# Galaxy

# for high-throughput sequence data analysis

http://usegalaxy.org

# The Galaxy Team







Guru Ananda



### Dannon Baker





Dan Blankenberg Ramkrishna Chakrabarty



Nate Coraor



Jeremy Goecks



Greg von Kuster

Kanwei Li



**Kelly Vincent** 

Supported by the National Human Genome Research Institute (HG005542, HG004909, HG005133), the National Science Foundation (DBI-0850103), Penn State University, Emory University, and the Pennsylvania Department of Public Health

# Galaxy: accessible analysis system



# What is Galaxy?

- A free (for everyone) web service integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage
- Open source software that makes integrating your own tools and data and customizing for your own site simple

# Integrating existing tools into a uniform framework



- Defined in terms of an abstract interface (inputs and outputs)
  - In practice, mostly command line tools, a declarative XML description of the interface, how to generate a command line
- Designed to be as easy as possible for tool authors, while still allowing rigorous reasoning

# **Galaxy analysis interface**



- Consistent tool user interfaces automatically generated
- History system facilitates and tracks multistep analyses

## Automatically tracks every step of every analysis

7: Map with Bowtie for Illumina on data 6 and o 9,073,928 lines, format: 9 database and Run this job again the aligne	● Ø <u>lata 5</u> sam, d.	*
	0	
1.QNAME	2.FLAG	3.1
HWI-EAS269:3:1:1449:913	99	ch
HWI-EAS269:3:1:1449:913	147	ch
HWI-EAS269:3:1:709:832	99	chi
HWI-EAS269:3:1:709:832	147	ch
HWI-EAS269:3:1:1422:1087	99	ch
HWI-EAS269:3:1:1422:1087	147	ch
$\Theta$	) 4	•

Map with Bowt	ie for Illumina
Will you select	a reference genome from your history or use a built-in index?
Use a built-in i	ndex 🗘
Built-ins were i	ndexed using default options
Select a refere	nce genome:
mm9	~
if your genome	of interest is not listed - contact Galaxy team
Is this library i	nate-paired?:
Paired-end	
Farmered FACTO	
Forward FASIC	t file:
S: E18 PE.1 Rea	as actual multiplication with ASCII officer 33
must nave sang	er-scaled quality values with ASCII offset 55
Reverse FASTO	file:
6: E18 PE.2 Rea	ds Gred, Trimmed
Must have Sang	er-scaled quality values with ASCII offset 33
Maximum inse	rt size for valid paired-end alignments (-X):
1000	
The upstream/ against the for	downstream mate orientation for valid paired-end alignment ward reference strand (fr/rf/ff):
FR (for Illumina	
Bowtie setting	s to use:
Commonly use	۵ 🔹
For most mapp Full parameter	ng needs use Commonly used settings. If you want full control use ist
Suppress the h	eader in the output SAM file:
Bowtie produce	s SAM with several lines of header information by default
(Execute)	

### As well as user-generated metadata and annotation...

istory		Ор	tions 🔻
			02
ariant Analysi	s for San	nple E18	
ags:			
snp × pile	) (x qu	bowtie ×	
demo x sa	mple:e18	× 4	
nnotation / N	otes:		
erform a variar	t analysis	with defa	ult
arameters to ic	entify var	riants in sa	mple
18 that lie in a	nnotated	genes.	



# Galaxy workflow system

🔴 🔿 🔿	edu/work	flow/editor%	Ga d=6ad323	laxy 258c4F75	dec .				c	Q+ Cos	ole	-
- Galaxy Analy	rze Data	Workflow	Deta Lib	araries	Lab.	Admin	Help	User				
Tools Get. Data Send. Data ENCODE. Tools Lift=Over Text. Manipulation Genvert. Formats Ither. and Sent Join. Subtract and Group Entract. Features Fetch. Alignments Get. Genomic. Scores Operate: on Genomic. Intervals Statistics Graph/Display. Data Regional. Variation Multiple: regression Multiple: Regression Mul	Workfi it datass uit nt datass Nut FAS1 Seque outp	et 35 et 35 et 35 FA-to-Tabu vert these ences ut (tabular) Add colum to Query out, file1 Tabular- Tabular- Tabular	I Metagen Selv Roa Qua out out out tar 32 n H to-FAST/ nited file Ista)	ect high ds lity score put1 (fas	quality es ta)	segment	s 14 Megabi Compa sequen output Negabi Compa sequen output	last re these ces L (tabula last re these ces L (tabula	2 0 2 2		Option	(
EIDUIS .						11	1111					15

- Workflows can be constructed from scratch *or* extracted from existing analysis histories
- Facilitate reuse, as well as providing precise reproducibility of a complex analysis

# Analyzing high throughput sequence data with Galaxy

- The Galaxy framework is generic, supporting a new type of analysis is as simple as integrating tools
- Galaxy is well suited to large-scale analysis
  - Allows tools to work with data in native, efficient formats
  - Integrates easily with cluster computing resources

