

Galaxy for NGS Analysis

What is Galaxy, and how can I use it in my lab?

Matt Shirley
Graduate Student
Johns Hopkins Medicine
Pevsner Lab

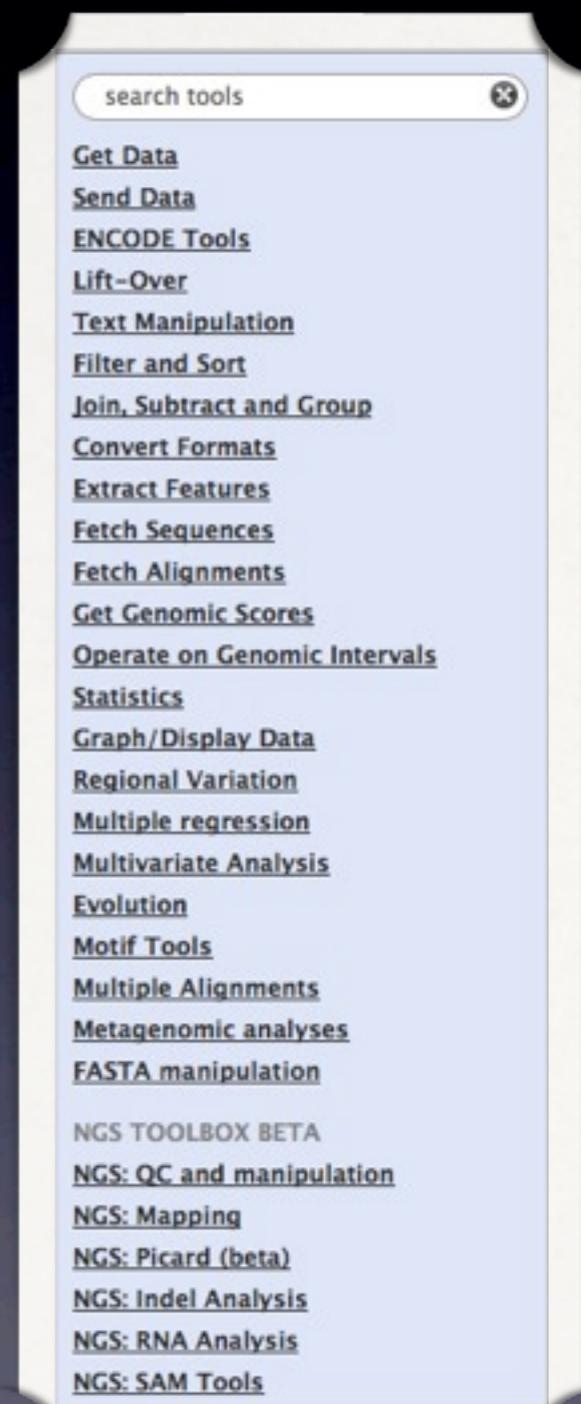
Slides available at <http://mattshirley.com/presentations>

Contents

- What is Galaxy?
- Interface elements
- Retrieving data
- Creating and running workflows
- A FASTQ quality statistics workflow
- Galaxy on Amazon Web Services (AWS)

What is Galaxy?

- Galaxy is a collection of bioinformatics tools for:
 - data conversion and manipulation
 - statistical analysis
 - next generation sequencing analysis
 - much, much more!....



What is Galaxy?

Citation

If you use this tool, please cite [Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A; Galaxy Team. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010 Jul 15;26\(14\):1783-5.](#)

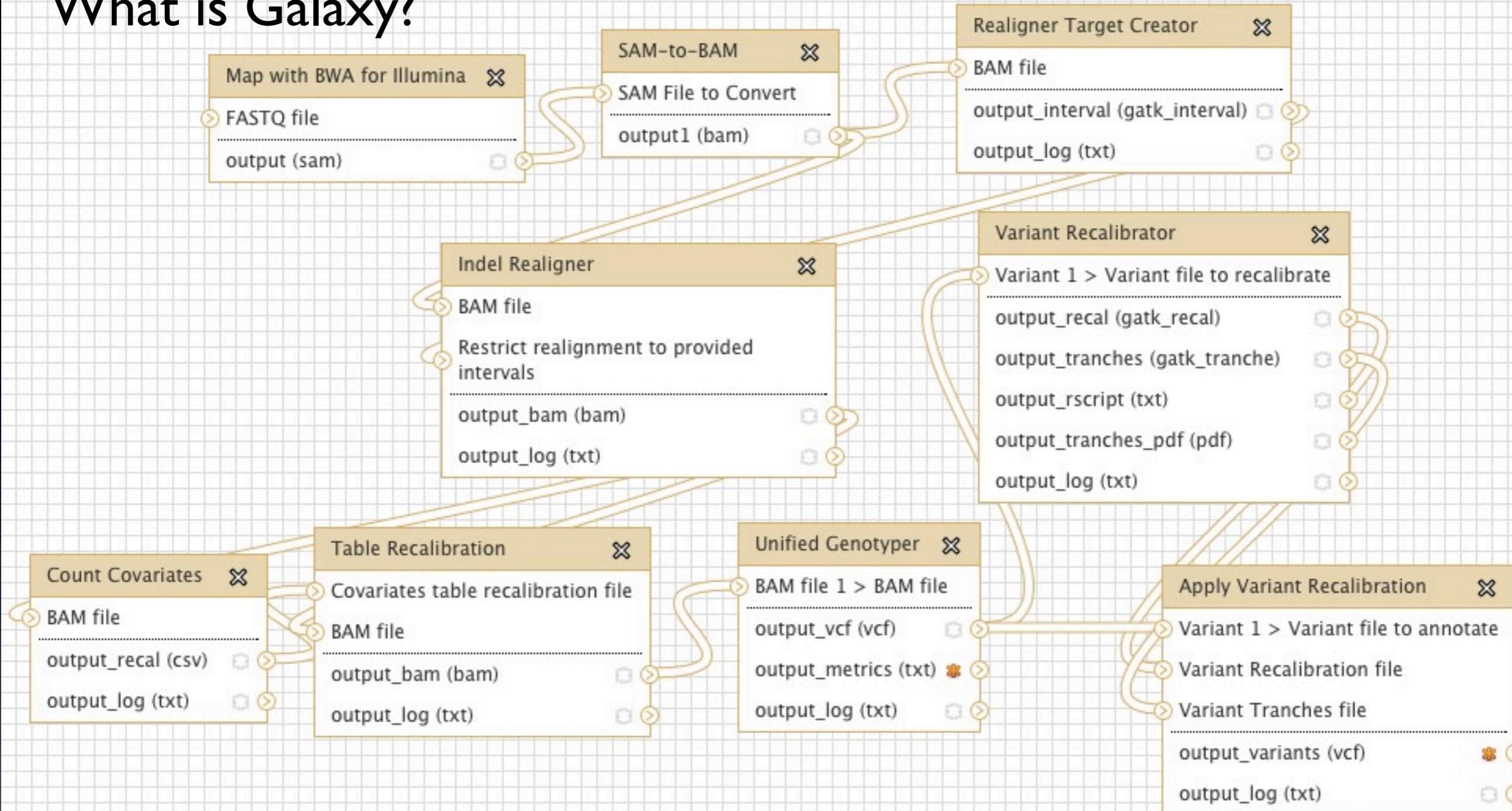
- Based on peer-reviewed and open-source implementations of each tool
- Galaxy provides integration with useful tools, targeted toward “bench” scientists
- Unified and consistent interface for easy exploration

What is Galaxy?

