Galaxy for Core Facilities

Association of Biomolecular Resource Facilities (ABRF2013)
March 4, 2013

ABRF2013
TOOLS FOR THE ADVANCEMENT OF CONVERGENCE SCIENCE

Galaxy

Dave Clements, Dannon Baker
Emory University
http://galaxyproject.org/

Talk Complements ABRF Satellite on Saturday

- Saturday featured a live demo
- This talk introduces Galaxy and how it can benefit core facilities.
- For detailed hands-on training material, see the Galaxy learning hub page:

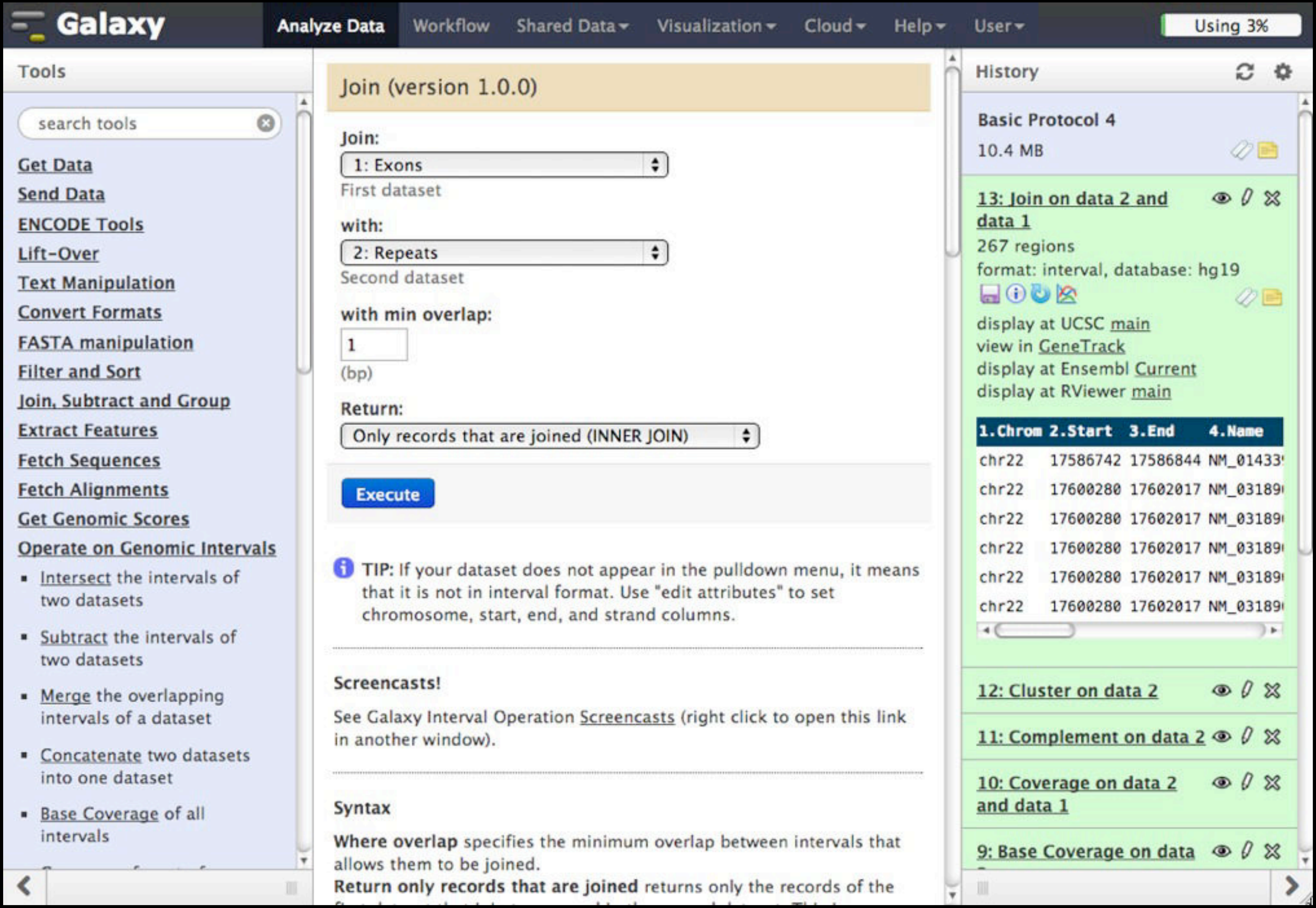
http://wiki.galaxyproject.org/Learn

What is Galaxy?

- An open, web-based platform for **accessible, reproducible,** and **transparent** computational biomedical research.
- Galaxy is available as:
  - **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
  - **Free cloud images** that can be deployed by informatics novices

http://galaxyproject.org

A free for everyone web service: usegalaxy.org



Get Data
Send Data
ENCODE Tools
Lift-Over
Text Manipulation
Convert Formats
FASTA manipulation
Filter and Sort
Join, Subtract and Group
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals

- Intersect the intervals of two datasets
- Subtract the intervals of two datasets
- Merge the overlapping intervals of a dataset
- Concatenate two datasets into one dataset
- Base Coverage of all intervals

Join (version 1.0.0)

Join: 1: Exons
First dataset

with: 2: Repeats
Second dataset

with min overlap: 1
(bp)

Return: Only records that are joined (INNER JOIN)

TIP: If your dataset does not appear in the pulldown menu, it means that it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.

Screencasts!

See Galaxy Interval Operation Screencasts (right click to open this link in another window).

Syntax

Where overlap specifies the minimum overlap between intervals that allows them to be joined.

Return only records that are joined returns only the records of the

Basic Protocol 4
10.4 MB

13: Join on data 2 and data 1
267 regions
format: interval, database: hg19
display at UCSC main
view in GeneTrack
display at Ensembl Current
display at RViewer main

1.Chrom	2.Start	3.End	4.Name
chr22	17586742	17586844	NM_01433
chr22	17600280	17602017	NM_03189
chr22	17600280	17602017	NM_03189
chr22	17600280	17602017	NM_03189
chr22	17600280	17602017	NM_03189
chr22	17600280	17602017	NM_03189

12: Cluster on data 2
11: Complement on data 2
10: Coverage on data 2 and data 1
9: Base Coverage on data

Open Source Software: getgalaxy.org

- Galaxy is designed for local installation and customization
  - Easily integrate new tools
  - Requires a computational resource on which to be deployed

http://getgalaxy.org

Local Galaxy Installs require a computational resource on which to be deployed

Control where tool execution happens

Galaxy works with any DRMAA compliant cluster job scheduler (which is most of them).

Galaxy is just another client to your scheduler.

CLUSTER RESOURCES
TORQUE

GRID ENGINE

Platform Computing

PBS Works

DRMAA
Distributed Resource Management
Application API — www.drmaa.org

Galaxy is available ...

- As open source software

  http://getgalaxy.org

- *On the Cloud*

  http://usegalaxy.org/cloud

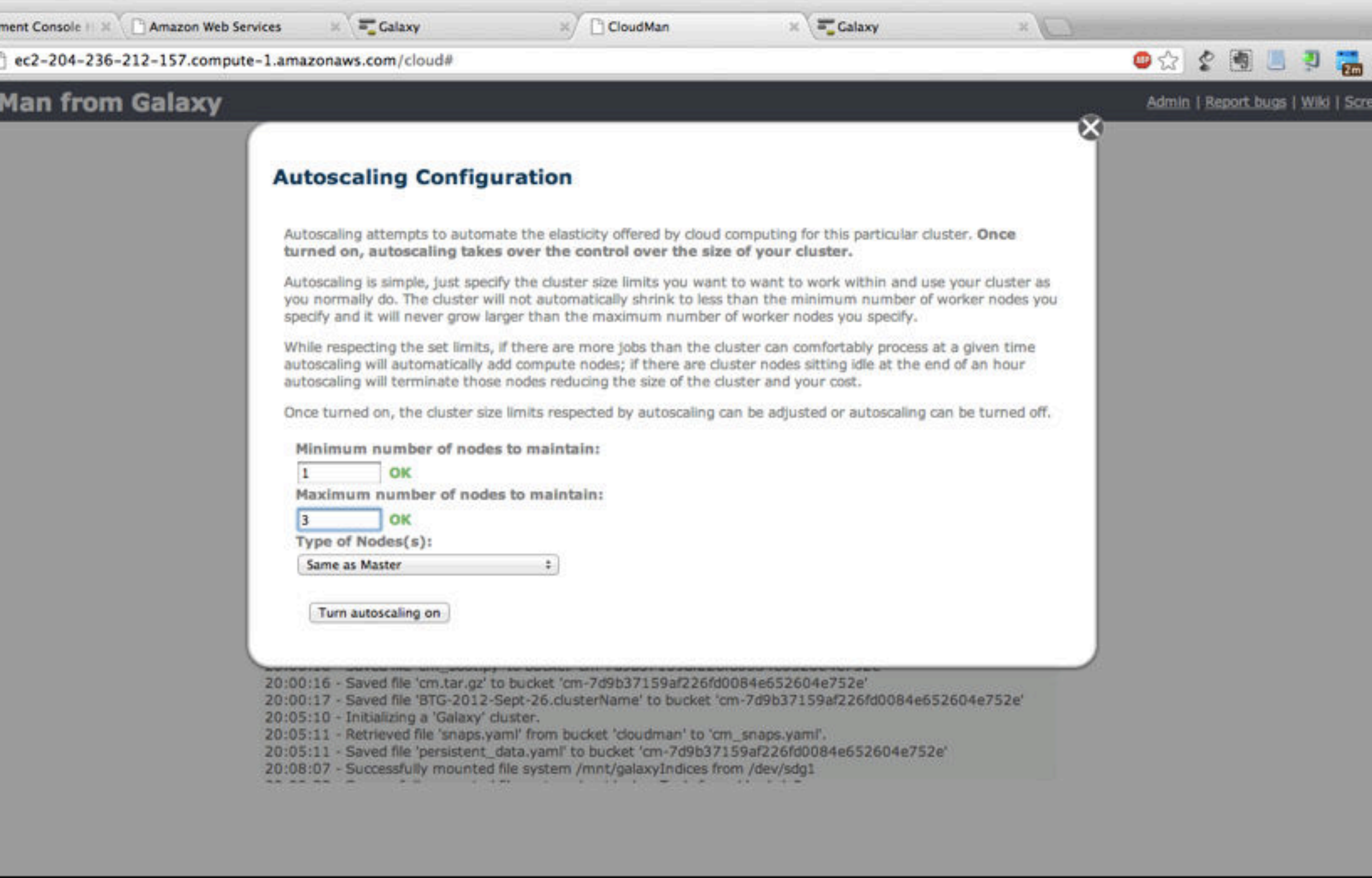
  We used this during the satellite workshop.

http://aws.amazon.com/education

On the Cloud: 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

Galaxy CloudMan



Autoscaling Configuration

Autoscaling attempts to automate the elasticity offered by cloud computing for this particular cluster. **Once turned on, autoscaling takes over the control over the size of your cluster.**

Autoscaling is simple, just specify the cluster size limits you want to want to work within and use your cluster as you normally do. The cluster will not automatically shrink to less than the minimum number of worker nodes you specify and it will never grow larger than the maximum number of worker nodes you specify.

While respecting the set limits, if there are more jobs than the cluster can comfortably process at a given time autoscaling will automatically add compute nodes; if there are cluster nodes sitting idle at the end of an hour autoscaling will terminate those nodes reducing the size of the cluster and your cost.

Once turned on, the cluster size limits respected by autoscaling can be adjusted or autoscaling can be turned off.