# (some) Galaxy tools for sequence data analysis

### NGS: QC and manipulation

**ILLUMINA DATA** 

- <u>FASTQ Groomer</u> convert between various FASTQ quality formats
- <u>FASTQ splitter</u> on joined paired end reads
- <u>FASTQ joiner</u> on paired end reads
- FASTQ Summary Statistics by column

ROCHE-454 DATA

- Build base quality distribution
- Select high quality segments
- <u>Combine FASTA and QUAL</u> into FASTQ

AB-SOLID DATA

- · Convert SOLiD output to fastq
- <u>Compute quality statistics</u> for SOLID data
- Draw quality score boxplot for SOLID data

### GENERIC FASTQ MANIPULATION

- <u>Filter FASTQ</u> reads by quality score and length
- FASTO Trimmer by column

### evolution

Metagenomic analyses Human Genome Variation EMBOSS

NGS TOOLBOX BETA

NGS: QC and manipulation NGS: Mapping ILLUMINA

- Map with Bowtie for Illumina
- Map with BWA for Illumina ROCHE-454
- <u>Lastz</u> map short reads against reference sequence
- <u>Megablast</u> compare short reads against htgs, nt, and wgs databases
- Parse blast XML output

AB-SOLID

Map with Bowtie for SOLID
 NGS: SAM Tools
 NGS: Indel Analysis

NGS: Peak Calling NGS: RNA Analysis

### RGENETICS

SNP/WGA: Data; Filters

#### NGS TOOLBOX BETA

#### NGS: QC and manipulation

NGS: Mapping

### NGS: SAM Tools

- <u>Filter SAM</u> on bitwise flag values
- · Convert SAM to interval
- <u>SAM-to-BAM</u> converts SAM format to BAM format
- <u>BAM-to-SAM</u> converts BAM format to SAM format
- Merge BAM Files merges BAM files together
- <u>Generate pileup</u> from BAM dataset
- Filter pileup on coverage and SNPs
- <u>Pileup-to-Interval</u> condenses pileup format into ranges of bases
- <u>flagstat</u> provides simple stats on BAM files

NGS: Indel Analysis NGS: Peak Calling NGS: RNA Analysis

### RGENETICS

SNP/WGA: Data; Filters

### NGS: SAM Tools

### NGS: Indel Analysis

- Filter Indels for SAM
- Extract indels from SAM
- Indel Analysis

### NGS: Peak Calling

- <u>MACS</u> Model-based Analysis of ChIP-Seq
- <u>GeneTrack indexer</u> on a BED file
- <u>Peak predictor</u> on GeneTrack index

### NGS: RNA Analysis

RNA-SEQ

- <u>Tophat</u> Find splice junctions using RNA-seq data
- <u>Cufflinks</u> transcript assembly and FPKM (RPKM) estimates for RNA-Seq data
- <u>Cuffcompare</u> compare assembled transcripts to a reference annotation and track Cufflinks transcripts across multiple experiments
- <u>Cuffdiff</u> find significant changes in transcript expression, splicing, and promoter use

FILTERING

utput	Tophat 🔉	
put dataset 🔀 utput	RNA-Seq FASTQ file RNA-Seq FASTQ file junctions (bed) accepted_hits (bam)	
Input dataset 🗱		(
	Tonhat %	
Input dataset 🕱 output	Tophat 24 RNA-Seq FASTQ file RNA-Seq FASTQ file junctions (bed) accepted_hits (barr)	
Input dataset 18 output	Tophat 24 RNA-Seq FASTQ file RNA-Seq FASTQ file junctions (bed) accepted_hits (bam)	

Cufflinks	x
SAM or BAM file of aligned RNA Seq reads	-
Reference Annotation	
genes_expression (tabular)	04
transcripts_expression (tabular)	
assembled_isoforms (gtf)	00
Cufflinks	35
SAM or BAM file of aligned RNA- Seq reads	
Reference Annotation	
Reference Annotation genes_expression (tabular)	00
Reference Annotation genes_expression (tabular) transcripts_expression (tabular)	00

Cuffcompare	32
GTF file produced by Cufflinks	
GTF file produced by Cufflinks	
Reference Annotation	
transcripts_accuracy (txt)	0
input1_tmap (tabular)	
input1_refmap (tabular)	
input2_tmap (tabular)	
input2_refmap (tabular)	
transcripts_tracking (tabul	ar)
transcripts_combined (gtf)	

	_
Cuffdiff	×
Transcripts	
SAM or BAM file of aligned RNA- Seq reads	
SAM or BAM file of aligned RNA- Seq reads	
soforms_exp (tabular)	
genes_exp (tabular)	
ss_groups_exp (tabular)	
cds_exp_fpkm_tracking (tabular)	
soforms_fpkm_tracking (tabular)	
genes_fpkm_tracking (tabular)	
tss_groups_fpkm_tracking (tabula	r)
cds_fpkm_tracking (tabular)	
splicing_diff (tabular)	
promoters_diff (tabular)	
cds_diff (tabular)	0