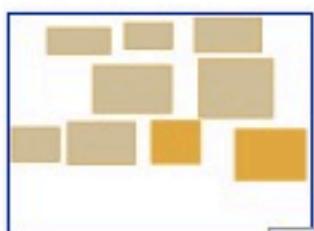
- Data library: management and sharing for collaborative analysis
- Sequencer interface for direct intake of raw data

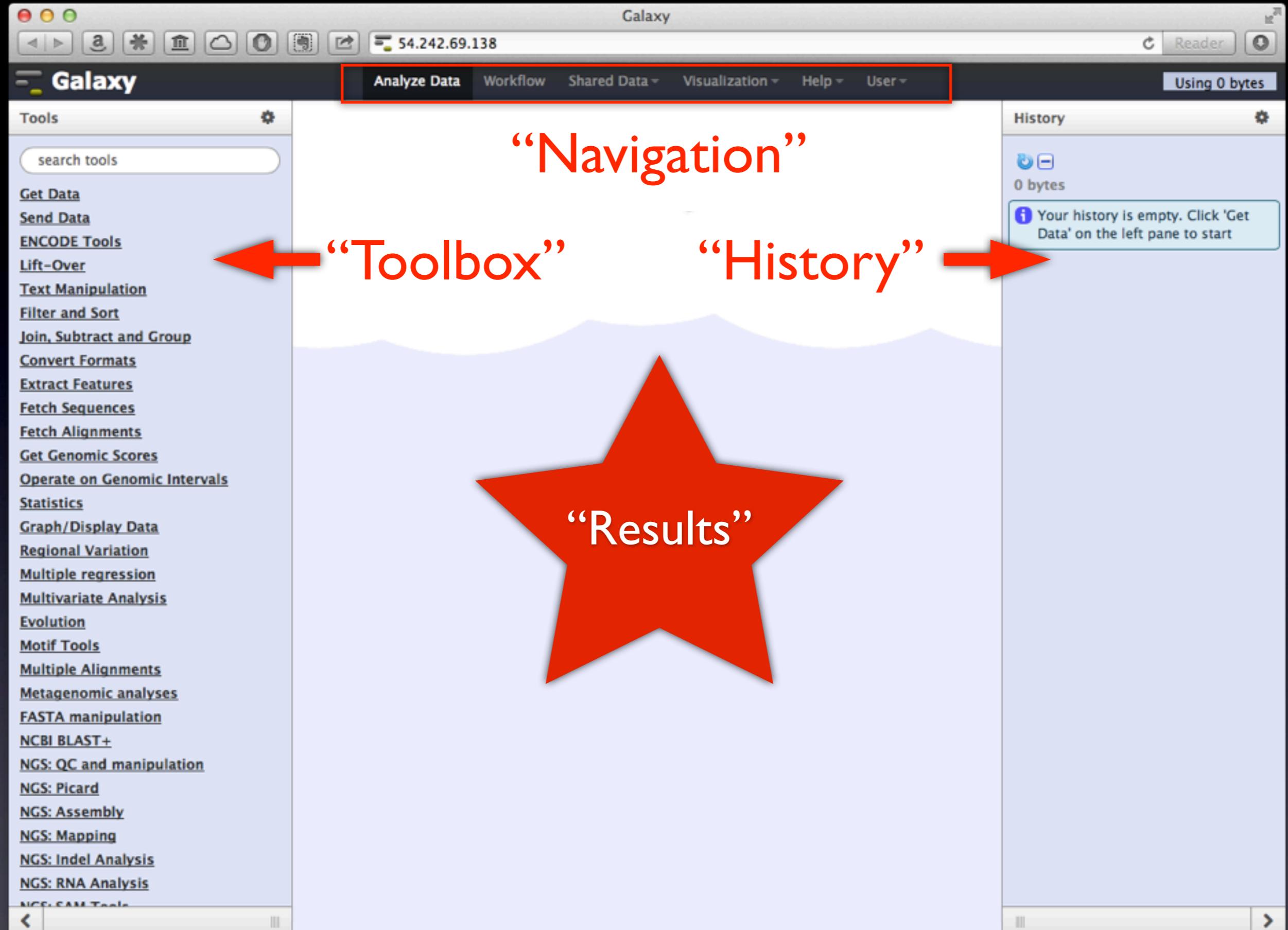
Data Libraries	
<input type="text" value="search dataset name, info, message, dbke"/> <input type="button" value=""/>	
Advanced Search	
Data library name ↓	Data library description
1000 Genomes	Data from the 1000 Genomes Project FTP site
AC-exome	
Bushman	Data for Nature Letter "Complete Khoisan and
ChIP-Seq Mouse Example	Data used in examples that demonstrate analy
Chobi	
CloudMap	Contains userguide, reference files, and config
Codon Usage Frequencies	
Coleman	IonPGM
Denisovan sequences	Files from 'A high-coverage genome sequence
Erythroid Epigenetic Landscape	Dynamics of the epigenetic landscape during e
Evolutionary Trajectories in a Phage	Experimental evolution (Illumina)
GATK	
GCAT	Consortium
Genome Diversity	Nucleotide polymorphisms for several threatene
guru_1000GP	
He-2010	
Heteroplasmy	Data for Genome Biology 2011 manuscript
iGenomes	Selected files from Illumina iGenomes collectio

What is Galaxy?



Workflows that enable
reproducible research





The “toolbox”



- Contains links for :
 - retrieving (“get”) data
 - manipulating data (lift-over, filter, sort, set operations, format conversions)
 - data analysis (statistics, sequence alignment, variant calling and annotation)

“Get” data

- In addition to uploading files from your computer, you may:
- Choose a file in the “shared data” library
- Import from UCSC, EBI SRA, BioMart, CBI Rice Map, modENCODE, Ratmine, Flymine, YeastMine, WormBase, EuPath, Microbial Genome Project, EncodeDB, EpiGRAPH, HbVar, GenomeSpace

Galaxy

54.242.69.138

Analyze Data Workflow Shared Data Visualization Help User Using 0 bytes

Tools

search tools

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- BX main browser
- Get Microbial Data
- BioMart Central server
- BioMart Test server
- CBI Rice Mart rice mart
- GrameneMart Central server
- modENCODE fly server
- Flymine server
- Flymine test server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- metabolicMine server
- modENCODE worm server
- WormBase server
- Wormbase test server
- EuPathDB server
- EncodeDB at NHGRI
- EpiGRAPH server
- EpiGRAPH test server

Upload File (version 1.1.3)

File Format:
Auto-detect
Which format? See help below

File:
Choose File no file selected
TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

URL/Text:
`https://s3.amazonaws.com/Sijung_Yun_NGS/1000_tags.fastq.gz`

Here you may specify a list of URLs (one per line) or paste the contents of a file.

Files uploaded via FTP:

File	Size	Date
------	------	------

Please [create](#) or [log in](#) to a Galaxy account to view files uploaded via FTP.

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at localhost using your Galaxy credentials (email address and password).

Convert spaces to tabs:
 Yes
Use this option if you are entering intervals by hand.

Genome:
Click to Search or Select

Execute

Auto-detect
The system will attempt to detect Axt, Fasta, Fastqsolexa, Gff, Gff3, Html, Lav, Maf, Tabular, Wiggle, Bed and Interval (Bed with headers) formats. If your file is not detected properly as one of the known formats, it may still be able to read it, but it is recommended to use the appropriate file type.