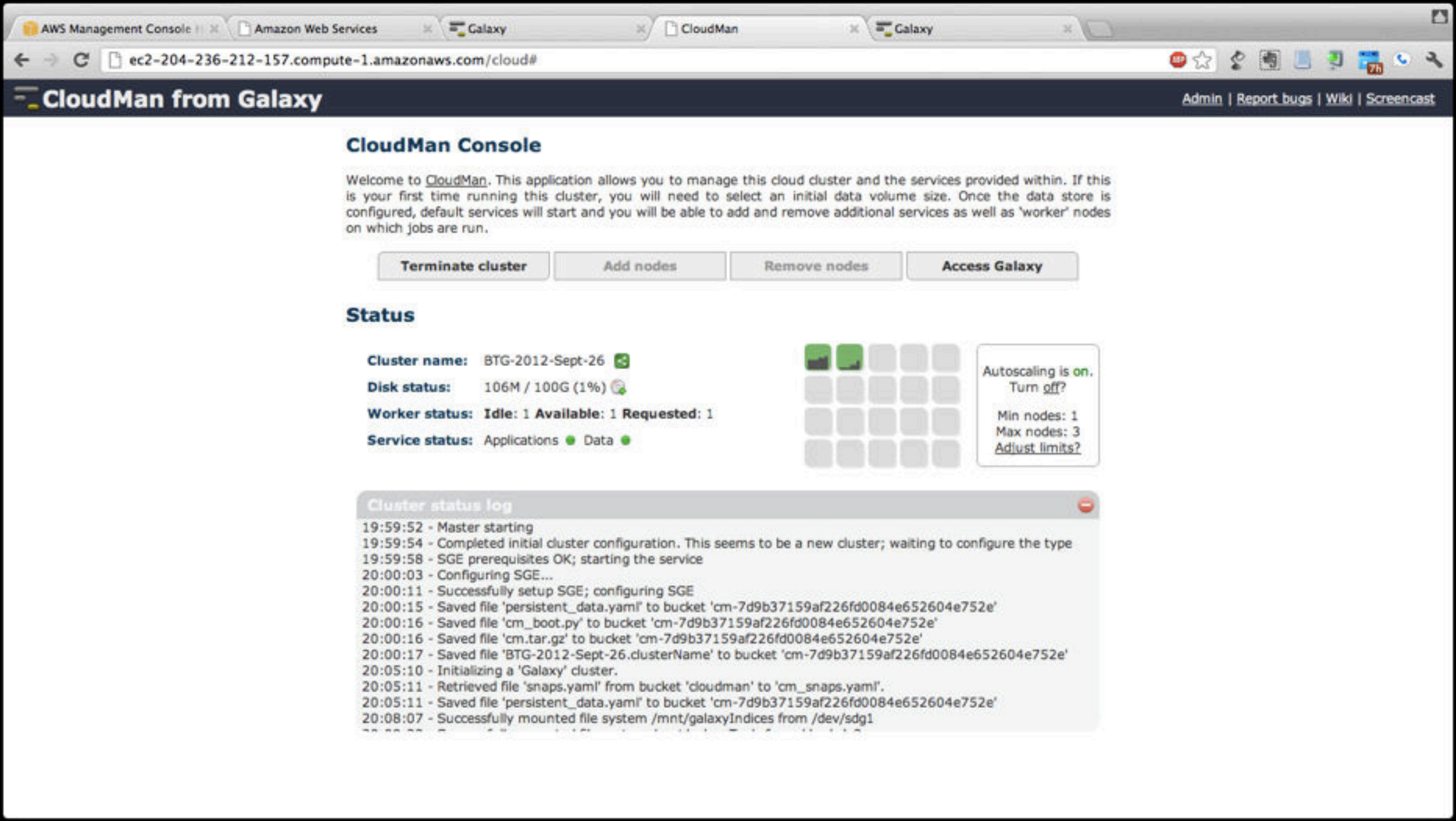
Minimum number of nodes to maintain: 1 OK

Maximum number of nodes to maintain: 3 OK

Type of Nodes(s): Same as Master

```
20:00:16 - Saved file 'cm.tar.gz' to bucket 'cm-7d9b37159af226fd0084e652604e752e'
20:00:17 - Saved file 'BTG-2012-Sept-26.clusterName' to bucket 'cm-7d9b37159af226fd0084e652604e752e'
20:05:10 - Initializing a 'Galaxy' cluster.
20:05:11 - Retrieved file 'snaps.yaml' from bucket 'cloudman' to 'cm_snaps.yaml'.
20:05:11 - Saved file 'persistent_data.yaml' to bucket 'cm-7d9b37159af226fd0084e652604e752e'
20:08:07 - Successfully mounted file system /mnt/galaxyIndices from /dev/sdg1
```

Galaxy CloudMan



CloudMan Console

Welcome to CloudMan. This application allows you to manage this cloud cluster and the services provided within. If this is your first time running this cluster, you will need to select an initial data volume size. Once the data store is configured, default services will start and you will be able to add and remove additional services as well as 'worker' nodes on which jobs are run.

Status

Cluster name: BTG-2012-Sept-26

Disk status: 106M / 100G (1%)

Worker status: **Idle:** 1 **Available:** 1 **Requested:** 1

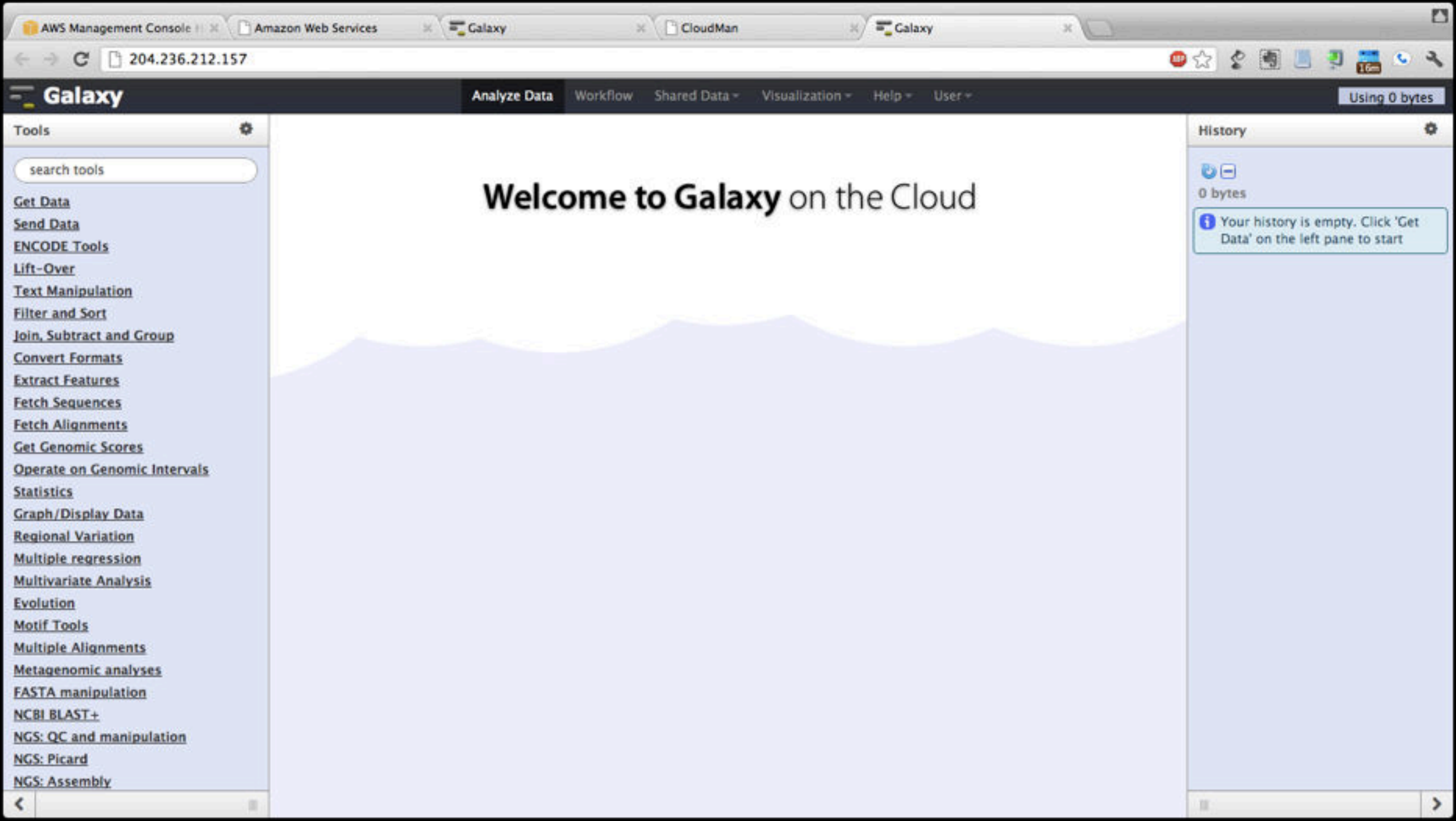
Service status: Applications ● Data ●

Autoscaling is on.
Turn off?

Min nodes: 1
Max nodes: 3
Adjust limits?

```
19:59:52 - Master starting
19:59:54 - Completed initial cluster configuration. This seems to be a new cluster; waiting to configure the type
19:59:58 - SGE prerequisites OK; starting the service
20:00:03 - Configuring SGE...
20:00:11 - Successfully setup SGE; configuring SGE
20:00:15 - Saved file 'persistent_data.yaml' to bucket 'cm-7d9b37159af226fd0084e652604e752e'
20:00:16 - Saved file 'cm_boot.py' to bucket 'cm-7d9b37159af226fd0084e652604e752e'
20:00:16 - Saved file 'cm.tar.gz' to bucket 'cm-7d9b37159af226fd0084e652604e752e'
20:00:17 - Saved file 'BTG-2012-Sept-26.clusterName' to bucket 'cm-7d9b37159af226fd0084e652604e752e'
20:05:10 - Initializing a 'Galaxy' cluster.
20:05:11 - Retrieved file 'snaps.yaml' from bucket 'cloudman' to 'cm_snaps.yaml'.
20:05:11 - Saved file 'persistent_data.yaml' to bucket 'cm-7d9b37159af226fd0084e652604e752e'
20:08:07 - Successfully mounted file system /mnt/galaxyIndices from /dev/sdg1
```

Galaxy CloudMan



Welcome to Galaxy on the Cloud

Your history is empty. Click 'Get Data' on the left pane to start

Why Galaxy?

Galaxy was originally designed for researchers to learn and run tools and iteratively build multi-step analyses.

Biology has rapidly become data intensive, and dependent on computational methods

How can we ensure that these methods are **accessible** to researchers?