**Example:** Workflow for differential expression analysis of RNA-seq using Tophat/Cufflinks tools

## **Community of tool developers**

00

### 💳 Galaxy Tool Shed / (beta)

Tools Help

Galaxy Tool Shed

User

C Q- Google

Community

Tools

- Browse by category
- Browse all tools
- · Login to upload

-							
a	te	a	n	rı	ρ	S	
~~~		3	~	•	~	-	

search	Q	Advanced Search

Name 4	Description	Tools
Convert Formats	Tools for converting data formats	5
Data Source	Tools for retrieving data from external data sources	1
Fasta Manipulation	Tools for manipulating fasta data	5
Next Gen Mappers	Tools for the analysis and handling of Next Gen sequencing data	7
Ontology Manipulation	Tools for manipulating ontologies	1
SAM	Tools for manipulating alignments in the SAM format	0
Sequence Analysis	Tools for performing Protein and DNA/RNA analysis	10
SNP Analysis	Tools for single nucleotide polymorphism data such as WGA	1
Statistics	Tools for generating statistics	1
Text Manipulation	Tools for manipulating data	3
Visualization	Tools for visualizing data	1

00

+ http://community.g2.bx.psu.edu/

### 💳 Galaxy Tool Shed / (beta)

Tools Help User

Galaxy Tool Shed

C Q- Google

Community

#### Tools

- Browse by category
- · Browse all tools
- · Login to upload

<u>Name</u>	Description	<u>Version</u>	Category	Uploaded By	<u>Type</u>	Average Rating
AGILE	Quickly match reads to a reference genome or sequence file	1.0.0	<ul> <li><u>Next Gen</u> <u>Mappers</u></li> <li><u>Sequence</u> <u>Analysis</u></li> </ul>	simonl	Tool	***
assemblystats	Summarise an assembly (e.g. N50 metrics)	1.0.1	<ul> <li><u>Next Gen</u> <u>Mappers</u></li> <li><u>Sequence</u> <u>Analysis</u></li> </ul>	konradpaszkiewicz	Tool	*** **
Divide FASTQ file into paired and unpaired reads	using the read name suffices	0.0.4	<ul> <li><u>Text</u> <u>Manipulation</u></li> <li><u>Sequence</u> <u>Analysis</u></li> </ul>	peterjc	Tool	*** **
FastQC	quality control checks on raw sequence data	1.0.0	<ul> <li>Fasta Manipulation</li> <li>Sequence Analysis</li> </ul>	jjohnson	Tool	***
Filter FASTA by ID	from a tabular file	0.0.3	<ul> <li>Fasta Manipulation</li> <li>Sequence Analysis</li> <li>Text Manipulation</li> </ul>	<u>peterjc</u>	Tool	***

00

+ = http://community.g2.bx.psu.edu/

#### Galaxy Tool Shed

C Q- Google

Tool Actions -

### 🚾 Galaxy Tool Shed / (beta)

Tools Help User

#### Community

Tools

- Browse by category
- · Browse all tools
- · Login to upload

3			
VIO	14/	0.01	
VIC	WW.	1001	21

This is the latest approved version of this tool suite

#### Mothur Metagenomics

Tool Id: Mothur\_toolsuite

Version:

1.15.1

#### Description: Mothur metagenomics commands as Galaxy tools

User Description:

Provides galaxy tools for the commands in the Hothur metagenomics package: http://www.mothur.org/wiki/Hain\_Page

#### Uploaded by:

**j**johnson

#### Date uploaded: about 22 hours ago

Categories:

Sequence Analysis

#### **Tool Contents**

- Mothur toolsuite 1.15.1.tar.gz
  - mothur/
  - mothur/tools/
  - mothur/tools/mothur/
  - mothur/tools/mothur/split.abund.xml

## Data management

## Everything can be shared and published

### Sharing and Publishing History 'Variant Analysis for Sample E18'

### Making History Accessible via Link and Publishing It

This history accessible via link and published.

Anyone can view and import this history by visiting the following URL:

http://main.g2.bx.psu.edu/u/jgoecks/h/variant-analysis-for-sample-e18 /

This history is publicly listed and searchable in Galaxy's Published Histories section.

You can:

#### **Unpublish History**

Removes history from Galaxy's Published Histories section so that it is not publicly listed or searchable.

#### Disable Access to History via Link and Unpublish

Disables history's link so that it is not accessible and removes history from Galaxy's <u>Published Histories</u> section so that it is not publicly listed or searchable.

### Sharing History with Specific Users

You have not shared this history with any users.

Share with a user

Back to Histories List



000			(	Galaxy							
	/main.g2.bx.psu.ed	u/library						¢	Q- Google		)
💳 Galaxy	Analyze Data	Workflow	Shared Data	Lab	Visualization	Admin	Help	User			
	ells 🔻										0
🗆 🕨 🔚 GIE-E	R4 Cells 🔻										
🗐 🕨 📴 MEL Ya	ale Cells 🔻										
🔲 🔻 📴 Enriched	v										m
🗐 🔻 📴 CTCF (	ChIP-seq 🔻										
🗏 🔻 📄 <u>сн</u>	12 Cells 👻										
