# Galaxy: The next decade



@galaxyproject / #usegalaxy

#### Three major concerns

Accessibility: Making use of large-scale data requires complex computational resources and methods. Can all researchers access these approaches? How can we make these methods available to *everyone* 

Transparency: Is it possible to communicate analyses and results in ways that are both easy to understand and provide all of the essential details

Reproducibility: Can analyses be precisely reproduced, to facilitate rigorous validation and peer review, and ease reuse?

# Reproducibility

#### **JWST Pipeline and Data Products**

Follow the links below to learn how to install and use the JWST pipeline on your computer and to learn more about the services at MAST for downloading data. JDox Z contains detailed documentation on the JWST data files produced by all stages of the pipeline. Also, simulated data that have been processed by the pipeline are available to familiarize yourself with the JWST data formats.



Nearly all science is computational and should be reproducible

increasingly

important

#### Leakage and the Reproducibility Crisis in ML-based Science

We argue that there is a reproducibility crisis in ML-based science. We compile evidence of this crisis across fields, identify data leakage as a pervasive cause of reproducibility failures, conduct our own reproducibility investigations using in-depth code-review, and propose a solution.





July 28 online workshop

### Collaboration

- Viewing exact details of an analysis is critical for understanding
- Building on prior work is a key component of science





### Accessibility

- Make complex analyses widely available via the Web and GUI
- Increase equality for computational tools + workflows

••• • • < >		Û	+ 88
📮 Galaxy / microGalaxy	🛠 Workflow Visualize Shared Data - Help - Login or Register 🞓 🌲 🏢	Us	sing 0 bytes
Tools ∷≣	Welcome to microGalaxy!	History	S \$
search tools		search datasets	00
1 Upload Data	microGalaxy is the place to go for anything microbiology in Galaxyl Whether you are analysing microBalaxy microbiome samples or bacterial isolates, long reads or short, shotgun or 16S, genomics transcriptomics or proteomics, this is the place to be!	Unnamed history (empty)	
Get Data			
Collection Operations	Content	This history is empty. You	can
GENERAL TEXT TOOLS	1. Welcome to microGalaxy!	load your own data or ge	t data
Text Manipulation	2. Join the µGalaxy Community	from an external source	
Filter and Sort	3. Workhows 1. Microbiome		
Join, Subtract and Group	2. Bacterial Isolates		
GENOMIC FILE MANIPULATION	5. Projects		
Convert Formats	6. Resources		
FASTA/FASTQ	Join the µGalaxy Community		
SAM/BAM	Anybody interested in microbiology in Galaxy is welcome to join our microGalaxy interest group! Everybody is		
BED	Welcome!		
COMMON GENOMICS TOOLS	Join our quarterly meetings! Agenda and minutes		
Operate on Genomic Intervals	Workflows		
GENOMICS ANALYSIS	Below are a list of curated Galaxy workflows for different kinds of microbial analysis. Many of these are		
^t-t	accompanied by comprehensive GTN Tutorials that will guide you through the analysis step by step.		>

History	Ø + 🗉	•
search datasets	0	0
TMA1-005_CycIF_ImmunePanel_Galaxy-MTI		
13 shown, 4422 deleted, 9961 hidden		
1.61 TB	2	
9908: Vitessee Visualization on collection 6210, collection 9819, and co a list with 88 items	liection 272	×
9819: Rename OME-TIFF Channels on collection 6120 a list with 88 items		×
6210: Initial (uncombined) phenotyping with Scimap a list with 88 items		×
6209: Single Cell Phenotyping: GMM-plots a nested list with 88 items		×
6120: Convert image on collection 4 a list with 88 items		×
6031: Convert McMicro Output to Anndata on collection 451 a list with 88 items		×
6030: TMA005_phenotype_20220606.csv	œ,	×
451: CellMaskQuant a list with 88 items		×
272: Mesmer on collection 4: Mask a list with 88 items		×
6: UNetCoreograph on data 1: TMA Map	æ /	×
4: UNetCoreograph on data 1: Images a list with 88 items		×
2: TMA1_005_immune_markers.csv	•/	×
1: OHSU_TMA1_005.ome.tif	•/	×



## Accessibility is power





# Fighting analytical inequality

# Galaxy makes analyses accessible

#### Past

...by early 2000's the big data in biology was genomic sequences and alignments. Penn State was central in developing alignment tools



Webb Miller



Ross Hardison

The basic question in the *early* 2000's was: What is aligned to what and does it overlap with anything interesting?