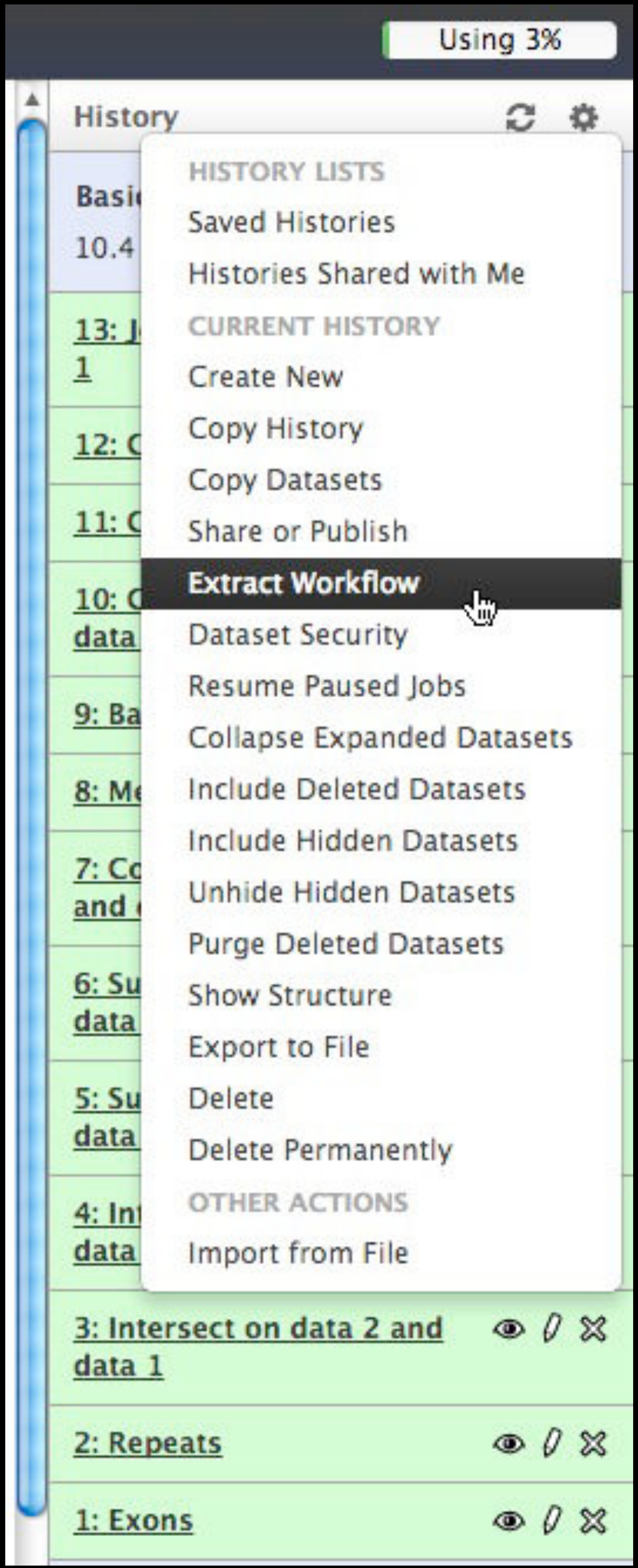
...while also ensuring that scientific results remain **reproducible**?

Why Galaxy

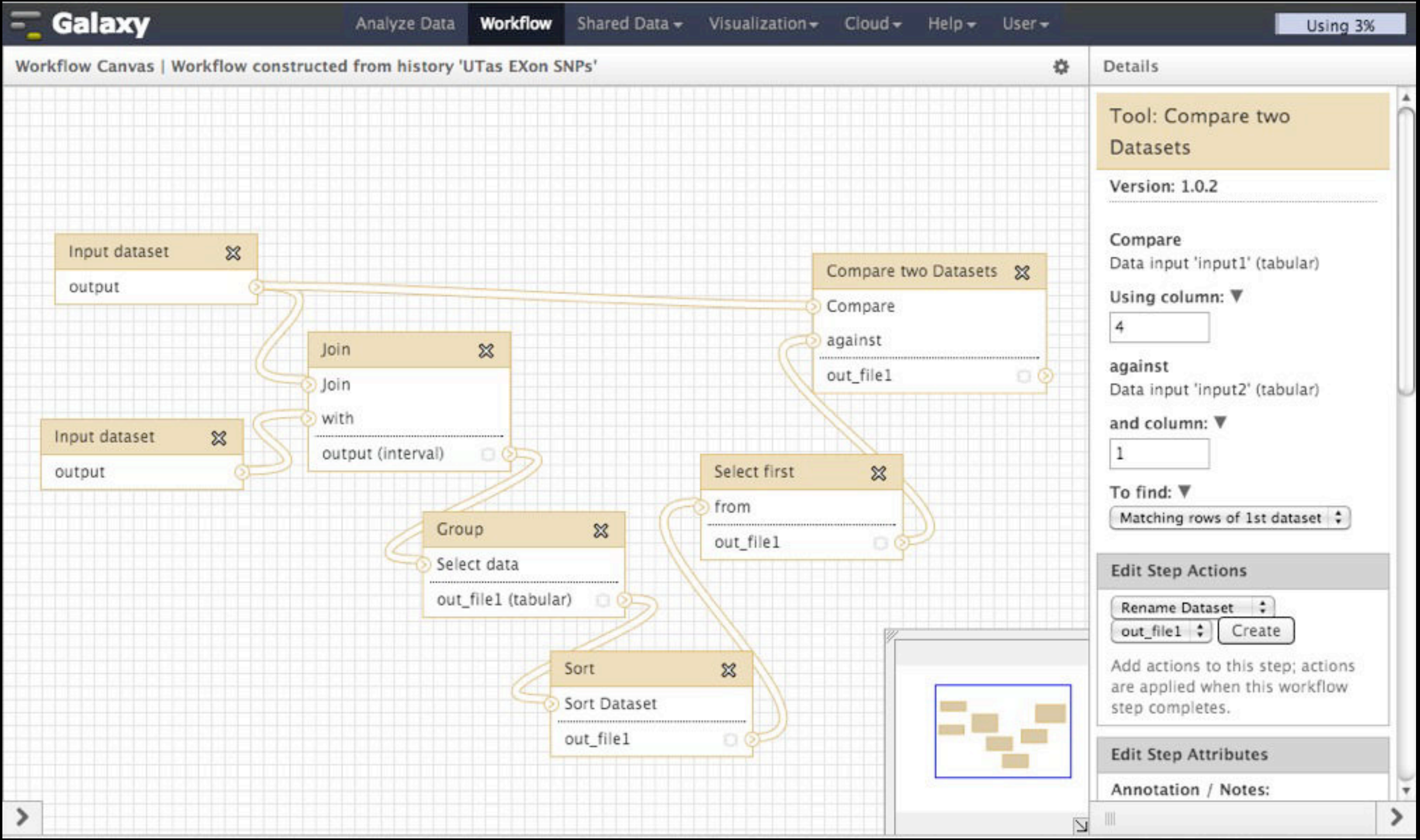
for Core Facilities?

1. Automate Your Analysis Pipelines

Define a reusable workflow (pipeline) by building/running an analysis interactively, and then extracting a reusable workflow from that analysis.



Can also define a workflow (pipeline) by building it directly using the workflow editor.



Workflow Canvas | Workflow constructed from history 'UTas EXon SNPs'

Tool: Compare two Datasets

Version: 1.0.2

Compare
Data input 'input1' (tabular)

Using column: 4

against
Data input 'input2' (tabular)

and column: 1

To find: Matching rows of 1st dataset

Add actions to this step; actions are applied when this workflow step completes.

Annotation / Notes:

Or define a workflow by loading a workflow definition via the Galaxy Application Programming Interface (API)

```
{
    "a_galaxy_workflow": "true",
    "annotation": "",
    "format-version": "0.1",
    "name": "Workflow constructed from history 'UTas EXon SNPs'",
    "steps": {
        "0": {
            "annotation": "",
            "id": 0,
            "input_connections": {},
            "inputs": [
                {
                    "description": "",
                    "name": "Input Dataset"
                }
            ],
            "name": "Input dataset",
            "outputs": [],
            "position": {
                "left": 10,
                "top": 10
            },
            "tool_errors": null,
            "tool_id": null,
            "tool_state": "{\"name\": \"Input Dataset\"}",
            "tool_version": null,
            "type": "data_input",
            "user_outputs": []
        },
        "1": {
            "annotation": "",
            "id": 1,
            "input_connections": {},
            "inputs": [
                {
                    "description": "",
                    "name": "Input Dataset"
                }
            ],
            "name": "Input dataset",
            "outputs": [],
```

Galaxy API

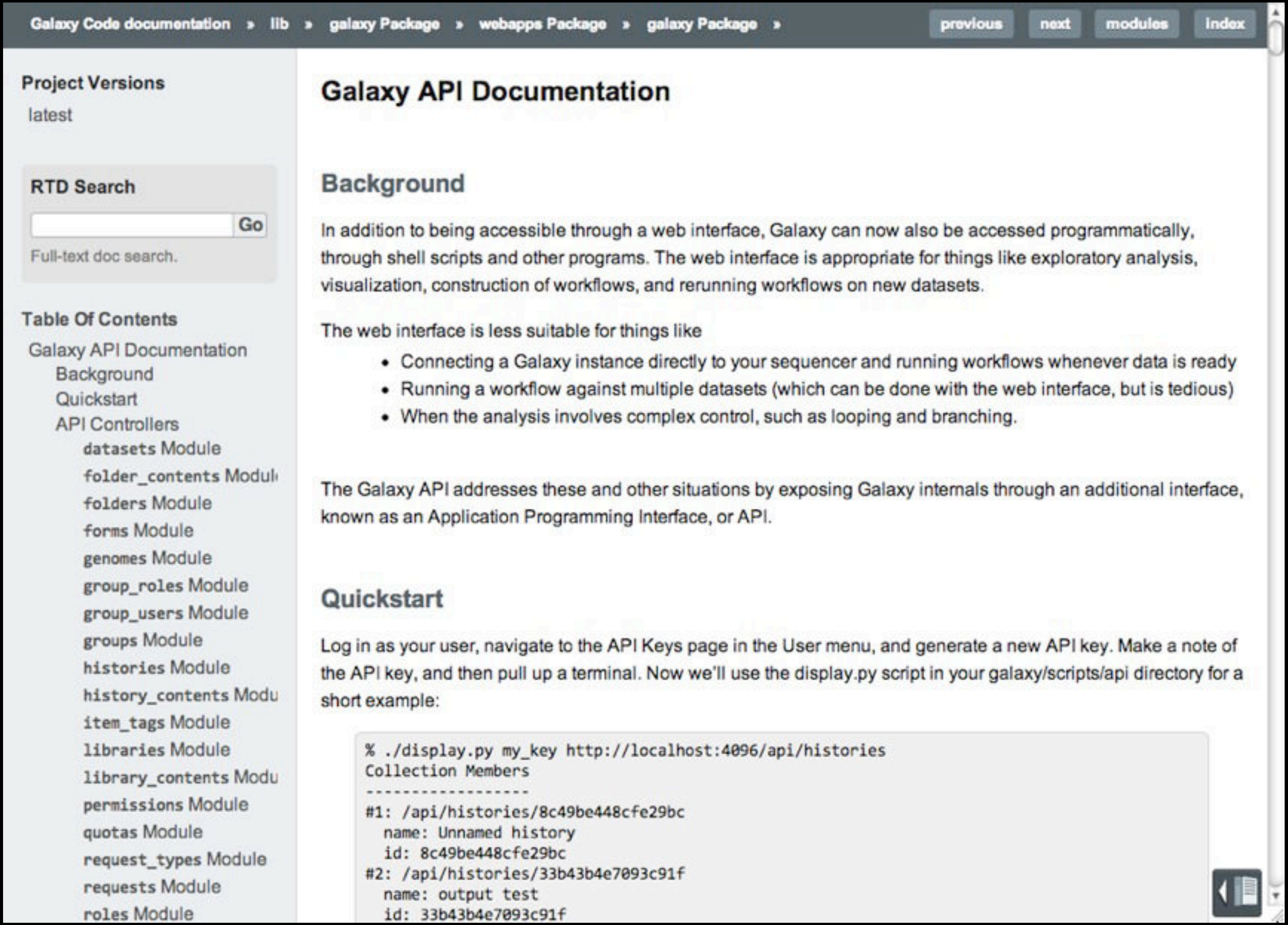


Table Of Contents

Galaxy API Documentation

Background

In addition to being accessible through a web interface, Galaxy can now also be accessed programmatically, through shell scripts and other programs. The web interface is appropriate for things like exploratory analysis, visualization, construction of workflows, and rerunning workflows on new datasets.

The web interface is less suitable for things like

- Connecting a Galaxy instance directly to your sequencer and running workflows whenever data is ready
- Running a workflow against multiple datasets (which can be done with the web interface, but is tedious)
- When the analysis involves complex control, such as looping and branching.