History

0 bytes

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

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 2.1 Mb

Tools

- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NCBI BLAST+
- NGS: QC and manipulation
 - FASTQC: FASTQ/SAM/BAM
 - Fastqc: Fastqc QC using FastQC from Babraham
 - ILLUMINA FASTQ
 - FASTQ Groomer convert between various FASTQ quality formats
 - FASTQ splitter on joined paired end reads
 - FASTQ joiner on paired end reads
 - FASTQ Summary Statistics by column
 - ROCHE-454 DATA
 - Build base quality distribution
 - Select high quality segments
 - Combine FASTA and QUAL into FASTQ
 - AB-SOLID DATA
 - Convert SOLID output to fastq
 - Compute quality statistics for SOLID data
 - Draw quality score boxplot for SOLID data

History

1: https://s3.amazonaws.com/Sijung_Yun_NGS/1000_tags.fastq 2.1 Mb

Edit Attributes

Name: https://s3.amazonaws.com/Sijung_Yun_NGS/1000_tags.fastq

Info: uploaded fastq file

Database/Build: Click to Search or Select

Save Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

Convert to new format

Convert FASTQ files to seek locations

This will create a new dataset with the contents of this dataset converted to a new format.

Convert

Change data type

New Type: fastqsanger

This will change the datatype of the existing dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

Save

54.242.69.138

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 2.1 Mb

Tools

- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NCBI BLAST+
- NGS: QC and manipulation
 - FASTQC: FASTQ/SAM/BAM
 - Fastqc: Fastqc QC using FastQC from Babraham
 - ILLUMINA FASTQ
 - FASTQ Groomer convert between various FASTQ quality formats
 - FASTQ splitter on joined paired end reads
 - FASTQ joiner on paired end reads
 - FASTQ Summary Statistics by column
 - ROCHE-454 DATA
 - Build base quality distribution
 - Select high quality segments
 - Combine FASTA and QUAL into FASTQ
 - AB-SOLID DATA
 - Convert SOLID output to fastq
 - Compute quality statistics for SOLID data
 - Draw quality score boxplot for SOLID data
- GENERIC FASTQ MANIPULATION

FASTQ Summary Statistics (version 1.0.0)

FASTQ File:
1: https://s3.amazonaws.com/Sijung_Yun_NGS/1000_tags.fastq

Execute

This tool creates summary statistics on a FASTQ file.

TIP: This statistics report can be used as input for the Boxplot and Nucleotides Distribution tools.

The output file will contain the following fields:

column = column number (1 to 36 for a 36-cycles read Solexa file)
count = number of bases found in this column.
min = Lowest quality score value found in this column.
max = Highest quality score value found in this column.
sum = Sum of quality score values for this column.
mean = Mean quality score value for this column.
Q1 = 1st quartile quality score.
med = Median quality score.
Q3 = 3rd quartile quality score.
IQR = Inter-Quartile range (Q3-Q1).
lW = 'Left-Whisker' value (for boxplotting).
rW = 'Right-Whisker' value (for boxplotting).
outliers = Scores falling beyond the left and right whiskers (comma separated list).
A_Count = Count of 'A' nucleotides found in this column.
C_Count = Count of 'C' nucleotides found in this column.
G_Count = Count of 'G' nucleotides found in this column.
T_Count = Count of 'T' nucleotides found in this column.
N_Count = Count of 'N' nucleotides found in this column.
Other_Nucs = Comma separated list of other nucleotides found in this column.
Other_Count = Comma separated count of other nucleotides found in this column.

For example:

#column	count	min	max	sum	mean	Q1	med	Q3	IQR	lW	rW	outliers	A_Count	C_Count
1	14336356	2	33	450600675	31.4306281875	32.0	33.0	33.0	33.0	1.0	31	33		
2	14336356	2	34	441135033	30.7703737965	30.0	33.0	33.0	33.0	3.0	26	34		
3	14336356	2	34	433659182	30.2489127642	29.0	32.0	33.0	33.0	4.0	23	34		
4	14336356	2	34	433635331	30.2472490917	29.0	32.0	33.0	33.0	4.0	23	34		
5	14336356	2	34	432498583	30.167957813	29.0	32.0	33.0	33.0	4.0	23	34		

History

Unnamed history 2.1 Mb

1: https://s3.amazonaws.com/Sijung_Yun_NGS/1000_tags.fastq

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Using 2.1 Mb

The following job has been successfully added to the queue:

2: FASTQ Summary Statistics on data 1

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

Unnamed history 2.1 Mb

2: FASTQ Summary Statistics on data 1

1: https://s3.amazonaws.com/Sijung_Yun_NGS/1000_tags.fastq

This screenshot shows the Galaxy web interface. The left sidebar contains a 'Tools' section with various bioinformatics tools categorized under 'Multivariate Analysis', 'Evolution', 'Motif Tools', 'Multiple Alignments', 'Metagenomic analyses', 'FASTA manipulation', 'NCBI BLAST+', and 'NGS: QC and manipulation'. Under 'NGS: QC and manipulation', there are sections for 'FASTQC: FASTQ/SAM/BAM', 'ILLUMINA FASTQ', 'FASTQ Groomer', 'FASTQ splitter', 'FASTQ joiner', 'FASTQ Summary Statistics by column', 'ROCHE-454 DATA', 'Build base quality distribution', 'Select high quality segments', 'Combine FASTA and QUAL into FASTQ', 'AB-SOLID DATA', 'Convert SOLID output to fastq', 'Compute quality statistics for SOLID data', 'Draw quality score boxplot for SOLID data', and 'GENERIC FASTQ MANIPULATION'. The main area displays a success message about a job being added to the queue, followed by a detailed description of the job. On the right, the 'History' panel shows a single entry for a FASTQ summary statistics job, with a red box highlighting it. The URL for the generated file is also shown.

Galaxy

54.242.69.138

Analyze Data Workflow Shared Data Visualization Help User Using 2.1 Mb

Tools

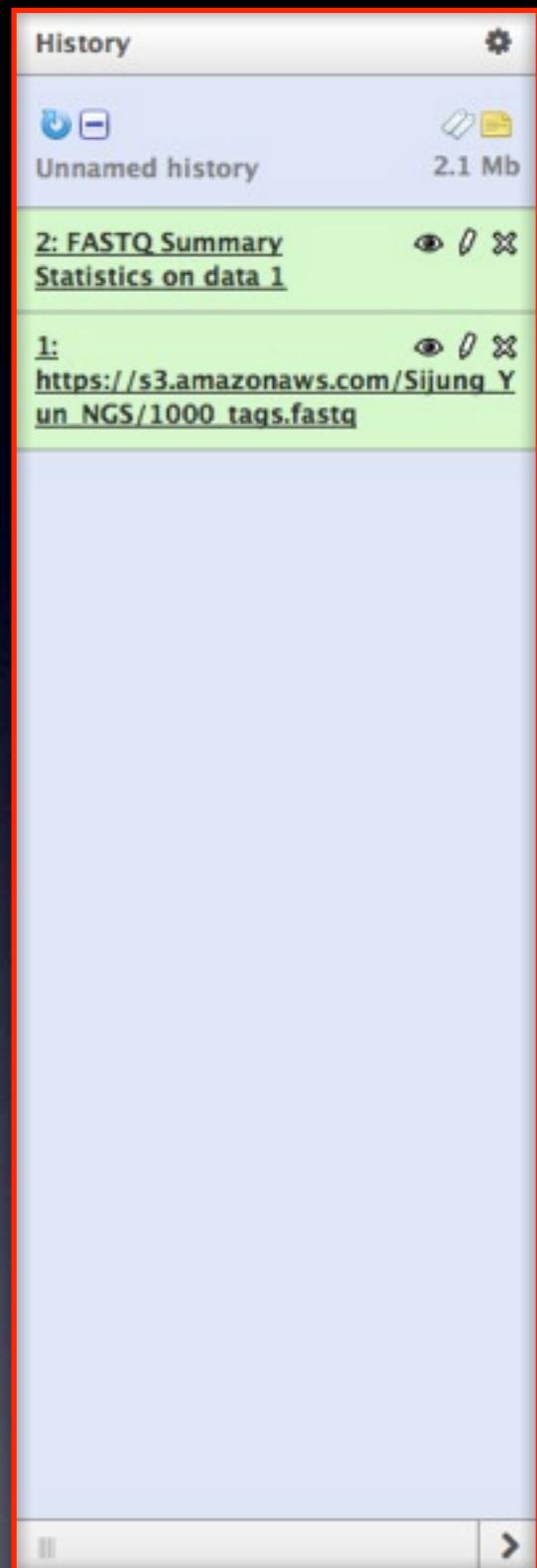
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NCBI BLAST+
- NGS: QC and manipulation
 - FASTQC: FASTQ/SAM/BAM
 - Fastqc: Fastqc QC using FastQC from Babraham
 - ILLUMINA FASTQ
 - FASTQ Groomer convert between various FASTQ quality formats
 - FASTQ splitter on joined paired end reads
 - FASTQ joiner on paired end reads
 - FASTQ Summary Statistics by column
 - ROCHE-454 DATA
 - Build base quality distribution
 - Select high quality segments
 - Combine FASTA and QUAL into FASTQ
 - AB-SOLID DATA
 - Convert SOLID output to fastq
 - Compute quality statistics for SOLID data
 - Draw quality score boxplot for SOLID data
 - GENERIC FASTQ MANIPULATION