#### 2003

#### **Resources**

#### GALA, a Database for Genomic Sequence Alignments and Annotations

# Belinda Giardine,<sup>1</sup> Laura Elnitski,<sup>1,2</sup> Cathy Riemer,<sup>1</sup> Izabela Makalowska,<sup>4</sup> Scott Schwartz,<sup>1</sup> Webb Miller,<sup>1,3,4</sup> and Ross C. Hardison<sup>2,4,5</sup>

Departments of <sup>1</sup>Computer Science and Engineering, <sup>2</sup>Biochemistry and Molecular Biology, <sup>3</sup>Biology, and <sup>4</sup>Huck Institute for Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

We have developed a relational database to contain whole genome sequence alignments between human and mouse with extensive annotations of the human sequence. Complex queries are supported on recorded features, both directly and on proximity among them. Searches can reveal a wide variety of relationships, such as finding all genes expressed in a designated tissue that have a highly conserved noncoding sequence 5' to the start site. Other examples are finding single nucleotide polymorphisms that occur in conserved noncoding regions upstream of genes and identifying CpG islands that overlap the 5' ends of divergently transcribed genes. The database is available online at http://globin.cse.psu.edu/ and http://bio.cse.psu.edu/.





GALA enabled query annotation information from the human genome, alongside alignments with the mouse genome, integrated with the UCSC browser, and allowed building up set queries using the results of previous queries (the birth of the History system)

#### Can GALA be extended to other analyses?

#### Can GALAXY be extended to other analyses?





#### Galaxy as a PERL script (~50,000 lines)

00	0	Table Browser	Galaxy: History Page
4.1	G A A C	🕴 🖶 http://hgwdev-giardine.cse. 📿 🖉 🖓 Gr Google	A A G + Ontp://www.bx.psu.edu/tpi-O Coops
Tat	ble Browser		Portal I History I About Galaxy I Example queries I FAQ I Contact us
Une cale con concernent conc	this program to ge utate intersections end by a track. See thois in this form. The ted period. de: <u>Vortexen</u> @ de: <u>Vortexen</u> @ de: <u>Vortexen</u> @ de: <u>Vortexen</u> @ de	the data associated with a track in text format, to between tracks, and to retrieve DNA sequence Using the Table Browser Page is still available for a genome: turnat @ assembly: May 2004 @ vestimate franks @ track: ccs: @ exother table strema: ENCODE @ position urg[ cessions): [user ter: usional text (leave blank to keep output in plain text _ gzip compressed	History Page Senome:            Genome:         Immer         Assembly:         Impl://mail.com/           Your Previous Queries:              • 1: codsGene ods (limit to chr112110531-2116578) [3 regions]            Action to Perform:              • 6: colput, [mailers 1 query]            Perform operations like intersoction, etc. [requires 1 or more queries]            Perform operations like intersoction, etc. [requires 1 or more queries]            Perform operations like intersoction, etc. [requires 1 or more queries]            Perform operations like intersoction, etc. [requires 1 or more queries]            Perform operations like intersoction, etc. [requires 1 or more queries]            Immer           Immer
0.0	0	City Courter	1000
4.1	QAAC	+ Or Goodh	A A C + Ohttp://www.bx.psu.adu/tgi-O Coods
Que	ery Operatio	ns	Portal   History   About Galaxy   Example queries   EAQ   Contact us
	Assembly: Hum	an, hg17	History Page
Selec	ted Queries: 1: ccdsGene cd: 2: snp (limit to cl	(limit to chr11:2110531-2116578) (3 regiona) hr11:2110531-2116578) (40 regiona)	Genome: Human (8) Assembly: hg17: May 2004 (8)
Opera	Union: Help     Union:     Intersection:	return whole regions from query #1 (2) , where overlap >=	Tour Previous Justness     ① 1: codsGene cds (limit to chr11:2110531-2116578) [3 regions]     ② 2: snp (limit to chr11:2110531-2116578) [40 regions]     ③ 3: regions from query 1 that intersect regions from query 2 [2 regions]
	Subtraction:     Complement:     Restrict     region size:     Proximity:	(inquires 1 query) (inquires 1 query)	Action to Perform: © Get output ○ Perform operations like intersection, etc. ○ Run analysis tools ○ Delete selected queries
0	Clustera:	[indires 1 divid)	(hore) Co (tatvash