The Galaxy API addresses these and other situations by exposing Galaxy internals through an additional interface, known as an Application Programming Interface, or API.

Quickstart

Log in as your user, navigate to the API Keys page in the User menu, and generate a new API key. Make a note of the API key, and then pull up a terminal. Now we'll use the display.py script in your galaxy/scripts/api directory for a short example:

```
% ./display.py my_key http://localhost:4096/api/histories
Collection Members
------------------
#1: /api/histories/8c49be448cfe29bc
  name: Unnamed history
  id: 8c49be448cfe29bc
#2: /api/histories/33b43b4e7093c91f
  name: output test
  id: 33b43b4e7093c91f
```

https://galaxy-dist.readthedocs.org/

Once the workflow is defined you can run it repeatedly with different inputs.

- Via the web interface
- Via the API

Simple example script that watches a folder for new files, imports that data to a data library, and then execute a workflow on it, creating a new history for each workflow invocation.

These

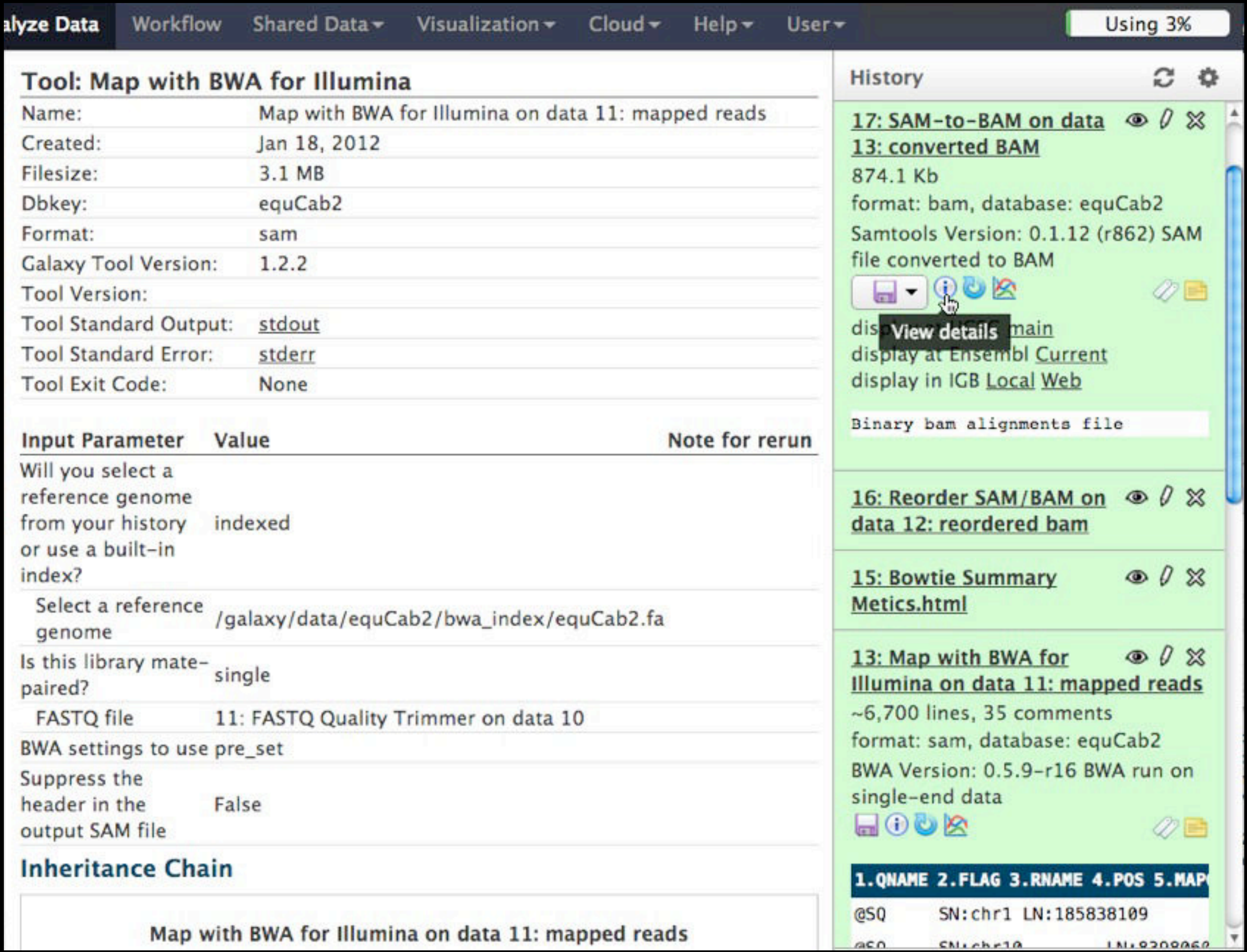
47 lines of (well commented) Python

automatically start an analysis whenever files are ready for processing.

```
#!/usr/bin/env python
"""
Simple example script that watches a folder for new files, imports that data to a data library, and then
execute a workflow on it, creating a new history for each workflow invocation.

This assumes a workflow with only one input, though it could be adapted to many.

Sample call:
python example_watch_folder.py <api_key> <api_url> /tmp/g_inbox/ /tmp/g_inbox/done/ "API Imports" f2db41e1fa331b3e

NOTE:  The upload method used requires the data library filesystem upload allow_library_path_paste
"""
import os
import shutil
import sys
import time
sys.path.insert( 0, os.path.dirname( __file__ ) )
from common import submit, display

def main(api_key, api_url, in_folder, out_folder, data_library, workflow):
    # Find/Create data library with the above name.  Assume we're putting datasets in the root folder '/'
    libs = display(api_key, api_url + 'libraries', return_formatted=False)
    library_id = None
    for library in libs:
        if library['name'] == data_library:
            library_id = library['id']
    if not library_id:
        lib_create_data = {'name':data_library}
        library = submit(api_key, api_url + 'libraries', lib_create_data, return_formatted=False)
        library_id = library[0]['id']
    folders = display(api_key, api_url + "libraries/%s/contents" % library_id, return_formatted = False)
    for f in folders:
        if f['name'] == "/":
            library_folder_id = f['id']
    workflow = display(api_key, api_url + 'workflows/%s' % workflow, return_formatted = False)
    if not workflow:
        print "Workflow %s not found, terminating."
```

```
        for fname in os.listdir(in_folder):
            fullpath = os.path.join(in_folder, fname)
            if os.path.isfile(fullpath):
                data = {}
                data['folder_id'] = library_folder_id
                data['file_type'] = 'auto'
                data['dbkey'] = ''
                data['upload_option'] = 'upload_paths'
                data['filesystem_paths'] = fullpath
                data['create_type'] = 'file'
                libset = submit(api_key, api_url + "libraries/%s/contents" % library_id, data, return_formatted = Fo
                #TODO Handle this better, but the datatype isn't always
                # set for the followup workflow execution without this
                # pause.
                time.sleep(5)
                for ds in libset:
                    if 'id' in ds:
                        # Successful upload of dataset, we have the ldda now.  Run the workflow.
                        wf_data = {}
                        wf_data['workflow_id'] = workflow['id']
                        wf_data['history'] = "%s - %s" % (fname, workflow['name'])
                        wf_data['ds_map'] = {}
                        for step_id, ds_in in workflow['inputs'].iteritems():
                            wf_data['ds_map'][step_id] = {'src':'ld', 'id':ds['id']}
                        res = submit( api_key, api_url + 'workflows', wf_data, return_formatted=False)
                        if res:
                            print res
                            # Successful workflow execution, safe to move dataset.
                            shutil.move(fullpath, os.path.join(out_folder, fname))
        time.sleep(10)

if __name__ == '__main__':
    try:
        api_key = sys.argv[1]
        api_url = sys.argv[2]
        in_folder = sys.argv[3]
        out_folder = sys.argv[4]
        data_library = sys.argv[5]
        workflow = sys.argv[6]
    except IndexError:
        print 'usage: %s key url in_folder out_folder data_library workflow' % os.path.basename( sys.argv[0] )
        sys.exit( 1 )
    main(api_key, api_url, in_folder, out_folder, data_library, workflow )
```

2. Enable clients to see analysis details



Tool: Map with BWA for Illumina

Name:	Map with BWA for Illumina on data 11: mapped reads
Created:	Jan 18, 2012
Filesize:	3.1 MB
Dbkey:	equCab2
Format:	sam
Galaxy Tool Version:	1.2.2
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	None

Input Parameter	Value	Note for rerun
Will you select a reference genome from your history or use a built-in index?	indexed	
Select a reference genome	/galaxy/data/equCab2/bwa_index/equCab2.fa	
Is this library mate-paired?	single	
FASTQ file	11: FASTQ Quality Trimmer on data 10	
BWA settings to use	pre_set	
Suppress the header in the output SAM file	False	

Inheritance Chain

Map with BWA for Illumina on data 11: mapped reads

History

17: SAM-to-BAM on data 13: converted BAM
874.1 Kb
format: bam, database: equCab2
Samtools Version: 0.1.12 (r862) SAM file converted to BAM
display at UCSC main
display at Ensembl Current
display in IGB Local Web

`Binary bam alignments file`

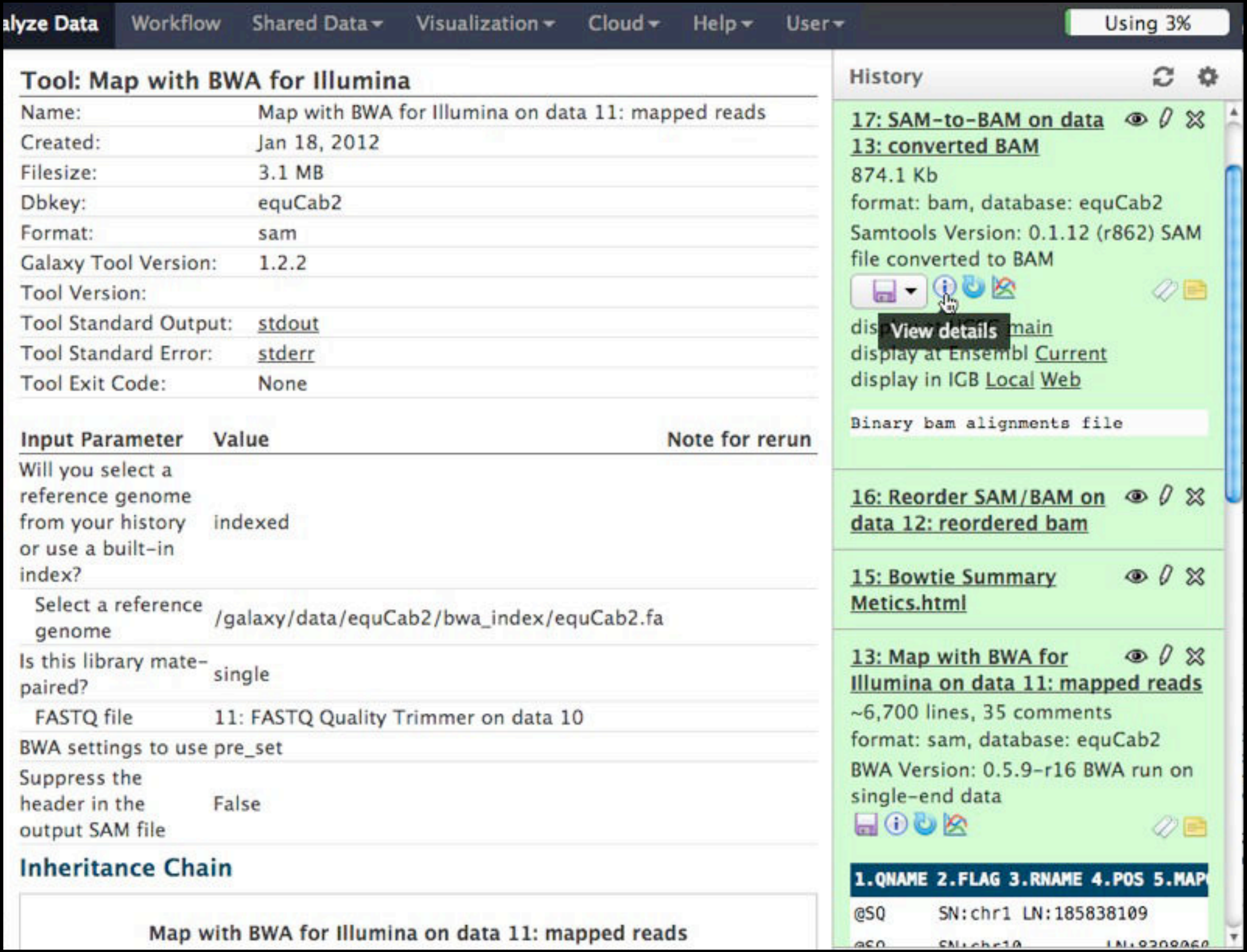
16: Reorder SAM/BAM on data 12: reordered bam