#column count min max sum mean Q1 med Q3 IQR IW rW History

#column	count	min	max	sum	mean	Q1	med	Q3	IQR	IW	rW	History
1	10000	2	40	358994	35.8994	35.0	38.0	39.0	4.0	29	40	Unnamed history 2.1 Mb
2	10000	2	40	356075	35.6075	35.0	38.0	39.0	4.0	29	40	
3	10000	2	40	355450	35.545	35.0	38.0	39.0	4.0	29	40	
4	10000	2	40	354268	35.4268	35.0	38.0	39.0	4.0	29	40	
5	10000	2	40	355691	35.5691	35.0	38.0	39.0	4.0	29	40	
6	10000	2	40	357066	35.7066	35.0	38.0	39.0	4.0	29	40	
7	10000	2	40	357335	35.7335	35.0	38.0	39.0	4.0	29	40	
8	10000	2	40	356799	35.6799	35.0	38.0	39.0	4.0	29	40	
9	10000	2	40	355891	35.5891	35.0	38.0	39.0	4.0	29	40	
10	10000	2	40	356335	35.6335	35.0	38.0	39.0	4.0	29	40	
11	10000	2	40	357109	35.7109	35.0	38.0	39.0	4.0	29	40	
12	10000	2	40	357572	35.7572	35.0	38.0	39.0	4.0	29	40	
13	10000	2	40	358956	35.8956	35.0	38.0	39.0	4.0	29	40	
14	10000	2	40	360000	36.0	35.0	38.0	39.0	4.0	29	40	
15	10000	2	40	356901	35.6901	35.0	38.0	39.0	4.0	29	40	
16	10000	2	40	350223	35.0223	34.0	38.0	39.0	5.0	27	40	
17	10000	2	40	351943	35.1943	34.0	38.0	39.0	5.0	27	40	
18	10000	2	40	351757	35.1757	34.0	38.0	39.0	5.0	27	40	
19	10000	2	40	351216	35.1216	34.0	38.0	39.0	5.0	27	40	
20	10000	2	40	351567	35.1567	34.0	38.0	39.0	5.0	27	40	
21	10000	2	40	349414	34.9414	34.0	38.0	39.0	5.0	27	40	
22	10000	2	40	350206	35.0206	34.0	38.0	39.0	5.0	27	40	
23	10000	2	40	350242	35.0242	34.0	38.0	39.0	5.0	27	40	
24	10000	2	40	351490	35.149	34.0	38.0	39.0	5.0	27	40	
25	10000	2	40	349360	34.936	34.0	37.0	39.0	5.0	27	40	
26	10000	2	40	353087	35.3087	35.0	38.0	39.0	4.0	29	40	
27	10000	2	40	351412	35.1412	35.0	38.0	39.0	4.0	29	40	
28	10000	2	40	352640	35.264	35.0	38.0	39.0	4.0	29	40	
29	10000	2	40	352542	35.2542	35.0	38.0	39.0	4.0	29	40	
30	10000	2	40	349193	34.9193	34.0	38.0	39.0	5.0	27	40	
31	10000	2	40	354332	35.4332	35.0	38.0	39.0	4.0	29	40	
32	10000	2	40	347471	34.7471	34.0	38.0	39.0	5.0	27	40	
33	10000	2	40	354691	35.4691	35.0	38.0	39.0	4.0	29	40	
34	10000	2	40	352285	35.2285	35.0	38.0	39.0	4.0	29	40	
35	10000	2	40	348150	34.815	34.0	38.0	39.0	5.0	27	40	
36	10000	2	40	350222	35.0222	35.0	38.0	39.0	4.0	29	40	

The “history”

- Displays a list of your analysis steps
- Allows interaction with analysis results
- Each item in the history is a “data-set”
- Multiple concurrent histories allowed
- Maintains the order of analysis steps, allowing extraction of workflows on-the-fly



Extracting workflows from histories

The screenshot shows a software interface with a context menu open over a dataset history. The menu includes options like 'Saved Histories', 'Clone', 'Copy Datasets', and 'Extract Workflow'. A red arrow points from the 'Extract Workflow' option to a separate window displaying a workflow diagram.

The workflow diagram consists of two main components connected by a directed edge:

- An "Input dataset" node with an "output" port.
- A "FASTQ Summary Statistics" node with a "FASTQ File" output port.

The "FASTQ File" port is connected to the "output" port of the "Input dataset" node. The "FASTQ File" port also has a "output_file (tabular)" connection.

Histories and workflows result in reproducible research

NGS analysis in Galaxy

- QC and manipulation: filter, trim, mask, and convert fastq files
- Picard: a Java implementation of many samtools functions
- Mapping: align to reference genome with BWA, Bowtie, Bowtie2, BFAST, PerM, Mosaik, Lastz
- RNA: Tophat, Cufflinks (gapped alignment and transcript assembly)
- GATK: advanced analysis tools from BROAD
- Peak Calling: ChIP-Seq analysis tools

[NGS: QC and manipulation](#)
[NGS: Picard \(beta\)](#)
[NGS: Mapping](#)
[NGS: Indel Analysis](#)
[NGS: RNA Analysis](#)
[NGS: SAM Tools](#)
[NGS: GATK Tools \(beta\)](#)
[NGS: Peak Calling](#)

Strengths and Weaknesses

- Strengths:
 - Each tool has similar user interface elements, leading to a much lower learning curve
 - Histories and workflows allow reproducibility
 - Active development, tools are peer-reviewed
 - Cluster and cloud compute-compatible
 - Extensible tool set via Python scripting
- Weaknesses:
 - Inefficient resource management (for now)

Local vs. Public

- Public Galaxy server is accessible at
<http://usegalaxy.org>
- Learn about installing local instances at
<http://getgalaxy.org>
- NGS analysis involves *large* data, and long compute times.
- For NGS analysis, a local (or cloud) installation of Galaxy is recommended.

Questions?