🗆 🕨 🚞	Pooled V										
🗐 🔻 🚘	Replicate 1 🔻										
01 rea	Feb2010 In7 CTCF	CH12 groome	<u>d</u> 👻 None	:		dan	@bx.psu.ec	du	2010- 10-06	2.0 Gb	U
	MACS peak calls (b	roadPeak) 🔻	None	2		dan	@bx.psu.ec	du	2010- 10-06	903.0 Kb	
	Mapped Tags (BAM	) <del>~</del>	None	2/		dan	@bx.psu.ec	du	2010- 10-06	493.0 Mb	
Θ	Tag Counts (bigWig	<u>a)</u> 🔻	None	2		dan	@bx.psu.ec	du	2010- 10-06	2.0 Gb	
🗆 🕨 🚞	Replicate 2 🔻										
🗆 🕨 📴 <u>G1</u>	E Cells 🔻										*
Display a menu											1



Display a menu

# Sample Tracking





illumina

e e e	Analyze Data
Data Library "Windsh	ield splatter"
Name	
□ Trip A Left Side QV ▼	
Trip A Left Side Reads V	
🗇 Trip A Right Side QV 🔻	
Trip A Right Side Reads	
□ Trip B Left Side QV ▼	
□ Trip B Left Side Reads ▼	
🗆 Trip B Right Side QV 🔻	
Trip B Right Side Reads 🔻	3

Lab: Sequencing Request Tracking

Analyze Data Workflow Shared Data Lab Visualization Admin Help User      Galaxy     Analyze Data Workflow Shared Data Lab Visualization Admin Help User      Browse requests      Create a new sequencing request      Select a request type configuration:     Pacific Biosciences ?      Contact the lab manager if you are not sure about the request type configuration. Name of the Experiment Snail transcriptome     (Required)      Description      (Optional)      Scientific Contact      Select one     (Optional)      Save Add samples	00			(	Galaxy						
Calaxy       Analyze Data       Workflow       Shared Data       Lab       Visualization       Admin       Help       User         Browse requests         Create a new sequencing request         Select a request type configuration:         Pacific Biosciences @       Contact the lab manager if you are not sure about the request type configuration.         Name of the Experiment       Snail transcriptome       (Required)         Description       (Optional)       Scientific Contact       (Optional)         Scientific Contact       Save       Add samples       Add samples	+ http://localhost:8080/	/requests								¢	Q* Google
Browse requests         Create a new sequencing request         Select a request type configuration:         Pacific Biosciences         Contact the lab manager if you are not sure about the request type configuration.         Name of the Experiment         Snail transcriptome         (Required)         Description         (Optional)         Scientific Contact         Select one         (Optional)         Save	- Galaxy	Analyze Data	Workflow	Shared Data	Lab	Visualization	Admin	Help	User		
Create a new sequencing request Select a request type configuration: Pacific Biosciences : Contact the lab manager if you are not sure about the request type configuration. Name of the Experiment Snail transcriptome (Required) Description (Optional) Scientific Contact Select one Select											Browse requests
Select a request type configuration: Pacific Biosciences Contact the lab manager if you are not sure about the request type configuration. Name of the Experiment Snall transcriptome (Required) Description (Optional) Scientific Contact Select one Coptional) Save Add samples	Create a new sequencing request										(
	Select a request type configuration: Pacific Biosciences : Contact the lab manager if you are no Name of the Experiment Snail transcriptome (Required) Description (Optional) Scientific Contact Select one (Optional) Save Add samples	t sure about the	request type	configuration.							
	Display a menu										

Customers create their own requests

00				Galaxy		
◄ ▶ + (	http://loc	alhost:8080/requests				C Q Google
💳 Galax	y	Analyze Data	Workflow Shar	ed Data Lab Visualizati	on Admin Help User	
						(Request Actions *)
Add Samples	to Seque	ncing Request "Snail to	ranscriptome"			
Name	State	Data Library	Folder	History	Workflow	
Sample 1		Select one	Select one	Lambda History	lambda 🛟	
(required)					Input BAM:	
					Coverage (gff) Reference Cenome:	
					lambda_ref.fasta	
Copy 1 Select the sam Add sample Click the Add	samples fr ple from wh Save (C sample but ples from c	nich the new sample should	be copied or leave s	election as <b>None</b> to add a new ton when you have finished ad	v "generic" sample. Iding samples.	
Display a menu						

Customers add samples, cand define library to deposit in and/or workflows to run

000		Ga	laxy Admi	inistration				
- Galaxy	Analyze Data	Workflow	Shared Da	ita Lab	Admin He	lp User		