15: Bowtie Summary Metics.html

13: Map with BWA for Illumina on data 11: mapped reads
~6,700 lines, 35 comments
format: sam, database: equCab2
BWA Version: 0.5.9-r16 BWA run on single-end data

```
1.QNAME 2.FLAG 3.RNAME 4.POS 5.MAP
@SQ     SN:chr1 LN:185838109
```

3. Galaxy's provenance and reproducibility for free



Tool: Map with BWA for Illumina

Name:	Map with BWA for Illumina on data 11: mapped reads
Created:	Jan 18, 2012
Filesize:	3.1 MB
Dbkey:	equCab2
Format:	sam
Galaxy Tool Version:	1.2.2
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	None

Input Parameter	Value	Note for rerun
Will you select a reference genome from your history or use a built-in index?	indexed	
Select a reference genome	/galaxy/data/equCab2/bwa_index/equCab2.fa	
Is this library mate-paired?	single	
FASTQ file	11: FASTQ Quality Trimmer on data 10	
BWA settings to use	pre_set	
Suppress the header in the output SAM file	False	

Inheritance Chain

Map with BWA for Illumina on data 11: mapped reads

17: SAM-to-BAM on data 13: converted BAM

874.1 Kb

format: bam, database: equCab2

Samtools Version: 0.1.12 (r862) SAM file converted to BAM

display at UCSC main

display at Ensembl Current

display in IGB Local Web

```
Binary bam alignments file
```

16: Reorder SAM/BAM on data 12: reordered bam

15: Bowtie Summary Metics.html

13: Map with BWA for Illumina on data 11: mapped reads

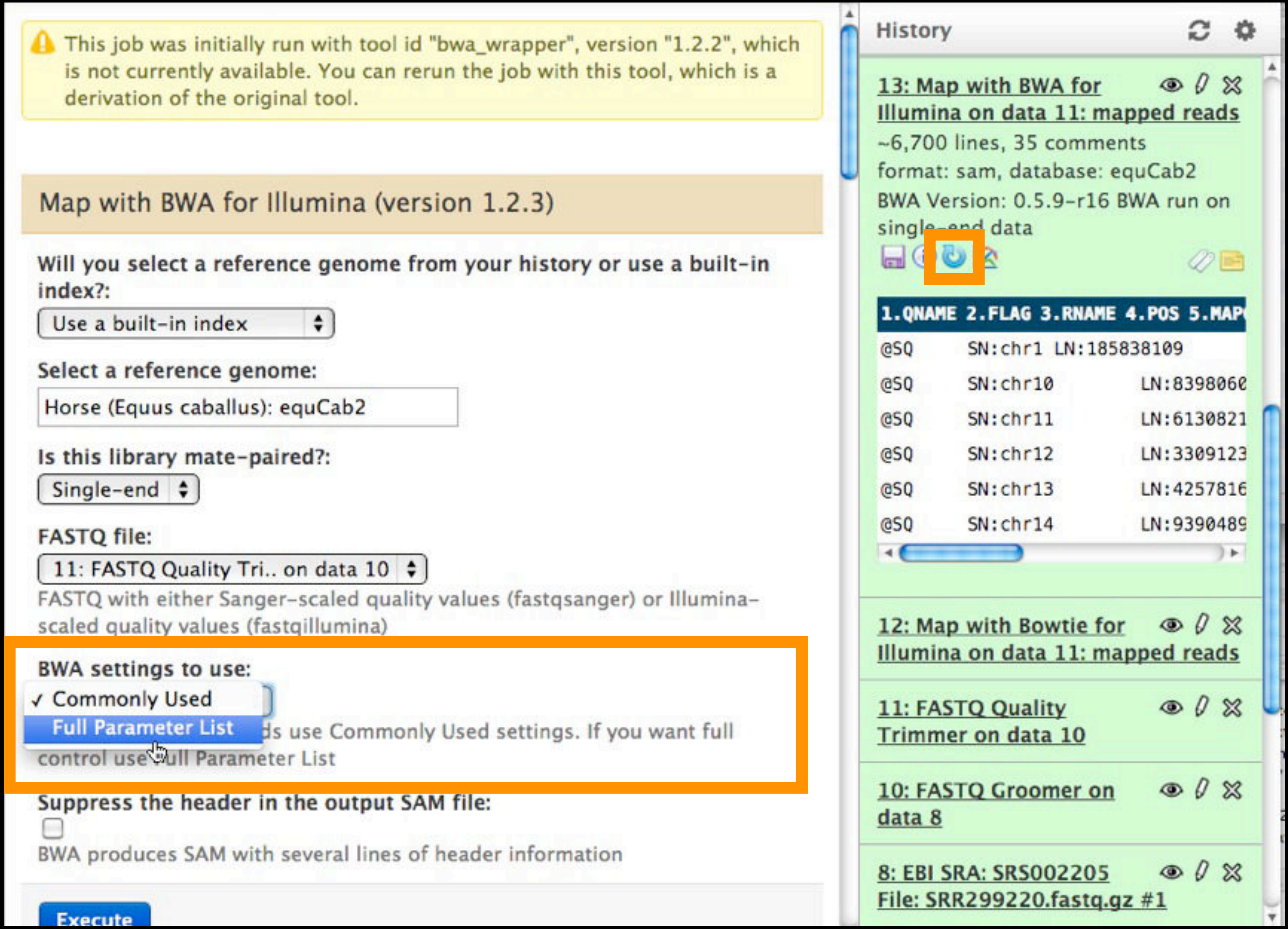
~6,700 lines, 35 comments

format: sam, database: equCab2

BWA Version: 0.5.9-r16 BWA run on single-end data

```
1.QNAME 2.FLAG 3.RNAME 4.POS 5.MAP
@SQ     SN:chr1  LN:185838109
```

4. Enable clients to reuse & modify your pipelines



This job was initially run with tool id "bwa_wrapper", version "1.2.2", which is not currently available. You can rerun the job with this tool, which is a derivation of the original tool.

Map with BWA for Illumina (version 1.2.3)

Will you select a reference genome from your history or use a built-in index?: Use a built-in index

Select a reference genome: Horse (Equus caballus): equCab2

Is this library mate-paired?: Single-end

FASTQ file: 11: FASTQ Quality Tri.. on data 10

FASTQ with either Sanger-scaled quality values (fastqsanger) or Illumina-scaled quality values (fastqillumina)

BWA settings to use:

- ✓ Commonly Used
- Full Parameter List

…ds use Commonly Used settings. If you want full control use Full Parameter List

Suppress the header in the output SAM file: ☐

BWA produces SAM with several lines of header information

History

13: Map with BWA for Illumina on data 11: mapped reads

~6,700 lines, 35 comments

format: sam, database: equCab2

BWA Version: 0.5.9-r16 BWA run on single-end data

```
1.QNAME 2.FLAG 3.RNAME 4.POS 5.MAP
@SQ     SN:chr1  LN:185838109
@SQ     SN:chr10        LN:8398060
@SQ     SN:chr11        LN:6130821
@SQ     SN:chr12        LN:3309123
@SQ     SN:chr13        LN:4257816
@SQ     SN:chr14        LN:9390489
```

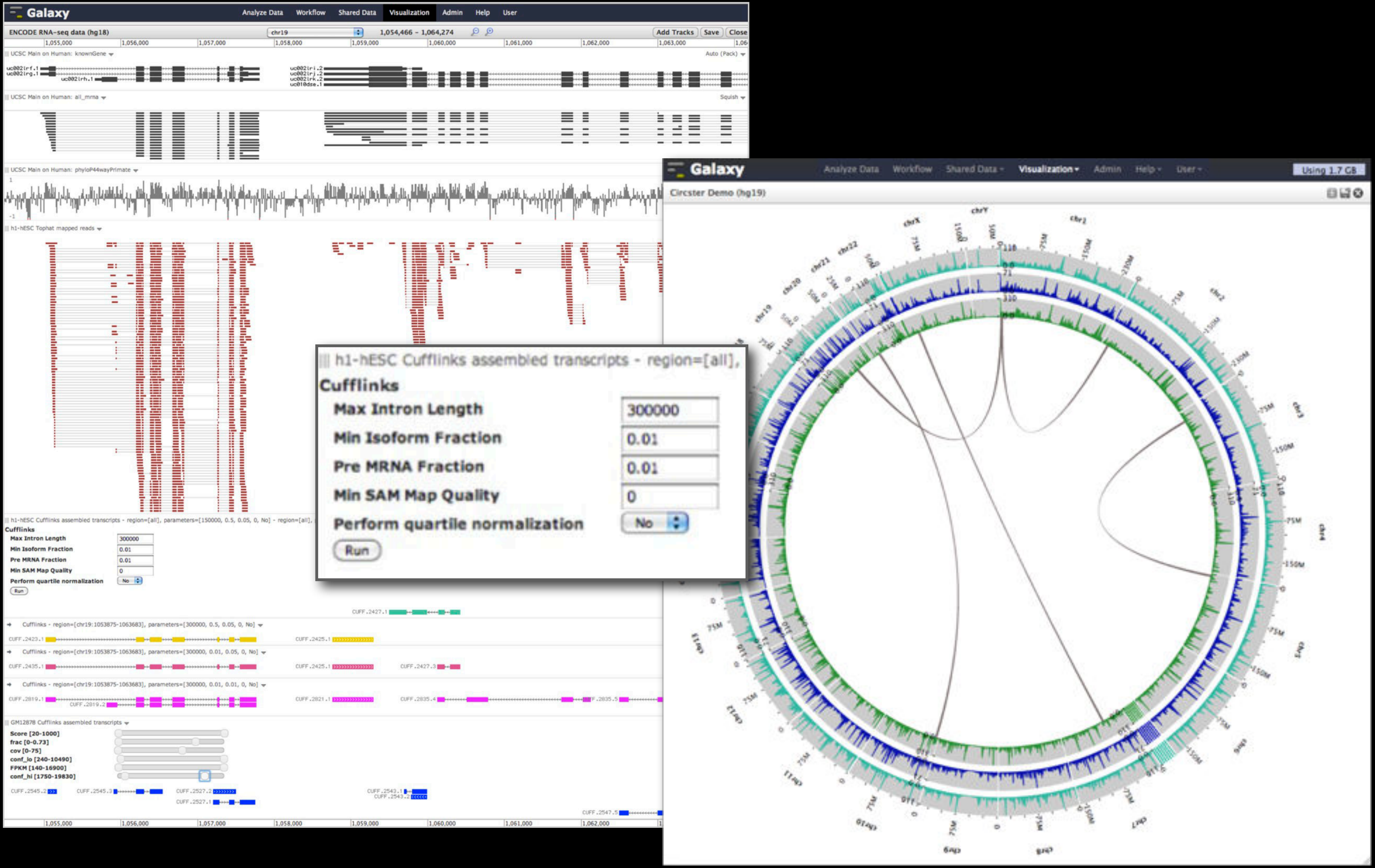
12: Map with Bowtie for Illumina on data 11: mapped reads

