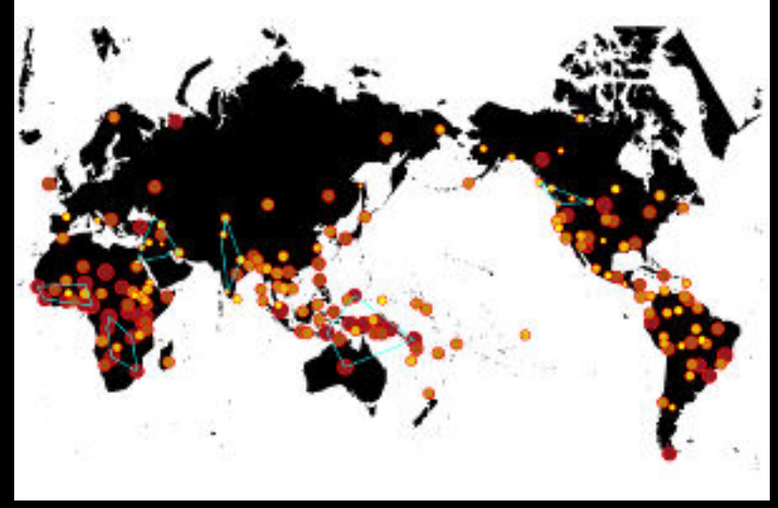
ballaxy

Powered by the Biochemical Algorithms Library Project



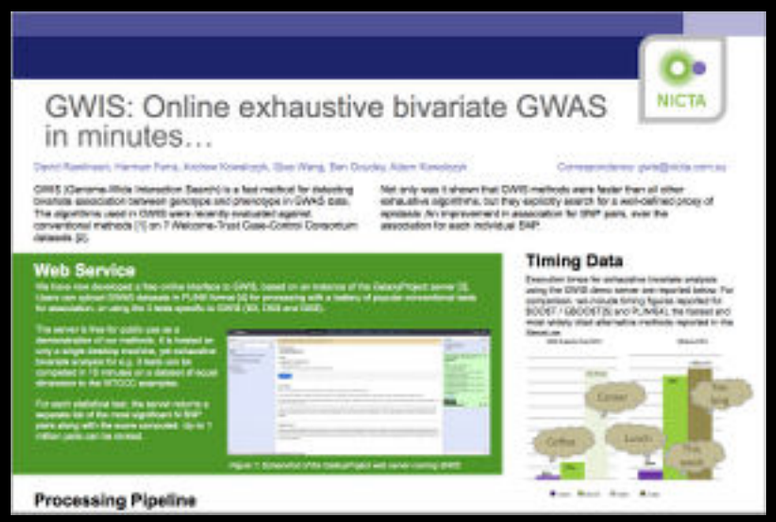


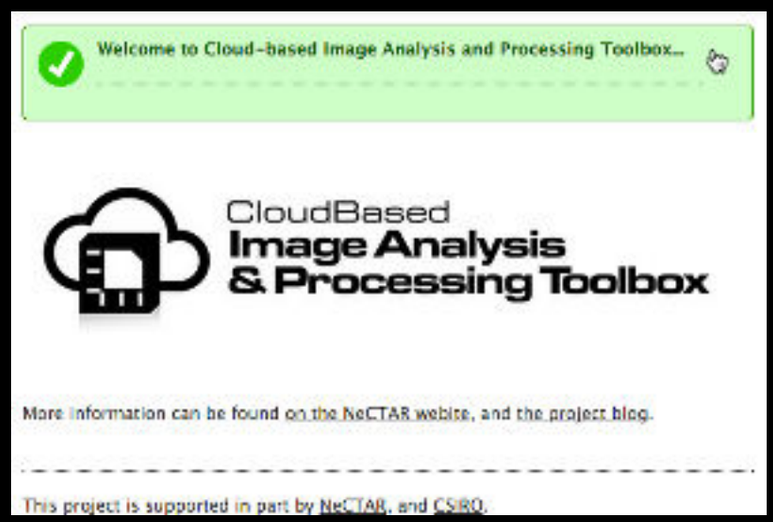
deepTools