# James Taylor joins the project

#### Early pythonic Galaxy circa 2005

Basic Statistics: Histogram: histogram.tool Scatter Plot: scatter.tool Filtering: filtering.tool Correlation: correlation.tool Region Length: region length.tool Score distribution: scoreGraph.tool Operations: Complement: complement.tool Restrict: restrict.tool Merge overlapping regions: merge.tool Cluster: cluster.tool Union: union.tool Intersect: intersect.tool Subtract: subtract.tool Proximity: proximity.tool Join Lists: joinLists.tool Vicinity: vicinity.tool Join Same Coordinates Region: joinSameCoor.tool Sequence Tools: Extract sequences: fasta-subseq-wrapper.tool Extract blastZ alignments: extractAxt wrapper.tool Data Sources: UCSC query: ucsc.tool Genbank: genbank.tool Encode DB: encodedb.tool Featured datasets: import.tool Format Converters: BED and xBED converter: bed convert.tool

00	History Page	
4 - 1 A A	C http://nekrut.bx.psu.edu:90( ^ Q+ Google	>>>
CC ORF BIZseq Pub History Page	Med Meetings * Funding * Genomics * Building *	cms ≫
Data		
Query	Genome	
cefresh		
Perform Action	15	
- Basic Statistics		
- Histogram	- builds histogram for any numeric column	[hele]
Scatter Plot	- builds scatterplot for any numeric column	(help]
• Filtering	- fitters data on any column using simple conditional expressions	[help]
• Correlation	- computes Pearson's correlation between any two numerical columns	[hele]
· Region Length	- computes length of bed intervals	[help]
• Score distribution	- display the score distribution of a selected score name	[help]
- Operations		
- Sequence Tools		
- Data Sources		
- Format Converte	18	
- Upload Data		



#### Dan Blankenberg and the letter *P*

#### **Connection with UCSC**



••• <> ••	5		a genome.uc	sc.edu		Ċ	₫ <b>₽</b> +
Genomes	Genome Browser	Tools Mirrors	Downloads	My Data	Help	About Us	
Table Browser							
Use this program to covered by a track. information and sar complex queries, yo send the data to GF usage restrictions a clade: Mammal group: Genes and G	retrieve the data assoc For help in using this a pple queries, and the O u may want to use <u>Gal</u> <u>EAT</u> . Send data to <u>Ger</u> ssociated with these dat <b>genome:</b> Human ene Predictions <b>track</b>	iated with a track in oplication see <u>Using</u> penHelix Table Brow axy or our <u>public My</u> <u>nomeSpace</u> for use v ta. All tables can be \$ (GENCODE v24	text format, to cal the Table Browse vser tutorial for a r SQL server. To ex with diverse comp downloaded in th assembly: De add cu	culate interse or for a descri- narrated press camine the bi- utational too leir entirety fr c. 2013 (GRCh38 stom tracks) t	ections be iption of th entation o ological fu ls. Refer to om the <u>Se</u> (hg38) rack hubs	tween tracks, and e controls in this f the software fea inction of your se to the <u>Credits</u> page equence and Ann	d to retrieve DNA sequence form, the <u>User's Guide</u> for general atures and usage. For more t through annotation enrichments, e for the list of contributors and <u>otation Downloads</u> page.
table: knownGene	describe table so	hema					
region: o genome	oposition chr1:1110283	-11267747 lo	okup define regions	5			
filter: create intersection: crea	e	upload list					
correlation: create	) D. however extensible data	A Soud				omoSpace	
output file: file type returned:	<ul> <li>plain text _ gzip co</li> </ul>	(leave blank to kee	ep output in brows	ser)		IOMESpace	
get output summary	/statistics t settings (including cus	stom tracks), <u>click he</u>	ere.				