11: FASTQ Quality Trimmer on data 10

10: FASTQ Groomer on data 8

8: EBI SRA: SRS002205 File: SRR299220.fastq.gz #1

5. Enable clients to directly visualize results.



h1-hESC Cufflinks assembled transcripts - region=[all],

Cufflinks

Max Intron Length	300000
Min Isoform Fraction	0.01
Pre MRNA Fraction	0.01
Min SAM Map Quality	0
Perform quartile normalization	No

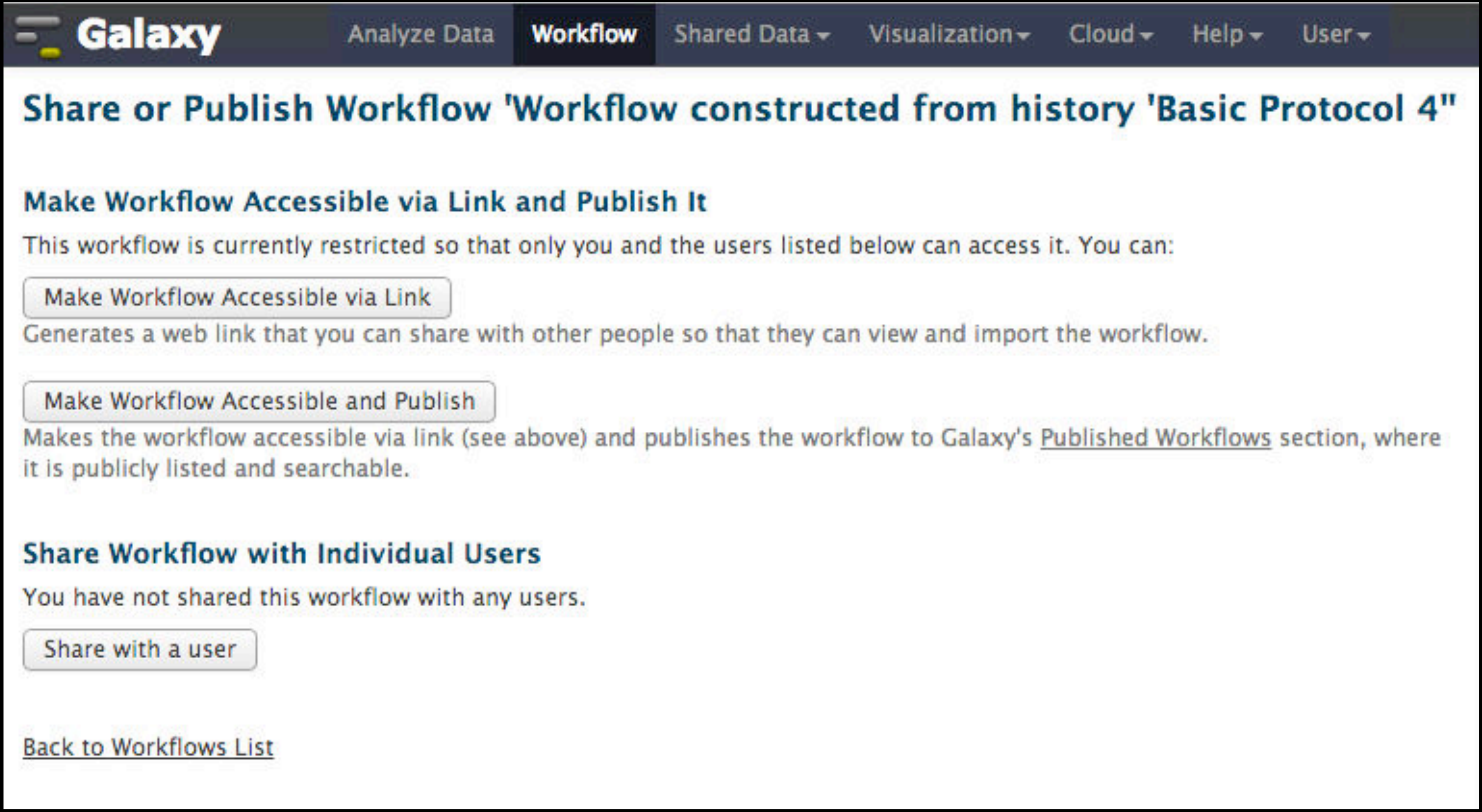
Run

6. Enable clients to develop their own analyses ***without***

- learning a programming language, command line / shell interfaces, Linux package management, ... or
- extensive ***hand-holding from core facility staff***

!

7. Enable sharing of methods and results.



Share or Publish Workflow 'Workflow constructed from history 'Basic Protocol 4''

Make Workflow Accessible via Link and Publish It

This workflow is currently restricted so that only you and the users listed below can access it. You can:

Generates a web link that you can share with other people so that they can view and import the workflow.

Makes the workflow accessible via link (see above) and publishes the workflow to Galaxy's Published Workflows section, where it is publicly listed and searchable.

Share Workflow with Individual Users

You have not shared this workflow with any users.

Back to Workflows List

Galaxy for Core Facilities

1. Automate your analysis pipelines
2. Enable clients to see details of how results were produced
3. Galaxy's provenance and reproducibility for free
4. Enable clients to explore, tweak, and reuse your pipelines
5. Enable clients to directly visualize results
6. Enable clients to develop their own analyses **without**
   - learning a programming language, command line interfaces and shell, Linux package management, ...
   - extensive hand-holding from core facility staff
7. Enable sharing of methods and results.

Registration & abstract submission are now open

http://galaxyproject.org/GCC2013


Enis Afgan


Dannon Baker


Dan Blankenberg


Dave Bouvier


Dave Clements


Nate Coraor


Carl Eberhard


Jeremy Goecks


Sam Guerler


Jen Jackson


Greg von Kuster


Ross Lazarus


Anton Nekrutenko


James Taylor

You

The Galaxy Team

http://wiki.galaxyproject.org/GalaxyTeam

Acknowledgements

Jim Vincent
Nalini Raghavachari
David Needleman

The Galaxy Team
especially
Dannon Baker

ABRF

NIH NSF Huck Institute
Penn State University   Emory University

Thanks

Galaxy is hiring post-docs and software engineers at both Emory and Penn State.

Please help.

http://wiki.galaxyproject.org/GalaxyIsHiring

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 Meting

http://wiki.galaxyproject.org

Mailing Lists

http://wiki.galaxyproject.org/MailingLists

Galaxy-Announce

- Project announcements, low volume, moderated
- Low volume (42 posts, 1600 members in 2012)

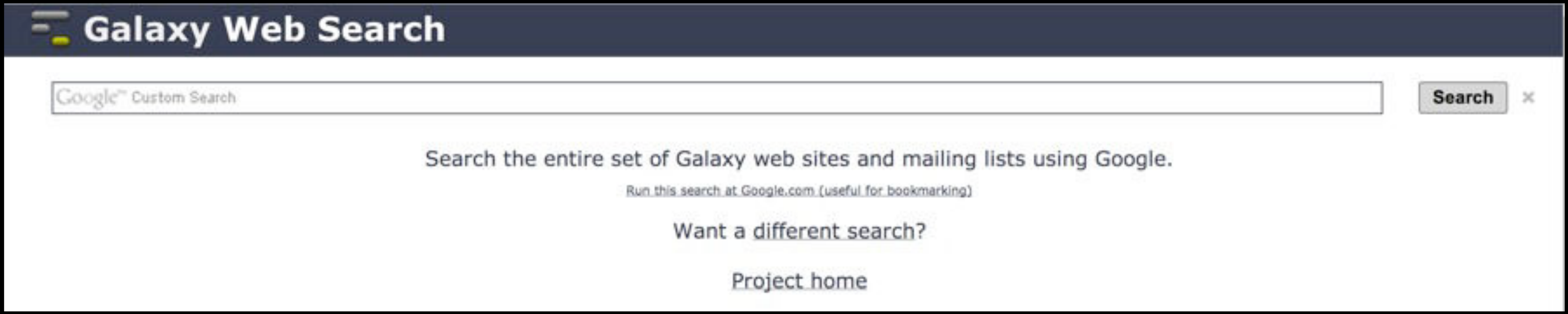
Galaxy-User

- Questions about using Galaxy and usegalaxy.org
- High volume (2900 posts, 2700 members in 2012)

Galaxy-Dev

- Questions about developing for and deploying Galaxy
- High volume (4500 posts,   850 members in 2012)

Unified Search: http://galaxyproject.org/search

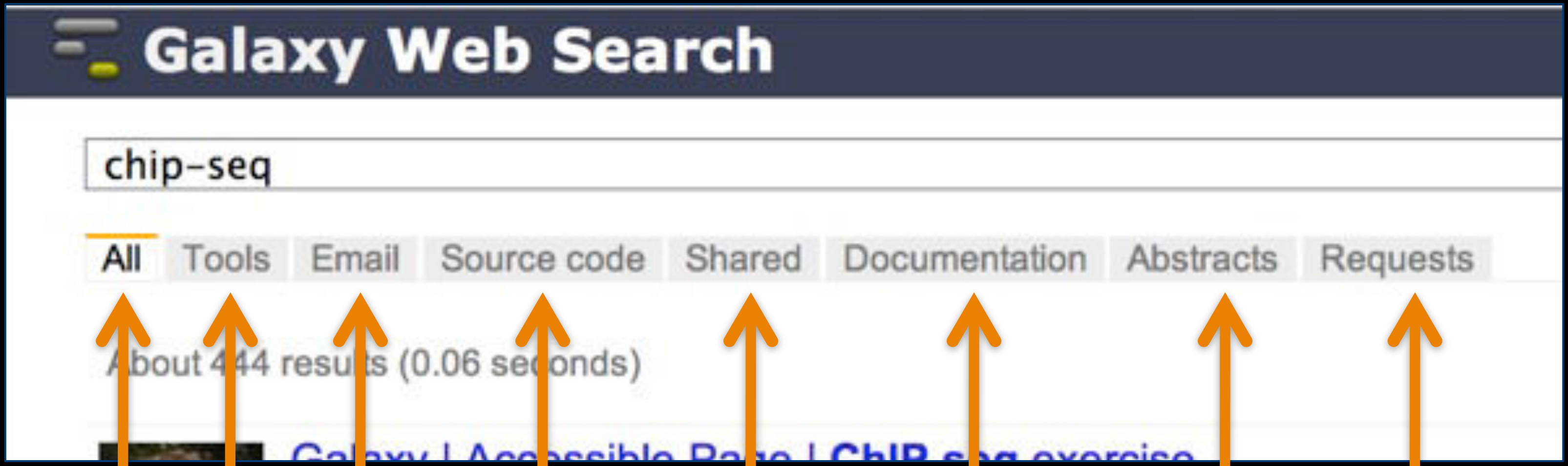


Search the entire set of Galaxy web sites and mailing lists using Google.

Run this search at Google.com (useful for bookmarking)

Want a different search?