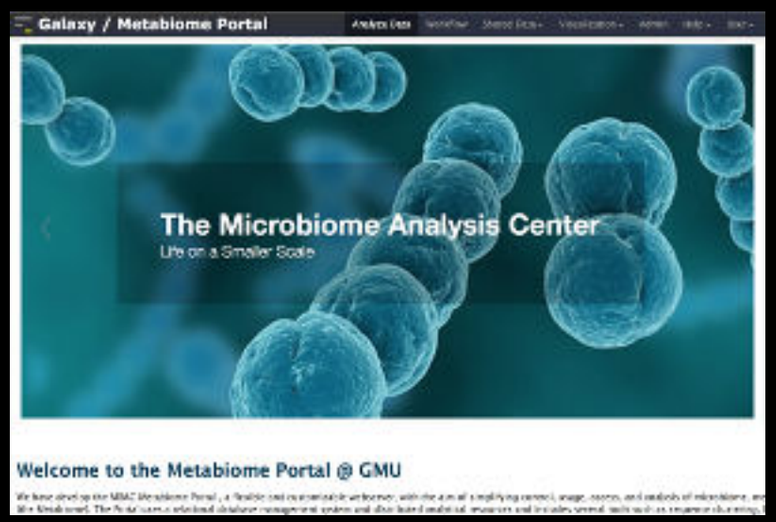


GENEONTOLOGY
Unifying Biology

galaxy.berkeleybop.org



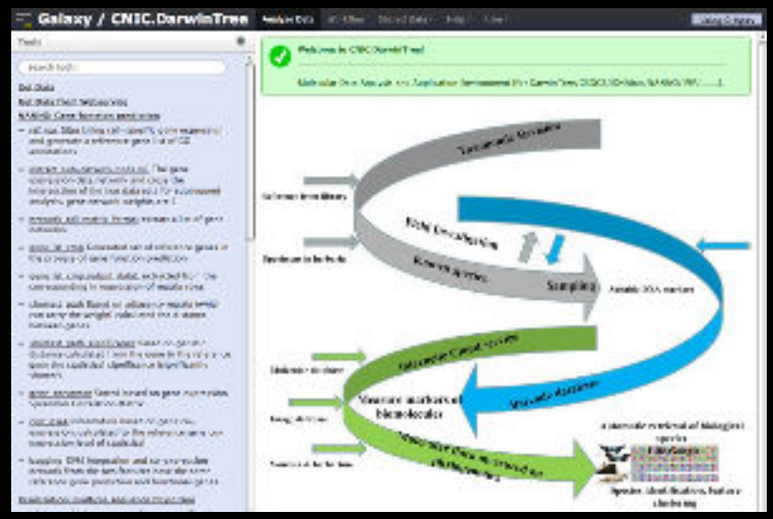




bit.ly/gxyServers

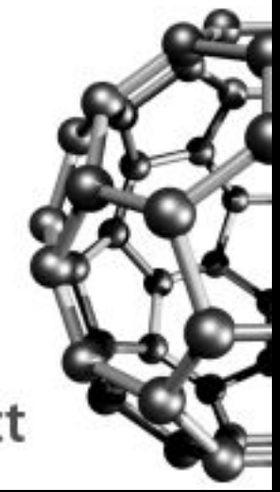