#### 2005: NGS begins: Il pleut des cordes

#### ARTICLES

#### Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies<sup>1\*</sup>, Michael Egholm<sup>1\*</sup>, William E. Altman<sup>1</sup>, Said Attiya<sup>1</sup>, Joel S. Bader<sup>1</sup>, Lisa A. Bemben<sup>1</sup>, Jan Berka<sup>1</sup>, Michael S. Braverman<sup>1</sup>, Yi-Ju Chen<sup>1</sup>, Zhoutao Chen<sup>1</sup>, Scott B. Dewell<sup>1</sup>, Lei Du<sup>1</sup>, Joseph M. Fierro<sup>1</sup>, Xavier V. Gomes<sup>1</sup>, Brian C. Godwin<sup>1</sup>, Wen He<sup>1</sup>, Scott Helgesen<sup>1</sup>, Chun He Ho<sup>1</sup>, Gerard P. Irzyk<sup>1</sup>, Szilveszter C. Jando<sup>1</sup>, Maria L. I. Alenquer<sup>1</sup>, Thomas P. Jarvie<sup>1</sup>, Kshama B. Jirage<sup>1</sup>, Jong-Bum Kim<sup>1</sup>, James R. Knight<sup>1</sup>, Janna R. Lanza<sup>1</sup>, John H. Leamon<sup>1</sup>, Steven M. Lefkowitz<sup>1</sup>, Ming Lei<sup>1</sup>, Jing Li<sup>1</sup>, Kenton L. Lohman<sup>1</sup>, Hong Lu<sup>1</sup>, Vinod B. Makhijani<sup>1</sup>, Keith E. McDade<sup>1</sup>, Michael P. McKenna<sup>1</sup>, Eugene W. Myers<sup>2</sup>, Elizabeth Nickerson<sup>1</sup>, John R. Nobile<sup>1</sup>, Ramona Plant<sup>1</sup>, Bernard P. Puc<sup>1</sup>, Michael T. Ronan<sup>1</sup>, George T. Roth<sup>1</sup>, Gary J. Sarkis<sup>1</sup>, Jan Fredrik Simons<sup>1</sup>, John W. Simpson<sup>1</sup>, Maithreyan Srinivasan<sup>1</sup>, Karrie R. Tartaro<sup>1</sup>, Alexander Tomasz<sup>3</sup>, Kari A. Vogt<sup>1</sup>, Greg A. Volkmer<sup>1</sup>, Shally H. Wang<sup>1</sup>, Yong Wang<sup>1</sup>, Michael P. Weiner<sup>4</sup>, Pengguang Yu<sup>1</sup>, Richard F. Begley<sup>1</sup> & Jonathan M. Rothberg<sup>1</sup>

#### The basic question in the *late* 2000's becomes: What would happen if I sequence the s\*\*\*\*t out of anything?\*

\*For metagenomic studies this was, in fact, precisely the question asked.

Calaxy     Calaxy   Calaxy Cal	000		Galaxy		0				
Galaxy	🔶 🧼 🥪 🔕	000		Galaxy		0			
Calaxy       - Galaxy       - Control of a low of the second low of the	Galaxy	🤹 🖗	💿 🏠 🎯 http://main.g2.bx.psu	u.edu/	▼ 🕨 (G• Google	۵) لا			
Done Query2: gen8 tunor-supressor gen8 tunor-supressor	Data Edit Filter and Sort Format Sequences and Alignments Scores Interval Functions Statistics Graph EMBOSS PHYLIP Cluster regions of a single Query Cluster and Merge regions within specified distance of each other Complement a single Query Coverage density of the regions of two queries Difference find non-overlapping segments between two Queries Intersect return overlapping segments of two Queries	Get Data Usional File from v USSC Main table 1 Get Microbial Data Get ENCODE Data ENCODE Tools Edit Queries Filter, Sort, Join and Convert Formats Fetch Sequences a Get Genomic Score Operate on Genom Statistics Graph Data EMBOSS HYPHY HYPHY	Convert Formats Extract Features Pattern-Matching Fetch Alignments Get Genomic Scores Operate on Genomic Intervals Statistics Graph Oisplay Data Regional Variation Fasta manipulation Regional Variation For	Join two Queries Join: 20. Subtract on data 17 and data 1 20. Subtract on data 17 and data 1 using column: c1 22. Build custom track on data 19 and column: c1 xeep lines of first input that do not jon No Xeep lines of first input that are inco No Xeep lines of first input that are inco No Xeep lines of first input that are inco No Xeep lines of first input that are inco No This tool will attempt to reuse the n click on the "edit attributes" link of the 1 Xexcute Xexute Xexute Xexute Xexute Xexute Xexute Xexute Xexute	Info: report bugs   wiki   scree in with second input: mplete: metadata from your first input. To change m history item generated by this tool. ise Text Manipulation->Convert mon field. An empty string ("") is not a valid id at do not join with your second input. ber. For example, 3 refers to the 3rd column of	encasts   biog Logger	Histor refrest 20: 5 and 1 19: 1 Hum (enc) 17: 11 Hum most	s@bx.psu.edu: man ry h   collapse.all build custom track ata 19 and data 20 Subtract on data 1 data 19 JCSC Main on an: knownGene ode) JCSC Main on an: Conserved28way	Dept   logout         Options         • Ø %         2 • Ø %         • Ø