Administration Security • Manage users	Sequencing search	Create new request						
Manage groups     Manage roles	Name	Desc	ription	Samples	Sequencer	Last Updated †	State	<u>User</u>
<ul> <li>Manage data libraries</li> <li>Server</li> </ul>	<u>Custome</u> <u>Lab 1</u>	<u>ir</u> *		1	<u>Core</u> Facility 454	14 minutes ago	In Progress	nate@bx.psu.edu
<u>Reload a tool's configuration</u> <u>Profile memory usage</u>	For 0 sel	ected requests	Delete	Undelete	D			
<ul> <li>Manage jobs</li> <li>Forms</li> <li>Manage forms</li> <li>Sample Tracking</li> <li>Sequencer configurations</li> <li>Sequencing requests</li> <li>Find samples</li> </ul>								

### Facility managers monitor requests

000		Galaxy	Administr	ation		
🚾 Galaxy	Analyze Data V	Workflow Sha	ired Data	Lab Admin I	Help User	
Administration Security • Manage users • Manage groups • Manage roles Data • Manage data libraries	Sequencing Request Informat	Request "C cility 454   User ation	Custome :: nate@bx.p	e <b>r Lab 1"</b> osu.edu   State: <i>In P</i>	Sequencing Req	uest Actions - Browse requests
Server <ul> <li><u>Reload a tool's configuration</u></li> <li><u>Profile memory usage</u></li> <li><u>Manage jobs</u></li> </ul> Forms <ul> <li><u>Manage forms</u></li> </ul> <li>Sample Tracking <ul> <li>Sequencer configurations</li> </ul> </li> <li>Sequencing requests</li> <li><u>Find samples</u></li>	Name     Sample 1     For selected sample     Change current     Sample Received     Comments     Optional     Save Cancel      Edit samples	Barcode	State New te	Data Library Customer Lab	Folder Customer Lab	O/O

### Lab staff manage sample state

000		C	Galaxy Administ	tration				
- Galaxy	Analyze Data	Workflow	Shared Data	Lab	Admin	Help	User	
Administration  Security  Manage users Manage groups Manage roles Data Manage data libraries Server Reload a tool's configuration Profile memory usage	Analyze Data Select files for Sample: Sample 1 : Select the samp Folder path on /data/run1/ run1.fa run1.qv	Workflow transfer le with which the sequence	Shared Data	Lab	Admin Sequencer the dataset (List conter	Help configu s	User ration "Core Facility 454"	Browse this request
<ul> <li>Manage jobs</li> <li>Forms</li> <li>Manage forms</li> <li>Sample Tracking</li> <li>Sequencer configurations</li> <li>Sequencing requests</li> <li>Find samples</li> </ul>	Select & show d	latasets) (Se	lect more					

### Simple data management: manual transfer to Galaxy Data Libraries

00			-		3	Galaxy							
< > + (	http://localhost	t:8080/reques	ts								c	Q+ Google	
💳 Galax	y		Analyze Data	Workflow	Shared Data	Lab	Visualization	Admin	Help	User			
												Reques	t Actions 👻
Sequencing re	ouest "Snail trai	nscriptome"											
Current state: Complete													
Description:													
User: james.taylor@e	mory.edu												
Request type: Pacific Bioscien	ices												
More													
Samples													
Name	Barcode	State	Data Library		Folder		н	istory		W	orkflow	Run Datasets	
Sample_1		Done	Pacific Bioscience	5	Pacific Bioso	ciences	L	ambda Hist	ory	la	mbda	<u>6</u>	
Additional in	formation												
Display a menu													

Automatic transfer, data loaded into Galaxy and workflow run with no intervention

## Sample tracking is completely extensible

- Track manually, with barcodes, or integrate with an existing LIMS
- Everything is configuration driven, capture whatever data and support whatever workflow you want
- Interaction with sequence instruments and secondary analysis is completely pluggable
  - For services that provide a web / REST API even easier

## Making Galaxy your own

## **Building local Galaxy instances**

- Galaxy is designed for local installation and customization
  - Just download and run, completely self-contained
  - Easily integrate new tools
  - Easy to deploy and manage on nearly any (unix) system
  - Run jobs on existing compute clusters

# Scale up on your cluster

- Move intensive processing (tool execution) to other hosts
- Frees up the application server to serve requests and manage jobs
- Utilize existing resources
- Supports any scheduler that supports DRMAA (most of them)
- It's easy
- But, requires an **existing computational resource** on which to be deployed





GRIDENGINE





# **Cloud computing / Infrastructure virtualization**

- Computing using resources acquired on demand