Cistrome

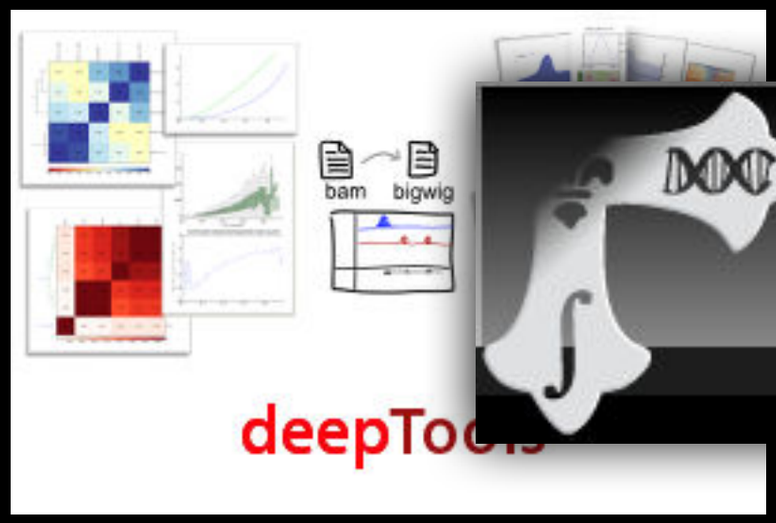
A Galaxy Server dedicated to ChIP-* analysis



ballaxy

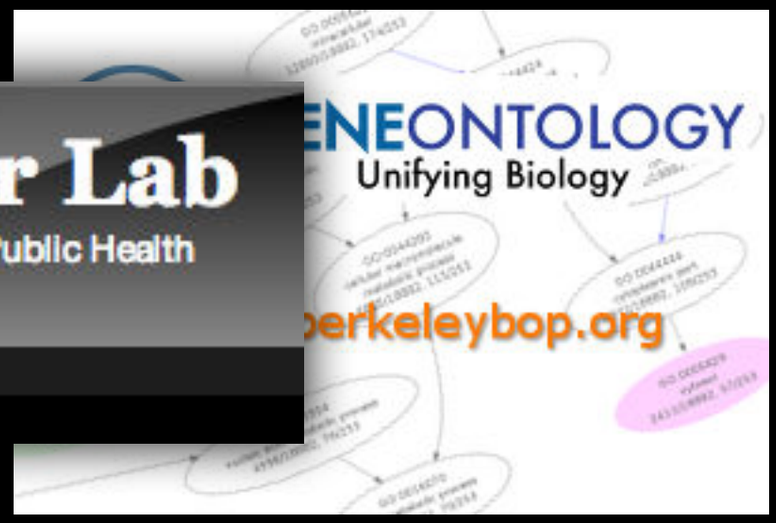
Powered by the Biochemical Algorithms Library Project