Slides available at <http://mattshirley.com/presentations>

Galaxy on AWS (“the cloud”)



<http://xkcd.com/1117/>

New! Two options for cluster initialization

1. Use the new cloud initialization tool from the main public instance.
2. Continue through the steps detailed in the remaining slides to manually configure a cluster through Amazon Web Services management console.

Launch a Galaxy Cloud Instance

To launch a Galaxy Cloud Cluster, enter your AWS Secret Key ID, and Secret Key. Galaxy will use these to present appropriate options for launching your cluster. Note that using this form to launch computational resources in the Amazon Cloud will result in costs to the account indicated above. See [Amazon's pricing](#) for more information. options for launching your cluster.

Key ID
 This is the text string that uniquely identifies your account, found in the [Security Credentials section of the AWS Console](#).

Secret Key
 This is your AWS Secret Key, also found in the [Security Credentials section of the AWS Console](#).

Galaxy on AWS (“the cloud”)

Steps adapted from <http://wiki.g2.bx.psu.edu/CloudMan>

I. Log in to AWS EC2 management console

<http://console.aws.amazon.com/ec2>

- Access your *Security Credentials* page
- Save your *Access Key ID* and *Secret Access Key*



Your Access Keys			
Created	Access Key ID	Secret Access Key	Status
October 3, 2012	AKIAJRBU4D36GXYROQWQ	Show	Active (Make Inactive)

Galaxy on AWS (“the cloud”)

2. Create a *Security Group* called “galaxy”,
description “galaxy AMI”

- Choose Key Pairs

Create Security Group

Name:

Description:

VPC:



Create Key Pair

Key Pair Name:

Galaxy on AWS (“the cloud”)

3. Add *Inbound Rules* for the services you want to access on your AMI

- HTTP, SSH, “Custom TCP Rule” (42284) (20-21) (30000-30100), “All TCP” source: galaxy

The screenshot shows the AWS Management Console interface for managing a security group named "galaxy". The "Inbound" tab is selected. On the left, there's a form to "Create a new rule" with dropdowns for "Custom TCP rule", "Port range" (set to "80 (HTTP)", "22 (SSH)", "42284", "0 - 65535", "20 - 21", and "30000 - 30100"), and "Source" ("0.0.0.0/0"). A green "Add Rule" button is at the bottom. A note at the bottom says "Your changes have not been applied yet." On the right, a table lists the existing rules:

TCP Port (Service)	Source	Action
80 (HTTP)	0.0.0.0/0	Delete
22 (SSH)	0.0.0.0/0	Delete
42284	0.0.0.0/0	Delete
0 - 65535	sg-da5918b2	Delete
20 - 21	0.0.0.0/0	Delete
30000 - 30100	0.0.0.0/0	Delete

Galaxy on AWS (“the cloud”)

4. From the EC2 dashboard, select AMIs, and search for “galaxy” under *Public Images*

- Choose “galaxy-cloudman-2011-03-22” and click *Launch*

Amazon Machine Images			
Launch Spot Request Register New AMI De-register Permissions			
Viewing:		Public Images	All Platforms
<input type="text"/> galaxy			
Name	AMI ID	Source	
<input type="checkbox"/>	empty	 ami-46b9412f	pb_secondary_ami/pacbio_galaxy_1_2_1.manifest.xml
<input type="checkbox"/>	empty	 ami-561bc93f	072133624695/galaxy-cloudman-2012-02-26
<input type="checkbox"/>	empty	 ami-5a759d33	115971652512/galaxy-cloud-2010-07-01
<input type="checkbox"/>	empty	 ami-78a00411	861460482541/galaxy-cloudman-2012-05-10
<input type="checkbox"/>	empty	 ami-9a7485f3	861460482541/galaxy-cloudman-2011-01-12
<input checked="" type="checkbox"/>	empty	 ami-da58aab3	861460482541/galaxy-cloudman-2011-03-22
<input type="checkbox"/>	empty	 ami-df60bf6	072133624695/CBL_NoGalaxyUser_32bit



matt.shirley@jhmi.edu

Galaxy on AWS (“the cloud”)

Request Instances Wizard

CHOOSE AN AMI **INSTANCE DETAILS** CREATE KEY PAIR CONFIGURE FIREWALL REVIEW

Provide the details for your instance(s). You may also decide whether you want to launch your instances as "on-demand" or "spot" instances.

Number of Instances: **Instance Type:** Large (m1.large, 7.5 GiB)

Launch as an EBS-Optimized instance (additional charges apply):

Launch Instances
EC2 Instances let you pay for compute capacity by the hour with no long term commitments. This transforms what are commonly large fixed costs into much smaller variable costs.
Launch into: EC2 VPC
Availability Zone: No Preference

Request Spot Instances

Set Number of Instances = 1
Instance Type = “Large”
Availability Zone may be arbitrary

Back **Continue**

Galaxy on AWS (“the cloud”)

Request Instances Wizard

CHOOSE AN AMI **INSTANCE DETAILS** CREATE KEY PAIR CONFIGURE FIREWALL REVIEW

Number of Instances: 1 **Availability Zone:** No Preference

Advanced Instance Options

Here you can choose a specific **kernel** or **RAM disk** to use with your instances. You can also choose to enable CloudWatch Detailed Monitoring or enter data that will be available from your instances once they launch.

Kernel ID: **RAM Disk ID:**

Monitoring: Enable CloudWatch detailed monitoring for this instance
(additional charges will apply)

User Data:
 as text as file

```
cluster_name: plato
password: eu_a-mousoi
access_key: <Access Key ID>
secret_key: <Secret Access Key>
```

base64 encoded

Termination Protection: Prevention against accidental termination.

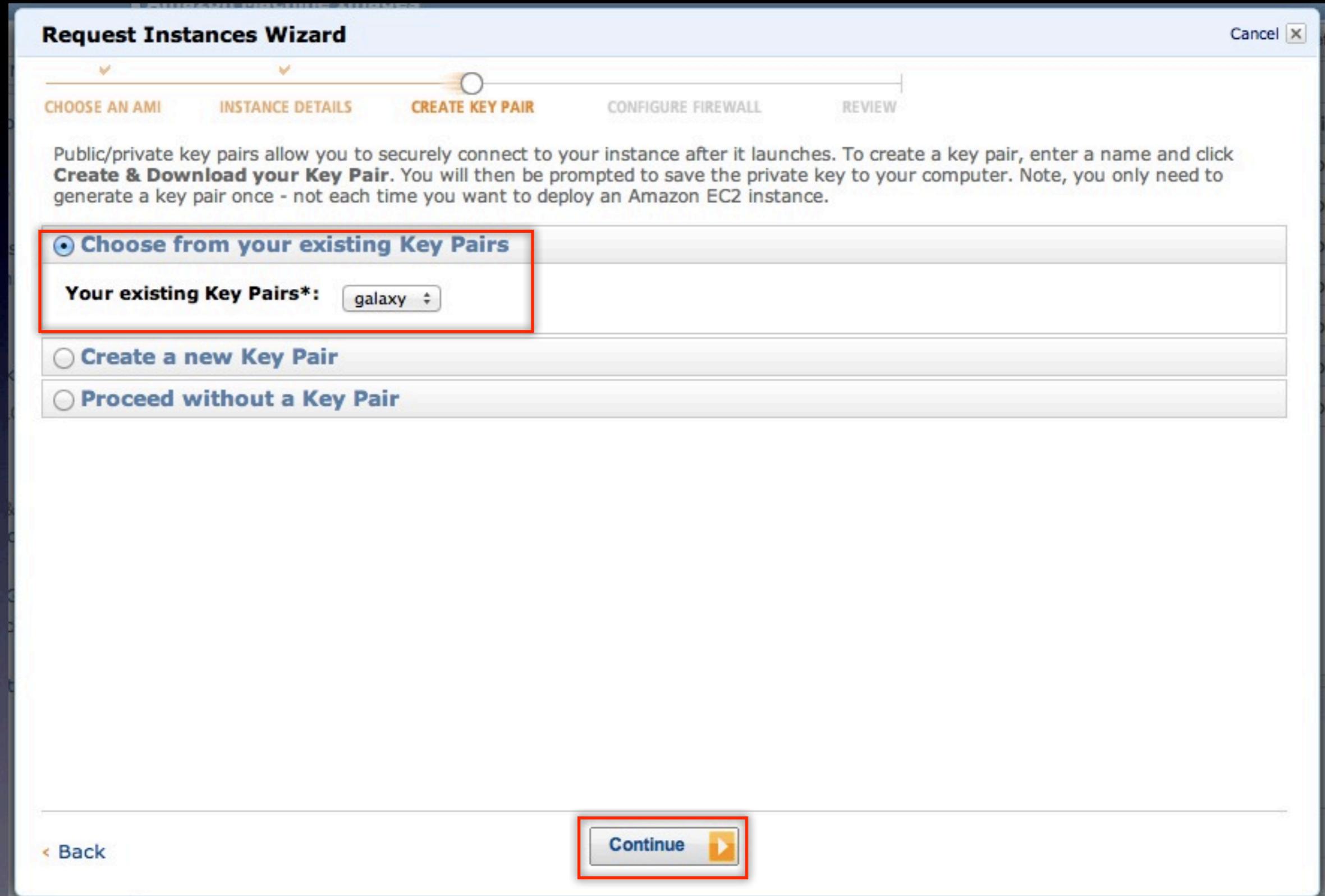
Shutdown Behavior:

IAM Role:

Continue 

Fill in User Data with information previously saved

Galaxy on AWS (“the cloud”)



Galaxy on AWS (“the cloud”)



Choose your “galaxy” security group

Galaxy on AWS (“the cloud”)

Request Instances Wizard

[Cancel](#)

CHOOSE AN AMI [INSTANCE DETAILS](#) [CREATE KEY PAIR](#) [CONFIGURE FIREWALL](#) [REVIEW](#)

Please review the information below, then click **Launch**.

AMI:  Other Linux AMI ID ami-da58aab3 (x86_64) [Edit AMI](#)

Number of Instances: 1

Availability Zone: No Preference

Instance Type: Large (m1.large)

Instance Class: On Demand [Edit Instance Details](#)

EBS-Optimized: No

Monitoring: Disabled **Termination Protection:** Disabled

Tenancy: Default

Kernel ID: Use Default **Shutdown Behavior:** Stop

RAM Disk ID: Use Default

Network Interfaces:

Secondary IP Addresses:

User Data: cluster_name: plato...

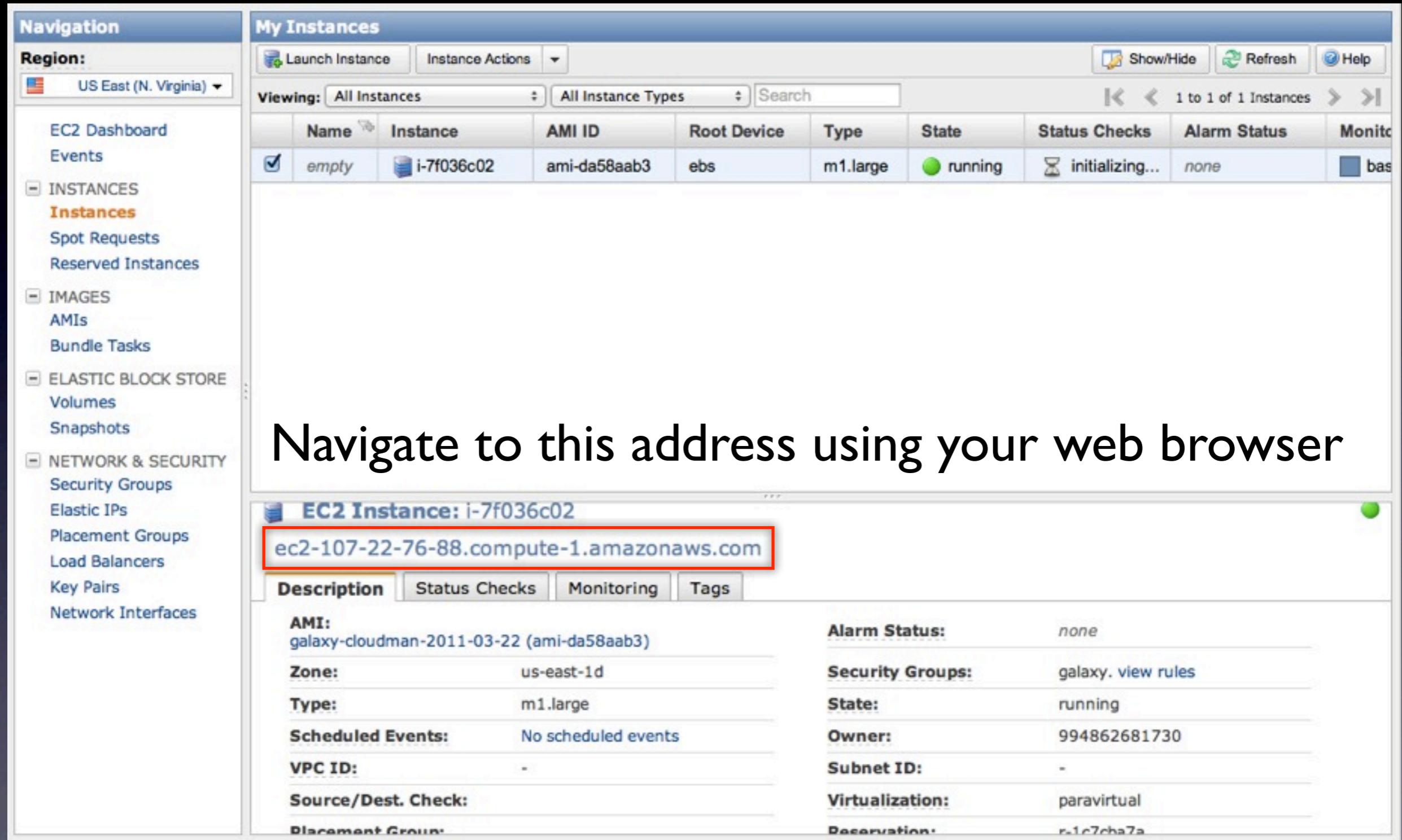
IAM Role: [Edit Advanced Details](#)

Key Pair Name: galaxy [Edit Key Pair](#)

Security Group(s): sg-da5918b2 [Edit Firewall](#)

[Back](#) [Launch](#) 

Galaxy on AWS (“the cloud”)



The screenshot shows the AWS Management Console interface for the EC2 service. The left sidebar contains a navigation menu with various AWS services like EC2 Dashboard, Events, Instances, Images, Elastic Block Store, Network & Security, and more. The main content area is titled "My Instances" and displays a table of running instances. There is one instance listed: "empty" (ID: i-7f036c02), which was launched using the "galaxy-cloudman-2011-03-22" AMI (AMI ID: ami-da58aab3). The instance is of type m1.large, using an ebs root device, and is currently running. A status check is shown as "initializing...".