Project home



About 444 results (0.06 seconds)

Find

Everything on …

Tools for …

Email about …

Source code for …

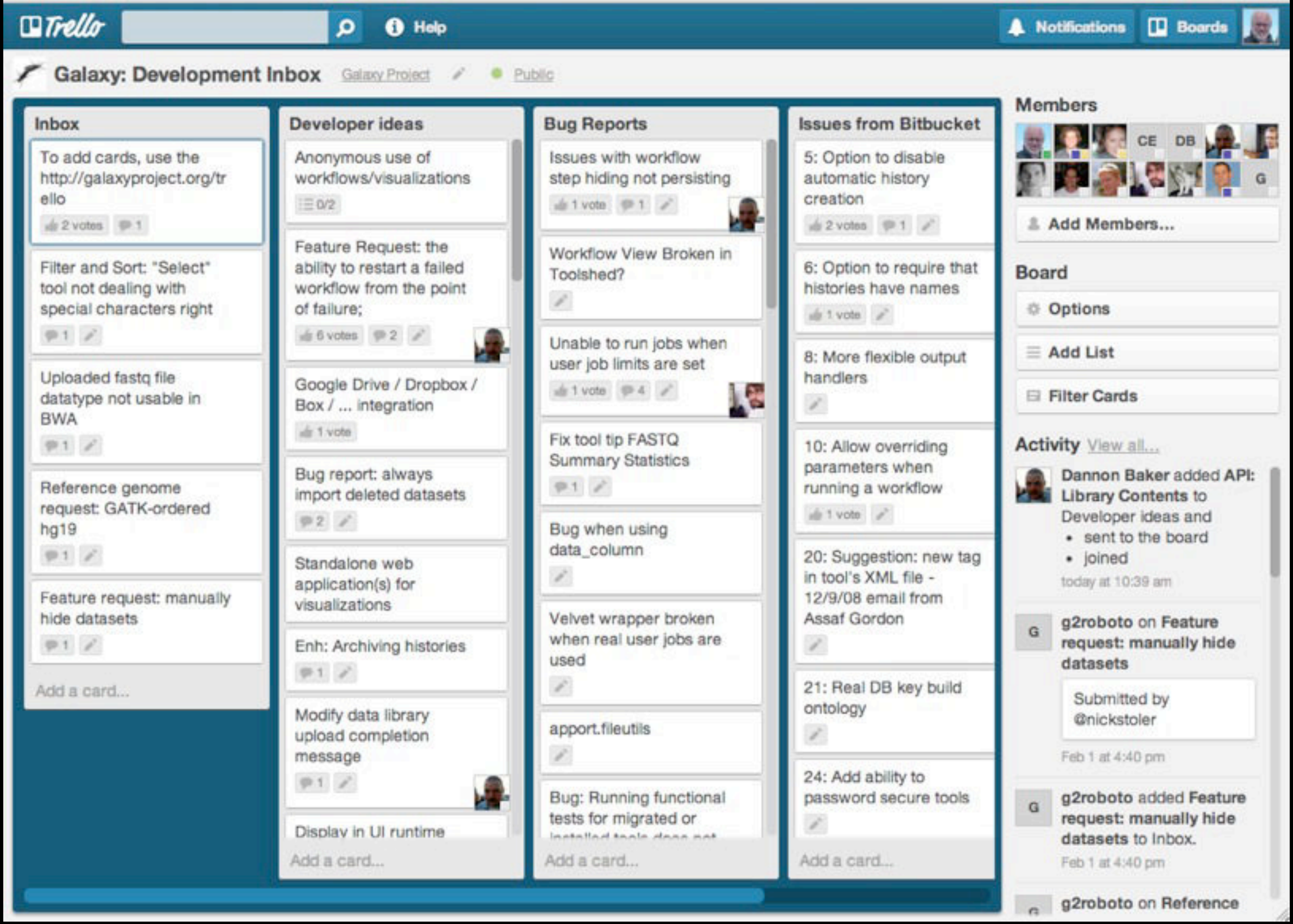
Published Histories, Pages, Workflows, about …

Documentation on …

Papers using Galaxy for …

Related feature requests

Community can create, vote and comment on issues



Galaxy: Development Inbox Galaxy Project · Public

Inbox
- To add cards, use the http://galaxyproject.org/trello — 2 votes, 1
- Filter and Sort: "Select" tool not dealing with special characters right — 1
- Uploaded fastq file datatype not usable in BWA — 1
- Reference genome request: GATK-ordered hg19 — 1
- Feature request: manually hide datasets — 1
- Add a card...

Developer ideas
- Anonymous use of workflows/visualizations — 0/2
- Feature Request: the ability to restart a failed workflow from the point of failure; — 6 votes, 2
- Google Drive / Dropbox / Box / ... integration — 1 vote
- Bug report: always import deleted datasets — 2
- Standalone web application(s) for visualizations
- Enh: Archiving histories — 1
- Modify data library upload completion message — 1
- Display in UI runtime
- Add a card...

Bug Reports
- Issues with workflow step hiding not persisting — 1 vote, 1
- Workflow View Broken in Toolshed?
- Unable to run jobs when user job limits are set — 1 vote, 4
- Fix tool tip FASTQ Summary Statistics — 1
- Bug when using data_column
- Velvet wrapper broken when real user jobs are used
- apport.fileutils
- Bug: Running functional tests for migrated or
- Add a card...

Issues from Bitbucket
- 5: Option to disable automatic history creation — 2 votes, 1
- 6: Option to require that histories have names — 1 vote
- 8: More flexible output handlers
- 10: Allow overriding parameters when running a workflow — 1 vote
- 20: Suggestion: new tag in tool's XML file - 12/9/08 email from Assaf Gordon
- 21: Real DB key build ontology
- 24: Add ability to password secure tools
- Add a card...

Members

CE DB G

Add Members...

Board

Options

Add List

Filter Cards

Activity View all...

Dannon Baker added **API: Library Contents** to Developer ideas and
- sent to the board
- joined

today at 10:39 am

g2roboto on **Feature request: manually hide datasets**

Submitted by @nickstoler

Feb 1 at 4:40 pm

g2roboto added **Feature request: manually hide datasets** to Inbox.

Feb 1 at 4:40 pm

g2roboto on **Reference**

http://bit.ly/gxyissues

Events

News

Galaxy Event Horizon

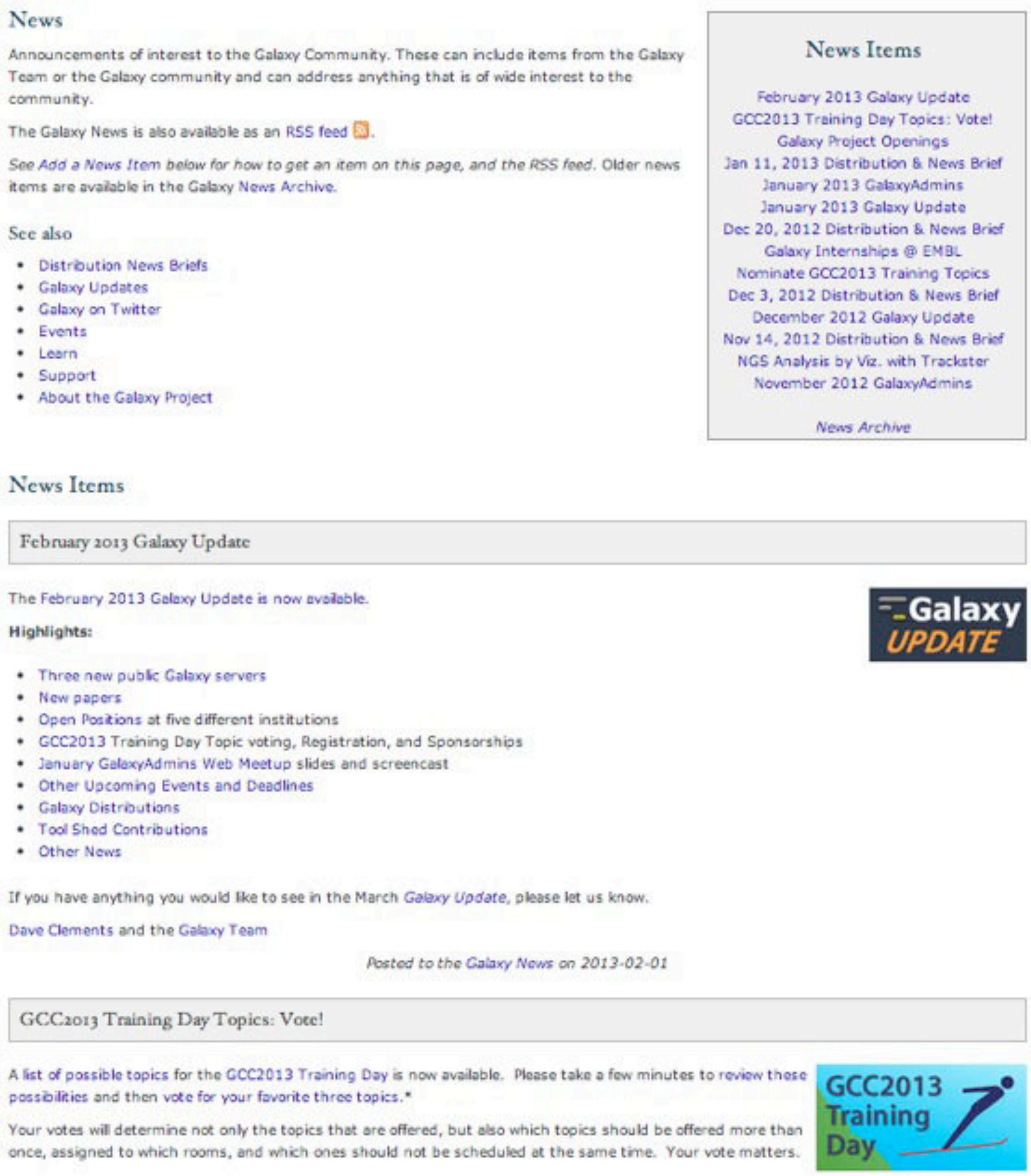
Events with Galaxy-related content are listed here.

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@glaxyproject.org .

Upcoming Events

Date	Topic/Event	Venue/Location
February 4	*Introduction to Galaxy Boot Camp*	UC Davis Bioinformatics Core Davis, California, United States
March 2-5	*Accessible, Transparent and Reproducible Analysis With Galaxy*, part of SW1: Application of NGS Platforms for Whole Transcriptome and Genome Analysis *Galaxy for Core Facilities*, part of "W6: Community Resource Solutions to Analyzing Large Genomic Data Sets"	**ABRF 2013** Palm Springs, California, United States
March 26-28	*RNA Technologies and Analysis Workshop*	DOE JGI User Meeting
April 5-6	*2013 GMOD Meeting*	Cambridge, United Kingdom, immediately prior to Biocuration 2013
April 7-10	*GO Galaxy Workshop*	Biocuration 2013, Cambridge, United Kingdom
April 9-11	Workshop: *Integrated Research Data Management for Next Gen Sequencing Analysis Using Galaxy and Globus Online Software-as-a-Service*	BioIT World, Boston, Massachusetts, United States
	Talk: *Integrated Research Data management and Analysis in NGS using Globus Online, Galaxy and Amazon Web Services*	
May 14-16	Tutorial: *Exploring and Enabling Biomedical Data Analysis with Galaxy*	Great Lakes Bioinformatics Conference (GLBIO) 2013, Pittsburgh, Pennsylvania, United States
May 21 **May 29**	*Initiation à l'utilisation de Galaxy* Les deux ateliers sont maintenant complets	Cycle "Bioinformatique par la pratique" 2013, INRA Jouy-en-Josas, France
May 22 **May 30**	Analyse de données issues de séquenceurs nouvelle génération sous Galaxy Les deux ateliers sont maintenant complets	
June 6-7	*Informatics on High Throughput Sequencing Data Workshop*	Toronto, Ontario, Cananda



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

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

News Items

February 2013 Galaxy Update
GCC2013 Training Day Topics: Vote!
Galaxy Project Openings
Jan 11, 2013 Distribution & News Brief
January 2013 GalaxyAdmins
January 2013 Galaxy Update
Dec 20, 2012 Distribution & News Brief
Galaxy Internships @ EMBL
Nominate GCC2013 Training Topics
Dec 3, 2012 Distribution & News Brief
December 2012 Galaxy Update
Nov 14, 2012 Distribution & News Brief
NGS Analysis by Viz. with Trackster
November 2012 GalaxyAdmins

News Archive

News Items

February 2013 Galaxy Update

The February 2013 Galaxy Update is now available.