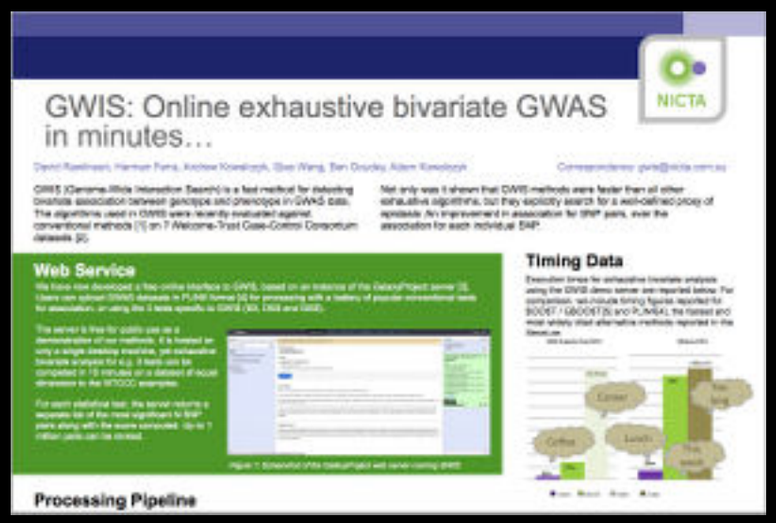


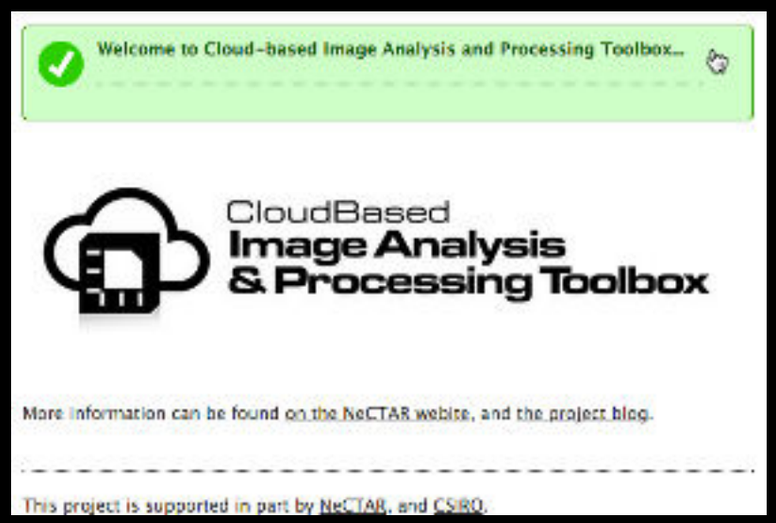


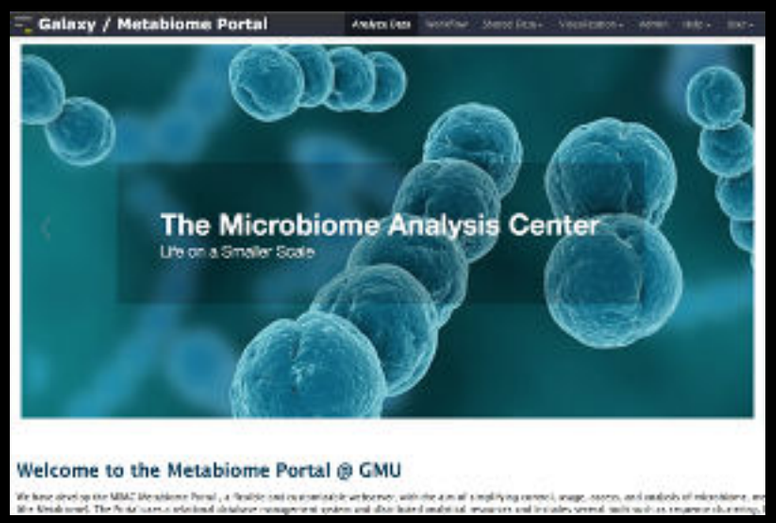
The Huttenhower Lab

Department of Biostatistics, Harvard School of Public Health









bit.ly/gxyServers

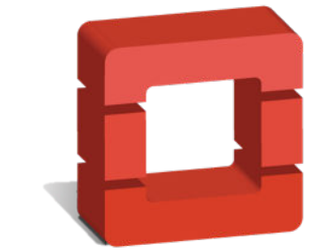
Galaxy is ...

...
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 web 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