# 



#### 2010- The modern Galaxy era

#### Meanwhile...



The great flood of 2012



#### Stability and sustainability crisis!

# ... in which Galaxy is saved by *public* infrastructure



- Australia - nector

#### Present
## usegalaxy.\*





## COMMITS









github.com/galaxyproject/galaxy

#### **GTN** Statistics



	5		
<b>28</b> Topics	269 Tutorials	263 Contributors	<b>7.0</b> Years
17 Scie	entific Topics	Growing C	Community rs over time
Assembly	13	300	
Climate	7		
Computational chemistry	8	250	1
Ecology	8	200	
Epigenetics	9		1
Genome Annotation	11	200	
Imaging	4		and the second
Introduction to Galaxy Analyses	10		
Metabolomics	7	150	
Metagenomics	8		
Proteomics	27	100 20000	
Sequence analysis	3	100	
Statistics and machine learning	16		
Synthetic Biology	3	50	
Transcriptomics	34		
Variant Analysis	12		
Visualisation	3		
0	5 10 15 20 25 30 35	47 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	

## Components

# Interaction with Galaxy





History	+	₽	•
search datasets		*	×
Gene analysis: cdkn2a_v2_100	)		1
€ 770 kB	<b>Q</b> 22	<b>R</b> 154	, S
<b>Z</b> ::			٠
176 : Collapse Collection a 175, data 174, and oth	on on da ners	t o	/ =
151 : Intersect on collection 126 a list with 24 bed datasets	ction 2 a	nd c	/ 1
<b>126 : Cut on collection</b> a list with 24 <b>bed</b> datasets	54		/ #
125 : Cut on data 124		0	/ =
124 : Join two Datasets	on data	9 0	/ 1
			>

History			+	#	•
search datas	sets			*	×
Filter by name:					
any name					
Filter by extension	on:				
any extensio	n				
Filter by tag:					
any tag					
Filter by state:					
any state					
Filter by item ind	ex:				
index greate	r	index lo	ower		
Filter by creation	time:			11	
created afte	r 🗖	created	d bef	or	
Deleted:					
Any Yes	No				
Visible:					
Any Yes	No				
Q Search	C Cance	el			
					>

Fast Infinite scrolling Advanced filtering Auto-input tracking Foundation for tree view Fast (!) Live today

### Collections!







**Collection Operations** 

**GENERAL TEXT TOOLS** 

**Text Manipulation** 

Filter and Sort

- By EDAM
- By BIO.TOOL
- By Training
- By IWC
- By ...

## Do we actually need permanently visible tool panel?

Major work on the tool panel in 22.09!



### **IWC - Intergalactic Workflow Commission**

Galaxy Workflow Tests for push and PR passing chat on gitter

The IWC maintains high-quality Galax

Workflows are categorized in the workflows directory, ar

All workflows are reviewed and tested before publication follow best practices and are versioned using github rele as:

- License
- Author
- Institutes

Additionally the IWC will collect further best practices, ti high-quality Galaxy workflows.

Dockstore Q Search A Organizations	s 💿 About 🧃	〕Docs     Forum	b
. Organizations / The Galaxy Intergalactic Workflow Comissio	on		
The Galaxy Intergalacti	ic Workflov	W Comission	* 26
Galaxy			/ Edit
Collections 5 Kembers 6 C Updates	10	About the Organization	/ Edit
naise and the second se		GD https://github.com/galaxyproject/iwc	
Variant Calling		IWC - Intergalactic Workflow	
1 Workflow	/ Edit	Commission	
Vertebrates Genome Project Workflows Galaxy Workflows for the Vertebrates Genome Project Workflow	ows	Reproducibility is important. Our goals are to: • foster workflow use • incorporate versioning	
2 Workflows	/ Edit	capture more metadata: (names, versions, authors, use cases, etc.)     help scientists find workflows!	
SARS-COV-2 Workflows			
7 Workflows	/ Edit		
Data Fetching Workflows Workflows that fetch data from remote resources			
1 Workflow	/ Edit		
Computational Chemistry Workflows Workflows for Computational Chemistry			
4 Workflows	/ Edit		