- Virtual infrastructure allows for (potential) economies of scale, and (definite) improvements to management automation
- Cloud-style deployment provides a solution both for users without dedicated compute resources, and for simplifying deployment and management

## Using Amazon EC2: Startup in 3 steps

Organization     O	00			AWS Managem	ent Console				-B.			
• Add Present Controls Support Activity       • Add Management Control         • Add Present Target Activity       • Add Management Control       • Outcome         • Add Present Target Activity       • Add Management Control       • Outcome         • Add Present Target Activity       • Add Present Target Activity       • Outcome         • Add Present Target Activity       • Add Present Target Activity       • Outcome       • Outcome         • Add Present Target Activity       • Add Present Target Activity       • Outcome       • Outcome       • Outcome         • Add Present Target Activity       • Add Present Target Activity       • Outcome       •	< > + https://cor	nsole.aws.amazon.co	om/ec2/home#c=EC2&s=Lau	nchinstanceWizard			C Q+ Coo	ogle				
Autor of Columnation       Alt Standament Columnation         Viscands       Image: Columnation       Image: Columnation       Image: Columnation       Image: Columnation         Viscands       Image: Columnation       Image: Columnat	🙀 aws.amazon.com	AWS Products	Developers Community S	uppert Account		Welco	me, James Taylor	Settings Sign Out	12			
NATURE Distance	Amazon EC2 Am	Request Inst				AWS Management Con	sole		_			
Repert Contraction   ************************************	Navigation	-0-	+ https://co	nsole.aws.amazon.com/ec	2/home#c=EC2&s=Launchinstan	ceWizard			c	Q+ Google		
C. C. Mathonis III     C. Mathonis IIII     Mathonis Taski      Database     Mathonis Taski      Database     Mathonis Taski      Database     Database     Mathonis Taski      Database     Databas	Region: 🔛 US Em	Choose an Ama	👔 awa.amazon.com	AWS Products Devi	obspare Community Support A	cebunt			Welcome, Jam	nn Taylor Settings Sig	er Cuit	
Instructions   * Specific Tasks   * Artis   * Substructions   * Standboord   * Standboord </td <td>* EC2 Deshboard</td> <td>Quick Start</td> <td>Amazon EC2</td> <td>Request Instance</td> <td>000</td> <td></td> <td></td> <td>AWS Mana</td> <td>gement Consol</td> <td>e</td> <td></td> <td>-</td>	* EC2 Deshboard	Quick Start	Amazon EC2	Request Instance	000			AWS Mana	gement Consol	e		-
A space Name of the Space N	INSTANCES	Viewing: All In	Manhantion	-	+ https://consol	e.aws.amazon.com/ec2/	home#c=EC2&s=ir	istances			c	Q+ Coogle
Indust   ANSe   ANSe   ANSE   CATIO Erados   MASS   Ansibility Zeit   Sapaholos   KETWORKING E SACUE   Sapaholos   MASS   Sapaholos<	* Spat Requests	AMI ID	Region: 🛃 US En	CHOOSE AN AM	i aws.amazon.com	AWS   Products   Develop	pers Community	Support   Account			Welcome, Jam	is Taylor Settings Sign
Advanced Inst   * Sopathods   * Sopathods   * Sopathods   * Attra stock IPs   * Attra stock IPs   * Backit IPs   * Societty Groups   * Sopathods   * Sopathods   * Sopathods   * Attra stock IPs   * Societty Groups   * Sopathods   * Societty Groups   * Societty Groups  *	AMIa     Bundle Taska	ami-53ab45	+ EC2 Dashboard	Availability Zone:	Amazon EC2 Amazon MapRe	Elastic Amazon duce CloudFrom	nt					
Vilument <td>HASTIC PLOCK STORE</td> <td></td> <td>+ Instances</td> <td>Advanced Insta</td> <td>Navigation</td> <td>My Instances</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	HASTIC PLOCK STORE		+ Instances	Advanced Insta	Navigation	My Instances						
• Snaphole • Snaphole • Snaphole • Sacurity Groupe • Societ of United Tasks • Societ of	+ Volumes	Spot Requests     Here you can choose     enter data that will			Region: US East + B Laurch Instance Actors + Reserved Instances +							ifride Refresh Street
MATE       AATE       Kernel 10: Us       Unitance       AMI ID       Root Device Type       Status       Lifecycle       Public DNS       Security Groups       Instance         * Ebstic TP6       * Bundle Tasis       * AM Disk 1D: User       Monitoring:       Instance       Mill D       Root Device Type       Status       Lifecycle       Public DNS       Security Groups       Instance         * Security Groups       * Volumes       * Security Groups       * Mills       Instance       Mill D       Root Device Type       Type       Status       * Lifecycle       Public DNS       Security Groups       *         * Security Groups       * Security Groups       * Security Groups       * Mills       Security Groups       * Securits S	+ Snapshots	mapshots (MAGES		enter data biat with	+ EC2 Dashboard	Viewing: Running Inst.	ances 🚺	All Instance Types	•		10 0	1 to 1 of 1 Instances > )