Highlights:

- Three new public Galaxy servers
- New papers
- Open Positions at five different institutions
- GCC2013 Training Day Topic voting, Registration, and Sponsorships
- January GalaxyAdmins Web Meetup slides and screencast
- Other Upcoming Events and Deadlines
- Galaxy Distributions
- Tool Shed Contributions
- Other News

If you have anything you would like to see in the March *Galaxy Update*, please let us know.

Dave Clements and the Galaxy Team

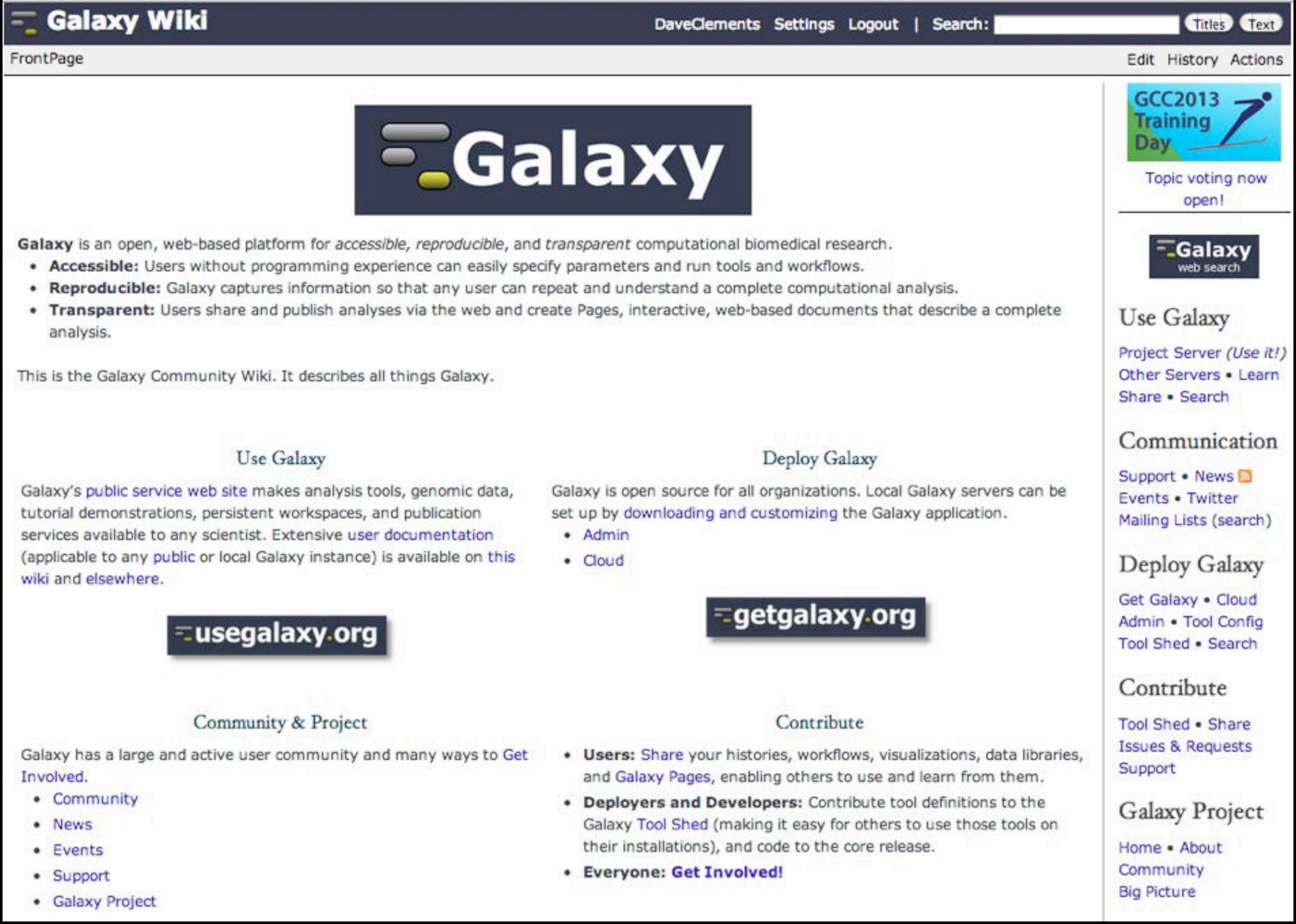
Posted to the Galaxy News on 2013-02-01

GCC2013 Training Day Topics: Vote!

A list of possible topics for the GCC2013 Training Day is now available. Please take a few minutes to review these possibilities and then vote for your favorite three topics.*

Your votes will determine not only the topics that are offered, but also which topics should be offered more than once, assigned to which rooms, and which ones should not be scheduled at the same time. Your vote matters.

http://wiki.galaxyproject.org



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.

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

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!**

Topic voting now open!

Use Galaxy

Project Server *(Use it!)*
Other Servers • 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

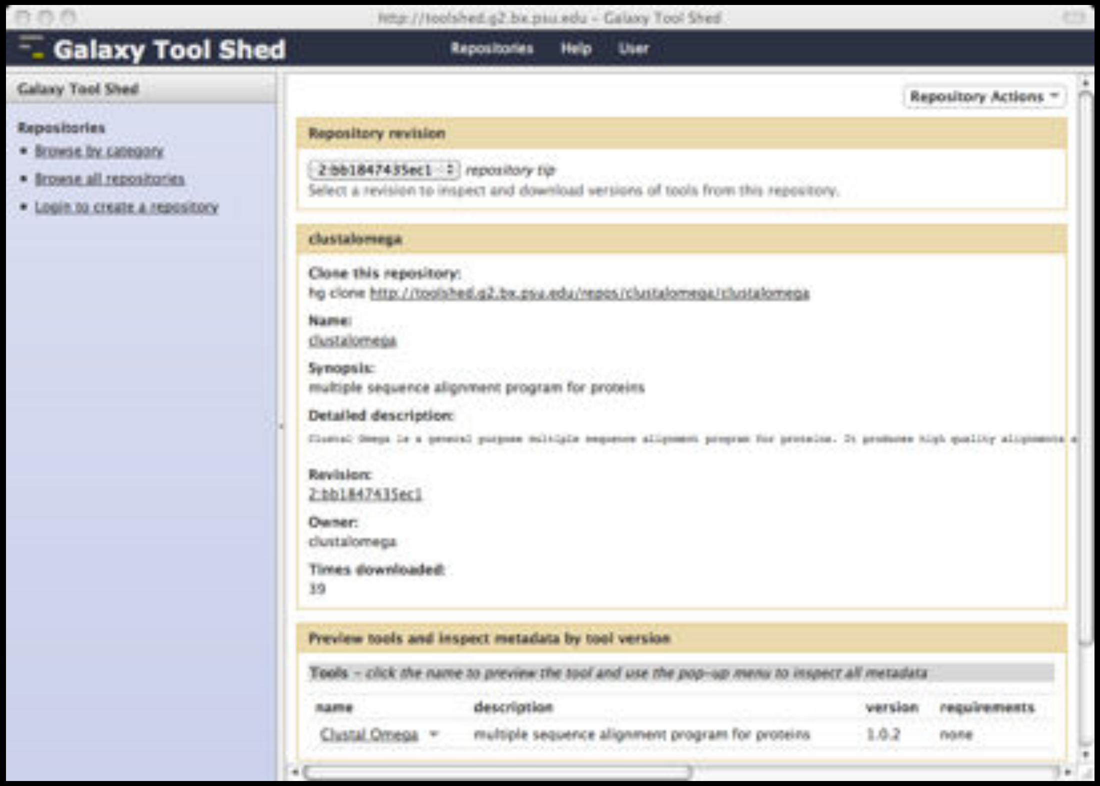
Home • About
Community
Big Picture

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



Repository revision

2:bb1847435ec1 repository tip

Select a revision to inspect and download versions of tools from this repository.

clustalomega

Clone this repository:
hg clone http://toolshed.g2.bx.psu.edu/repos/clustalomega/clustalomega

Name:
clustalomega

Synopsis:
multiple sequence alignment program for proteins

Detailed description:
`Clustal Omega is a general purpose multiple sequence alignment program for proteins. It produces high quality alignments`

Revision:
2:bb1847435ec1

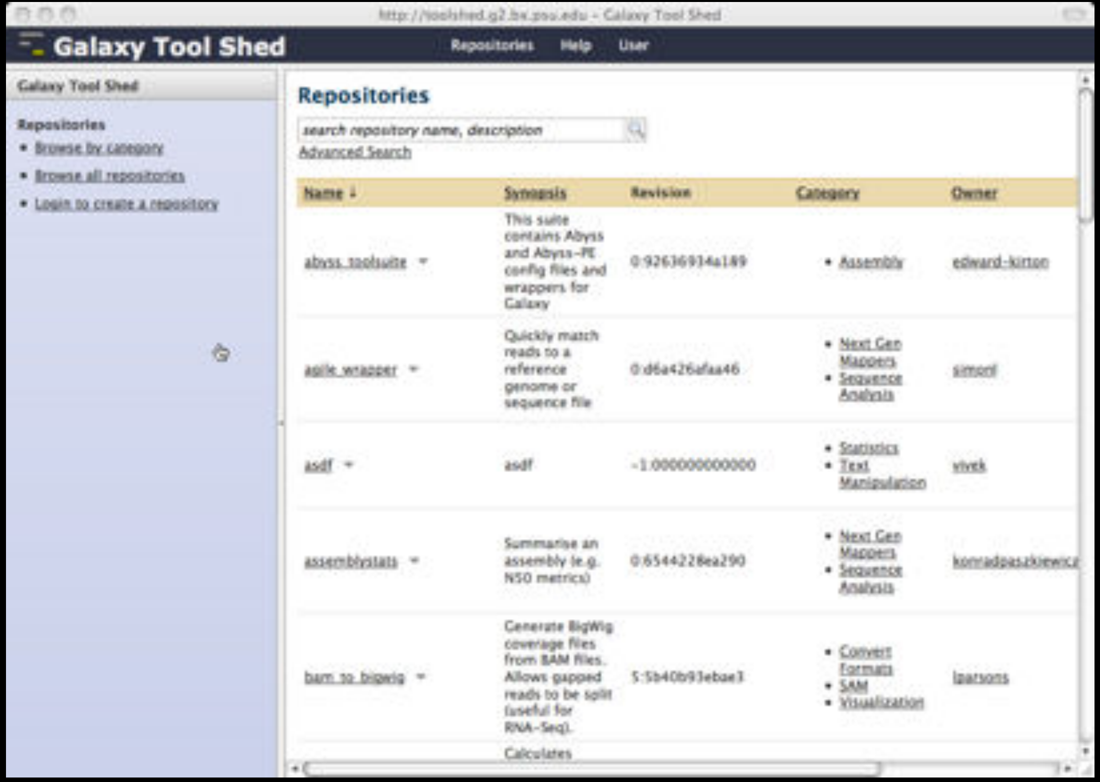
Owner:
clustalomega

Times downloaded:
39

Preview tools and inspect metadata by tool version

Tools - *click the name to preview the tool and use the pop-up menu to inspect all metadata*

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



Repositories

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	konradpaszkiewic
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
	Calculates			

Encourage Public Galaxy Instances

http://wiki.galaxyproject.org/PublicGalaxyServers

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 is hiring post-docs and software engineers at both Emory and Penn State.

Please help.

http://wiki.galaxyproject.org/GalaxyIsHiring