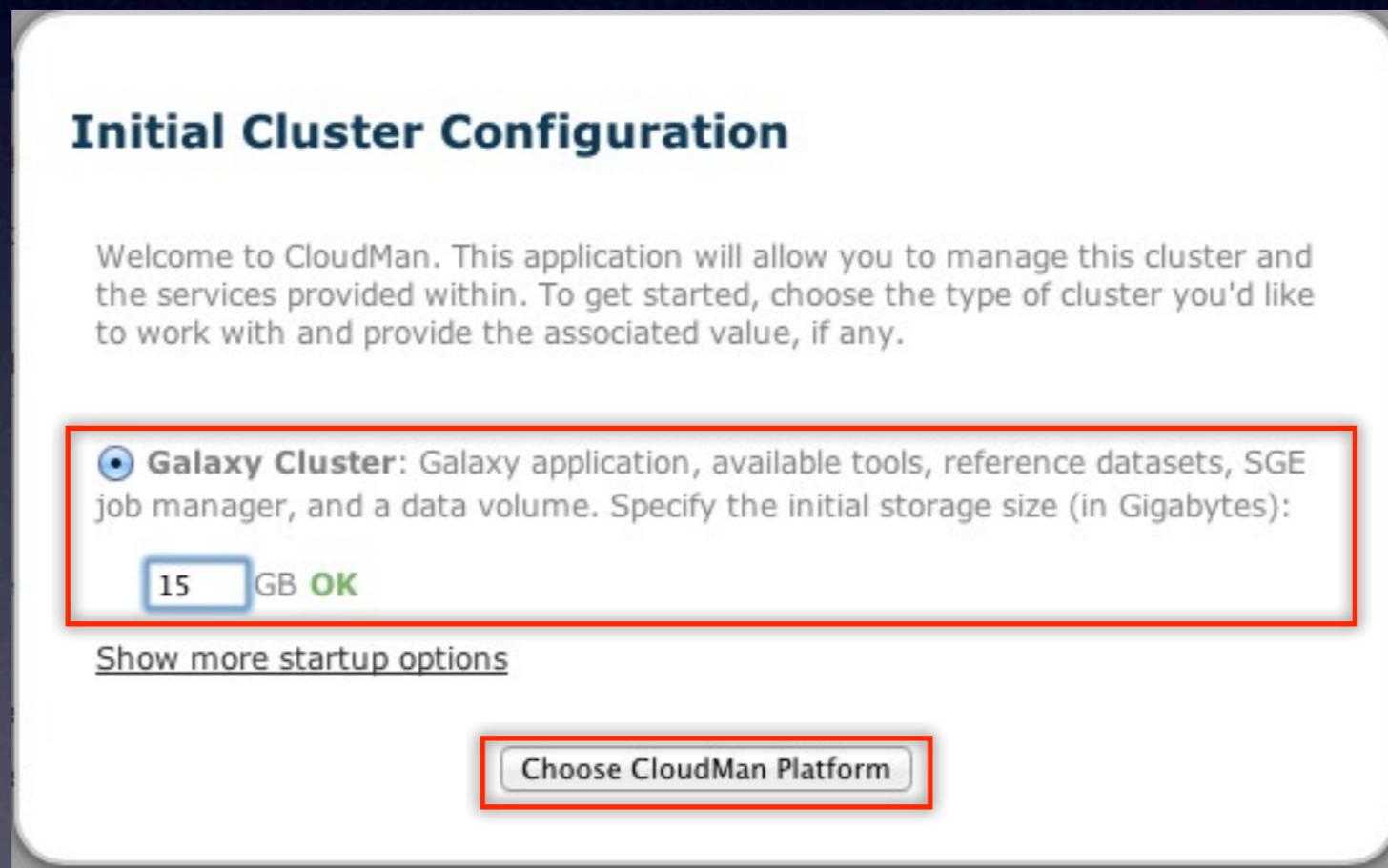
Navigate to this address using your web browser



The screenshot shows the detailed view for the EC2 instance i-7f036c02. The instance's public IP address, "ec2-107-22-76-88.compute-1.amazonaws.com", is highlighted with a red box. Below the IP address, the instance's configuration details are listed in two columns. The left column includes: AMI (galaxy-cloudman-2011-03-22), Zone (us-east-1d), Type (m1.large), Scheduled Events (No scheduled events), VPC ID (-), Source/Dest. Check, and Placement Group. The right column includes: Alarm Status (none), Security Groups (galaxy, view rules), State (running), Owner (994862681730), Subnet ID (-), Virtualization (paravirtual), and Reservation (-).

Galaxy on AWS (“the cloud”)

5. After logging in using the previously specified “cluster name” and “password”, specify the initial storage for the Galaxy cluster



Galaxy on AWS (“the cloud”)

6. After a few minutes, the *Access Galaxy* button will become accessible, signaling success
- Note that performance will be improved if autoscaling is turned on, but at a higher cost

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.

[Terminate cluster](#) [Add nodes ▾](#) [Remove nodes](#) [Access Galaxy](#)

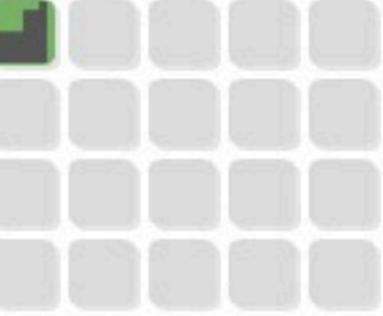
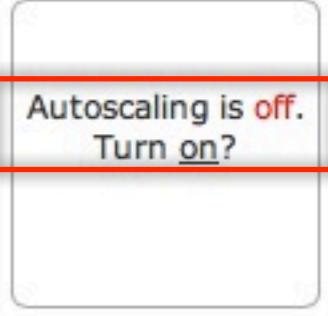
Status