Elastic IPs     Security Groups     Load Balancers     Status     Source I 2009, Amazem Wetr     Dupir a new      Dupir a new      Source I 2009, Amazem Wetr     Security Groups	NETWORKING & SECURE		* AMEs	Kernel ID:	INSTANCES	Instance	AMI ID	Root Device Type	Туре	Status • Lifecycle	Public DNS	Security Groups
Security Groups     Society of model     Starphiles     Subsection     Security Groups     Security G	+ Elastic IPs		• Bundle Tasks	RAM Disk ID: U	+ Instances	1453b742e	ami-ed03ed84	ebs	m1 large	<ul> <li>running normal</li> </ul>	ec2-184-73-52-147.comp	ute-1 galaxy-web, default
• Load Balancers • Snapphas • Snapphas • Snapphas • Snapphas • Snapphas • Security Groups • Security Groups • Load Balancers • Dodi - 2000, Amazem Webt Serie • Dodi - 2000,	Security Groups     New Pairs		ELASTIC BLOCK STORE	Monitoring:	Spot Requests							
Status <ul> <li></li></ul>	· Load Balancers		+ Snapshots	User Data: jan	IMAGES							
Status Status Status Security Groups Status Security Groups Status Social - 2008, Amazim Web Serve Doptry a menu Elastic IPs Social - 2008, Amazim Web Serve Doptry a menu Elastic IPs Security Groups <	a second de la seconda de la		HETWORKING & SECUR		AMIs     Bundle Teste							
Status       * Security Groups       > Volumes         * Kay Pairs       * Load Balancers       > Volumes         * Dody a terms       * Back       > Eastic IPs         * Dody a terms       * Back       > Kay Pairs         * Dody a terms       * Dody, Amazon Web Serve       > Eastic IPs         * Dody a terms       * Dody, Amazon Web Serve       > Kay Pairs         * Dody a terms       * Dody, Amazon Web Serve       > Kay Pairs         * Dody a terms       * Load Balancers       * Kay Pairs         * Load Balancers       * Load Balancers       > Select an Instance above		-	+ Elastic IPs									
Status          • Koy Pairs           • Snapshots             • 2008 - 2009, Amazen Wetz           • Back           • Snapshots             • Back           • Back           • Cod Balancers             • Back           • Back           • Cod Balancers             • Dobigs a memu           • Dobigs a memu           • Snapshots			+ Security Groups		Volumes							
© 2008 · 2009, Amazon Web 3       • Back       • Etastic IPs         • Back       • Back       • Security Groups         • Dosiny a memu       • Dosiny a memu       • Loed Balancers		Status	Key Pairs     Load Balancers	_	> Snapshots							
© 2009, Amazon Wett         (objer a menu         • Back         • Back         © 2008 - 3000, Amazon Wett Serve         © 2008 - 3000, Amazon Wett Serve         Disting a menu             Disting a menu             Back             0 EC2 Instances selected             Select an Instance above		Contractor		_	NETWORKING & SECURITY							
Back     > Security Groups       Display a menu     > 2006 - 2006, Amazaan Web Server       Display a menu     > Security Groups       Display a menu     > Security Groups	@ 2008 - 3	2009, Amazon Web I			> Elastic IPs							
Back      Key Pairs      O EC2 Instances selected      Load Balancers     Select an Instance above	Olity A menu			-	Security Groups							
(D 2008 – 3000), Avusatan Web Server Depisy a menu				Back	Key Pairs     Load Balancers	Select an instances sel	ected					
© 2008 – 3009, Amazim Web Servic Doplay a menu					A DECEMBER OF A DECEMBER OF A							
Doptry a menu			ip 2006 - 3	1009, Amazon Web Salvis								
			Display a menu									

0		Galaxy Cloud			
+ http://www.second.com/	//ec2-174-129-103-	-83.compute-1.amazonaws.com/cloud	0	Q+ Google	
	Galaxy Clou Welcome to Galaxy this is your first tim configured, default s which jobs are run,	dman Console Cloudman. This application will allow you to manage this cl e running this cluster, you will need to select an initial data services will start and you will be add and remove additiona	oud and the services volume size. Once ti l services as well as 'v	provided within. If he data store is worker' nodes on	
	Terminal	Initial Cluster Configuration			