Scaling the Project: Community Gatherings

Scaling the Project: Support

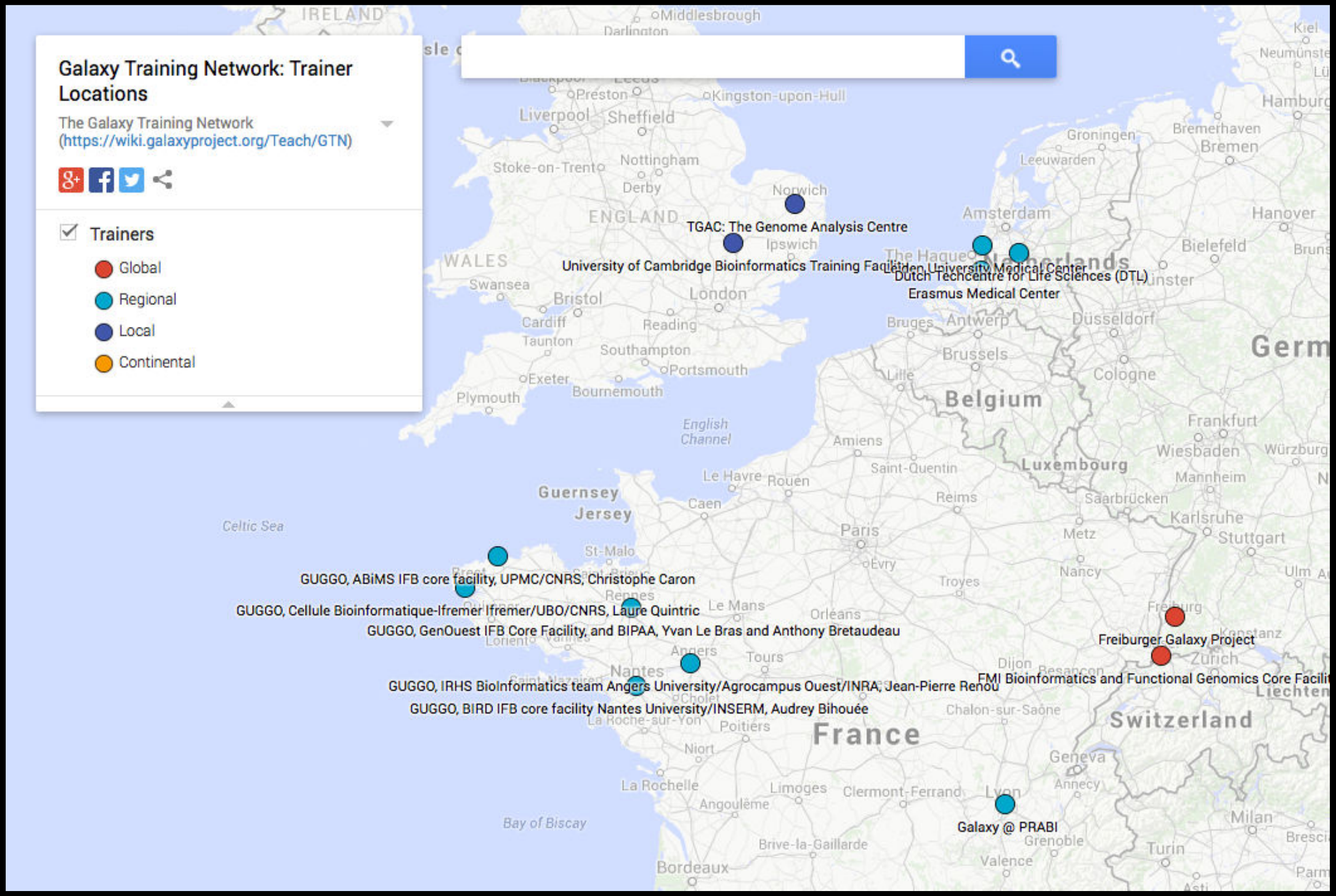
Tens of thousands of users leads to a lot of questions.

Absolutely have to encourage community support.

Project traditionally used mailing list

This year we moved user support to Galaxy Biostar, a gamefied online forum.

Scaling the Project: Training



Galaxy Training Network launched last week.

bit.ly/gxygtn

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

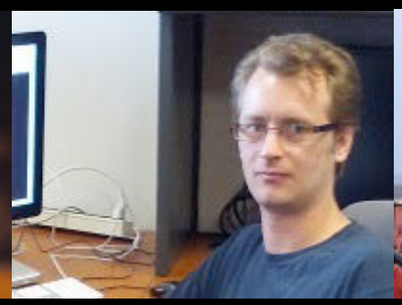
This talk will introduce the Galaxy platform and discuss the project's recent work and plans going forward. Time allowing, there will also be a brief demonstration.

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

Galaxy is hiring post-docs and software engineers

Please help.

http://wiki.galaxyproject.org/GalaxyIsHiring

Also Thanks To

THE INSTITUTE OF ECOLOGY AND EVOLUTION

Heather Archer
Michelle Wood
John Conery

Patrick Phillips
Bill Cresko

CAS

GEOL
Geological Sciences

Doug Toomey
Emilie Hooft

Thanks

Dave Clements

Galaxy Project

Johns Hopkins University

clements@galaxyproject.org