/ Edit



••• • • < >		0		🔒 usegalaxy.c	rg	9 <b>4</b>			<b>(</b> ) + 88
🚍 Galaxy			📣 Workflow	Visualize Shared Data <del>-</del>	Help 👻 User 👻 🛕	*			Using 2%
Galaxy has recently been updated to a	Download	fi Type to S	Search	×					n error (red) datasets, or email galaxy-
bugs.	Regular	FIP Dire	ectory Galaxy U	Jser's FIP Directory			_		
Tools		Library L	User Galaxy's	user library import directory	1				/ + ≓ *
search tools		Market Dire	pobox Your Dro	nhov files - configure an acc	sess taken via the user n	references			datasets × ×
1. Upload Data		Files							analysis: 🛛 🖌
Get Data	📝 New	F EBI FTP	server Europea	n Bioinformatic Institute FTP	server			۵	!a_v2_100
Send Data		SARS-C	oV-2 SARS-C	oV-2 variants and consensus	sequences				🛛 22 🔌 154 💋
Collection Operations		NCBI FT	P server NCBI FT	P server					¢
GENERAL TEXT TOOLS			BL FTP ENSEME	BL FTP server					lapse Collection on dat 🛛 🖉 👕
Text Manipulation		server	A.J. A		414/0				ta 174, and others
Filter and Sort		■ Genome	Ark Access 1	o Genome Ark open data on	AWS.	intion including SND	~		rsect on collection 2 and c 🧪 👕
Join, Subtract and Group		1000 06	enomes Accession etructure	al variants and their hanlotw	ne context	ation, including SNP	5,		74 hed datasets
Datamash						Cancel	Ok		
GENOMIC FILE MANIPULATION									on collection 54 🧪 🗃
FASTA/FASTQ		L.	Choose local files	Choose remote files	Paste/Fetch data	Start Pause	Reset	Close	24 bed datasets
FASTQ Quality Control		processing							on data 124 🛛 💿 🖊 👕
SAM/BAM						_		124 - 10	ain two Datacate an data 🖉 🖌 🖷
BED		+ ttt	۲	22 days ago				124 . Ju	d data 82
VCF/BCF	щ	+ Calling variants in	Variant X	11 months ago	¢			123 · Co	ollanse Collection on dat 💿 🖌 🗃 >

Download	l from web or u	pload from	disk				
Regular	Composite	Collection	Rule-based				
		You added 2	file(s) to the queue. A	Add more files or click 'Sta	rt' to proceed.		
	Name	Size	Туре	Genome	Settings	Status	
🕞 fAci	Rut3.mat.cur.20	<b>466.5</b> MB	Auto-de 🔻 Q	unspecified (?)	\$		创
► mH	omSap3.pat.cur.	<b>786</b> MB	Auto-de v Q	Unspecified (?)	ad configuration onvert spaces to t se POSIX standard efer dataset resolu	abs d ution	Û
	Тур	e (set all):	Auto-detect 💌	Q Genome (set all):	unspecified (?)	v	
	Choo	se local files	Choose remote	files Paste/Fetch da	ata Start F	Pause Reset	Close

### After 15 years, a new color added!









#### Global Alliance for Genomics & Health

#### **Tool Registry Service (TRS)**



#### Standard API for exchanging tools and workflows to analyze, read, and manipulate genomic data

Status: Supported



#### Data Repository Service (DRS)

Allows data consumers to access datasets regardless of the repository in which they are stored or managed

#### Status: PR Open



#### Beacon

Web-accessible service that users may query for information about a specific genomic allele

Status: Prototype done, PR coming

#### Task Execution Service (TES)



Schema and API for describing batch execution tasks with files, containers, and commands to run

Status: PR Open

#### Workflow Execution Service (WES)



Lets users run a single workflow on multiple different platforms, clouds, and environments

**Status: In discussion** 

#### Passports



Standards for communicating the data access authorizations that a user has based on role, affiliation, or access status.