	Status Cluster name Disk status: Worker status Service status	Welcome to Galaxy Cloudman. This application will allo cluster and the services provided within. To get started cluster you'd like to work with and specify the size of ye storage, if any. Start a full Galaxy Cluster. Specify initial storage size 100 GB OK Show more startup options	w you to manage this I, choose the type of our persistent data e (in Gigabytes)		
		(Start Cluster)			

•





Can use like any other Galaxy instance, with additional compute nodes acquired and released (*automatically*) in response to usage

Visualization (beta)



Integration with existing popular browsers, including mirrors and local browsers

- Galaxy	Analyze Data	Workflow Sha	red Data L	ab Visualization	Admin	Help	User			
Unnamed (lambda_NEB3011)	lan	bda_NEB3011	•	9,456 - 11,821	Θ	Ð		Add Tracks	Save	Close
	10,000				1	1,000				
Aligned reads 👻										Auto -
278	263	184 98		15			286 - 2	71 194 147 269 128		125
		623								

	10,000	11,000
Display a menu		

### Visualizing aligned reads in trackster

🚾 Galaxy	Analyze Data Workflo	ow Shared Data	Visualization	Help User			
Visual Analytics Test (test_build)						Add Tracks	Save Close
	test_chromosome	•	0 - 650	Ø Ø			
0  100	200	300		400	500		600
Cufflinks on data 77: assembled transcripts + Cufflinks Max Intron Length [300000] Min Isoform Fraction [0.05] Pre MRNA Fraction [0.05] Min SAM Map Quality [0] (Run) CUFF.3.1			**				Auto 👻
Cufflinks, region=[test_chromosome:0-650], parameter CUFF.3.1	rs=[1, 0.05, 0.05, 0] 😽						Auto 🔻
Cufflinks, region=[test_chromosome:0-650], parameter CUFF.3.1	rs=[90, 0.05, 0.05, 0] 🔫						Auto 👻
Cufflinks, region=[test_chromosome:0-650], paramete	rs=[100, 0.05, 0.05, 0] 👻					_	Auto 🔻
Cufflinks, region=[test_chromosome:0-650], paramete	rs=[150, 0.05, 0.05, 0] 🔻		**				Auto 👻

Visualization integrated with tools: visual analytics in trackster

# **Publishing analysis**

# Sharing and publishing

Galaxy   Published Page   Winds					eld Splatter			
- Galaxy	Analyze Data	Workflow Data Lib	raries Lab	Admin	Help	User		
Published Pages   aun1	Windshield Splatter							
Windshield s A live supple	platter analy ment	ysis with t	he Gala	axy m	etag	enomic	pipelin	e:
SERGEI KOSAKOVSKY POND <sup>1,2</sup> , NEXRUTENKO <sup>1,3</sup> and THE GAL/	", SAMIR WADHAWAN <sup>1,6*</sup> , R XY TEAM <sup>1,</sup>	RANCESCA CHIÁRONONT	TE <sup>4</sup> , CURUPRASA	D ANANDA <sup>1</sup>	, WEN-YU	CHUNG <sup>1,3,2</sup> , JAN	MES TAYLOR <sup>1,1</sup> , AN	TON
Cirrespondence should addressed	14 SSE. 27. or 88.							
How to use this a	locument							- 1
This document is a live cop in the paper, so you can pl the two histories and one v these items into your Galax a Galaxy account (unless yo	y of supplementary mati ay with them by re-runn vorkflow found below. W y workspace and start u ou already have one) - a	trials for the manuscr ing, changing parame ou can view these iter sing them; click on th hassle-free procedur	igt, it provides ters, or even a ns by clicking e green plus t e where you a	access to applying the on their na o import an re only ask	the exact me to your i item. To ed for a us	analyses and w rown data. Sp and them. You import workflo ername and p	vorkflows discus ecifically, we pro a can also import aws you must gr assword.	sed vide t sate
This is the Galaxy history d	letailing the comparison	of our pipeline to ME	GAN:					- 1
Đ	Corr	Galaxy History   Gal sparison of Galaxy v	EXY VS MEGAT	N eline.			0.8	
This is the Galaxy history s section of the manuscript a	howing a generic analysi nd Figure 3A)	s of metagenomic dat	a. (This corre	sponds to t	he "A com	plete metagen	omic pipeline"	
	G	laxy History   metar	enomic analy	ala.			0.8	- 1
This is the Galaxy workflow the manuscript and Figure	for generic analysis of ( 38):	metagenomic data. (T	his correspond	fs to the "A	complete	metagenomic	pipeline" section	of
٠	Gal	axy Workflow   metr	genomic ana	baia				- 1
Supplemental An	alysis							
Comparison between	Galaxy pipeline and	Megan						-
Loading "mitp //main.g2.bx.psu	edulation Lip/windshield-sp	latter", sompleted 5 of 6 if	19994					

- All analysis components

   (datasets, histories, workflows)
   can be *shared* among Galaxy
   users and *published*
- Pages and annotation allow analaysis to be augmented with textual content and provided in the form of an integrated document

# Sharing and publishing







Try it now: http://usegalaxy.org Develop and deploy: http://getgalaxy.org