Cluster name: plato 

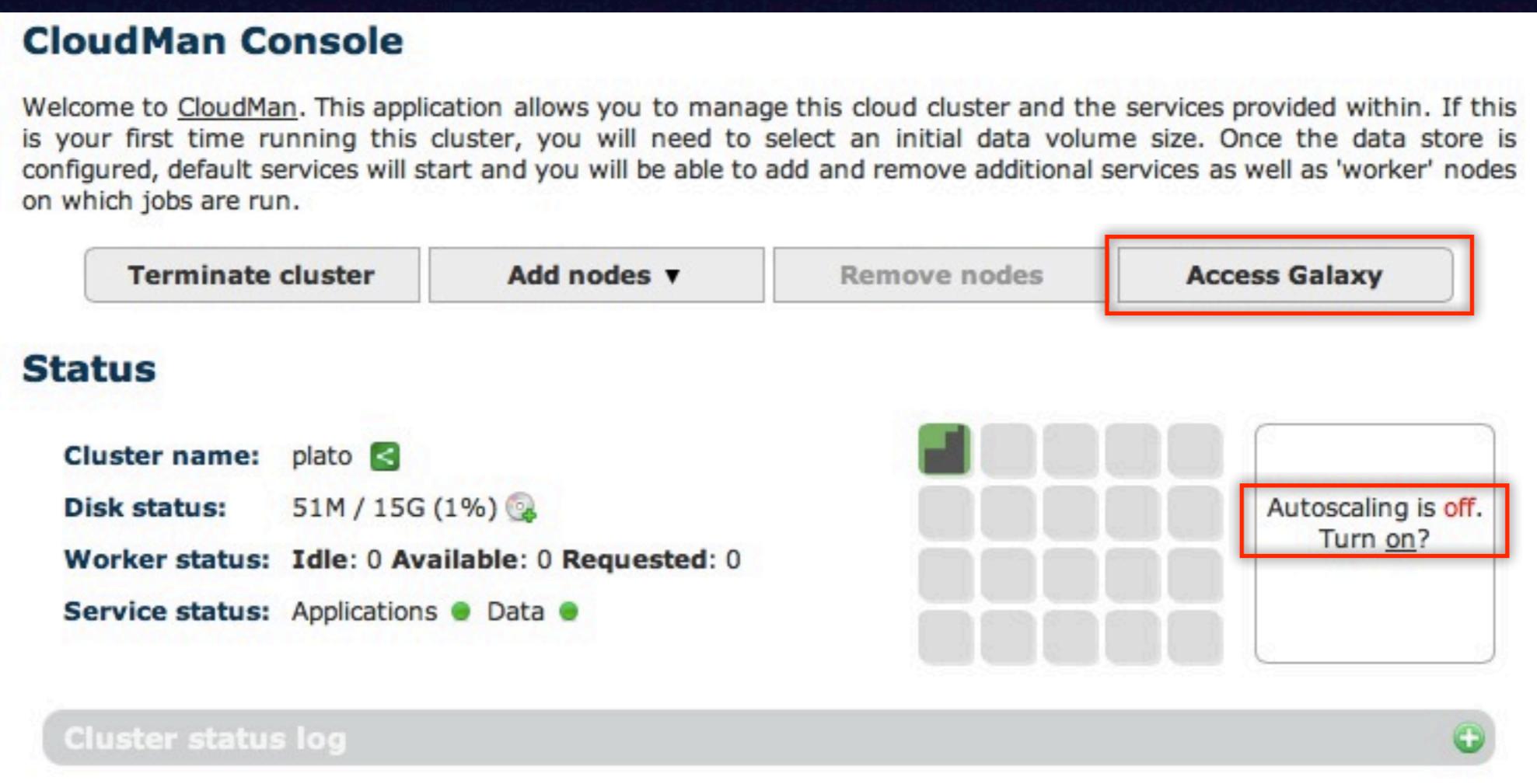
Disk status: 51M / 15G (1%)  

Worker status: Idle: 0 Available: 0 Requested: 0

Service status: Applications  Data 

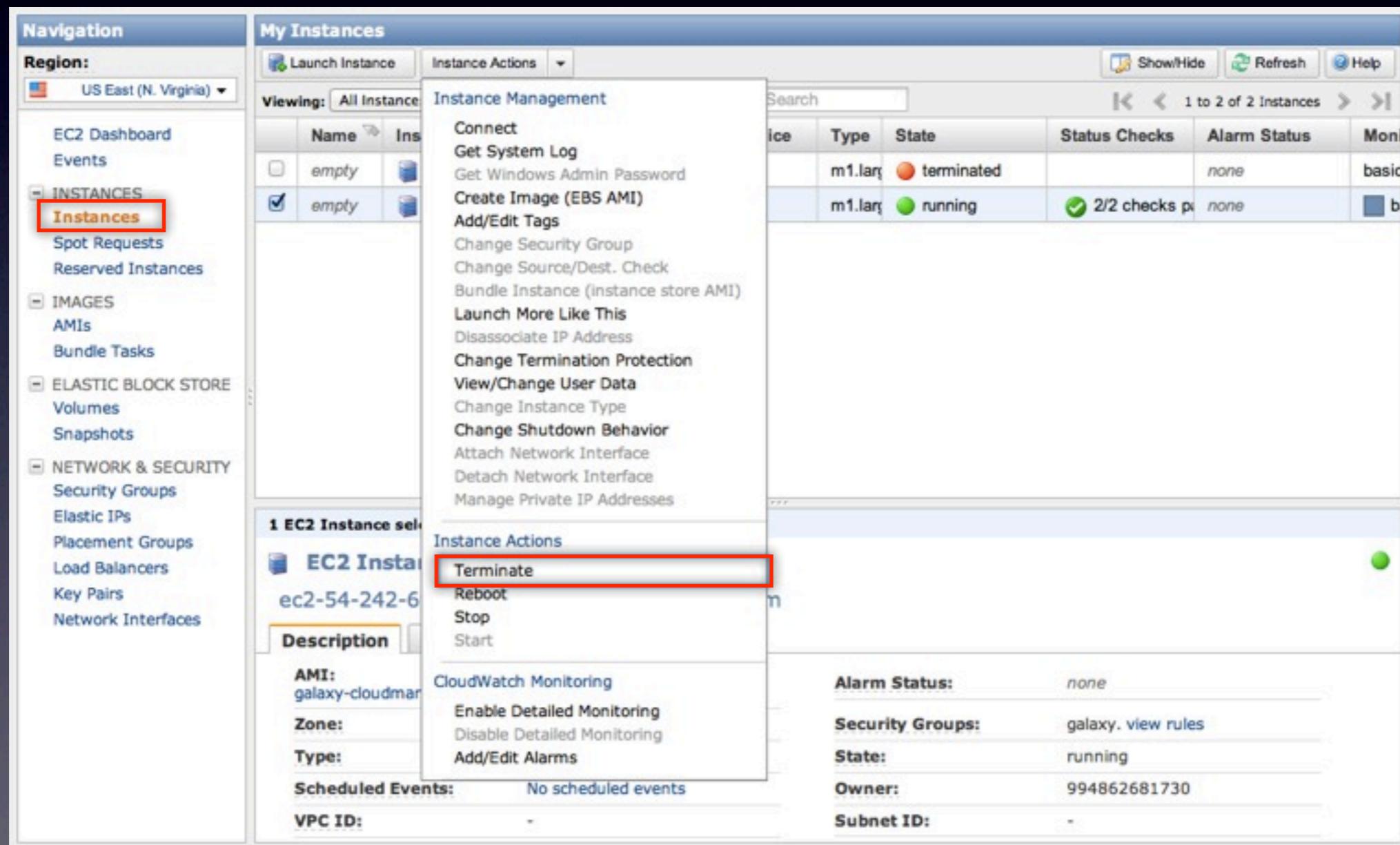
 
Autoscaling is off.
[Turn on?](#)

[Cluster status log](#) 



Galaxy on AWS (“the cloud”)

7. After you are finished with your Galaxy cluster, remember to *terminate* it, as AWS charges for compute time



Galaxy on AWS (“the cloud”)

8. Also remember to delete any EBS volumes that persist

The screenshot shows the AWS Management Console interface for managing EBS volumes. The left sidebar, titled 'Navigation', includes links for EC2 Dashboard, Events, Instances, Spot Requests, Reserved Instances, Images, AMIs, Bundle Tasks, Elastic Block Store (with 'Volumes' highlighted by a red box), and Network & Security. The main content area is titled 'EBS Volumes' and displays a table of four volumes. The columns are: Name, Volume ID, Capacity, Volume Type, Snapshot, Created, Zone, State, and Alarm. The volumes listed are:

Name	Volume ID	Capacity	Volume Type	Snapshot	Created	Zone	State	Alarm
empty	vol-f2abf088	15 GiB	standard	snap-a95003c5	2012-10-07T00:02:01	us-east-1d	in-use	none
empty	vol-c5da81bf	700 GiB	standard	snap-5b030634	2012-10-07T00:10:58	us-east-1d	in-use	none
empty	vol-5dd88327	10 GiB	standard	snap-cf3746b3	2012-10-07T00:11:16	us-east-1d	in-use	none
empty	vol-16d8836c	15 GiB	standard	--	2012-10-07T00:11:25	us-east-1d	in-use	none

At the bottom of the main pane, it says '0 Volumes selected' and 'Select a volume above'.