Status: In discussion; Supported in AnVIL

## APIs enable automatic QC, deployment, and creativity!

Galaxy ecosystem is increasingly about services and less about raw software



# The ad hoc analytics







Calaxy         Worker         Younget         Here         United           I classes	••• •	Ø			⊜ usegala	axy.org	ිකු උ		① + 器	
Calcy base recently base vectored as par-evectored base (2,000, par-you regiver theorem as a low calcy base recently all approximation of the state of the s	🚍 Galaxy		😭 Work	flow Visualize	Shared Data	🝷 Help 👻 User 👻 🐥	a III		Using 2%	
Tools         Dim         Start         End         Name         Score         Stard         ThickStart         Thi           search Lools         Mardleot         scarlfold_B         19350674         19350674         1         search Ldatasets         1           Lupleod Data         Mardleot         scarlfold_B         19350674         1         +         search Ldatasets         Centom         Search Ldatasets         Centom	I Galaxy has recently been updated to a pre-release bugs.	of the next Galaxy st	able release (22.05),	and you might enc	ounter problems a	s we iron out the wrinkles. Please	report any issues	you encounter using t	ne bug icon on error (red) datasets, or email galaxy-	
search tools         Newlored         scaffold, 8         19350076         scaffold, 8         1935076         sc	Tools 🗠 🔹	Chrom	Start	End	Name	Score	Strand	ThickStart Thic	History + ≓ -	
Detected         Buble(Gec)         scatified, 8         19350579         19350588         scatified, 8         1935527         1935522         scatified, 8         1935527         1935522         scatified, 8         1935527         1935528         scatified, 8         1935527         1935528         scatified, 8         1935528	search tools	bMelGeo1	scaffold_8	19350141	19350576	scaffold_8_ORF.56032	1 -	+	search datasets	
± Upload Data       EMAIGENT       scaffold, 8       1935247       1935242       scaffold, 8       07565051       1       +         Get Data       bMelGent       scaffold, 8       1935042       1935043       Scaffold, 8       0756721221       2       +         Send Data       bMelGent       scaffold, 8       1935042       1935043       Control 1       2       +         Collection Operations       bMelGent       scaffold, 8       1935043       076122220       2       +       0       170 : Collaps Collection on dat       •       •         Join, Subtract and Group       bMelGent       scaffold, 8       1935049       1935049       076122209       2       +       0       170 : Collaps Collection on dat       •       •       0         Join, Subtract and Group       bMelGent       scaffold, 8       1935207       1935241       19352639       caffold, 8       07636852       3       +         Join, Subtract and Group       bMelGent       scaffold, 8       1935207       1935241       caffold, 8       07659033       1       -         FASTA/FASTQ       bMelGent       scaffold, 8       1935207       1935243       caffold, 8       07659033       1       -	search tools	bMelGeo1	scaffold_8	19350579	19350858	scaffold_8_ORF.56033	1 -	+	search datasets	
DMMEGeol       scarfield_8       1936274       1936272       2       4       100 </td <td>1 Upload Data</td> <td>bMelGeo1</td> <td>scaffold_8</td> <td>19355247</td> <td>19355622</td> <td>scaffold_8_ORF.56051</td> <td>1 -</td> <td>+</td> <td>Gene analysis:</td> <td></td>	1 Upload Data	bMelGeo1	scaffold_8	19355247	19355622	scaffold_8_ORF.56051	1 -	+	Gene analysis:	
Get Data       DMM(leent)       scattrol_8       19350598       scattrol_8_0.0F:2/221       2       +         Send Data       bMe(leent)       scattrol_8       19350546       19355565       scattrol_8_0.0F:2/227       2       +         Collection Operations       bMe(leent)       scattrol_8       19362709       1936280       scattrol_8_0.0F:2/2250       2       +       E       C         General_TEXT TOOLS       bMe(leent)       scattrol_8       19363579       scattrol_8_0.0F:2/2250       2       +       T/6: Collaps Collection on dat ● / ■       a         Join, Subtract and Group       bMe(leent)       scattrol_8       19362309       scattrol_8_0.0F:2/3226       3       +         Join, Subtract and Group       bMe(leent)       scattrol_8       19362309       scattrol_8_0.0F:368561       3       +         Join, Subtract and Group       bMe(leent)       scattrol_8       19362309       scattrol_8_0.0F:368562       3       +         Datamash       bMe(leent)       scattrol_8       19362369       scattrol_8_0.0F:560503       -       -         FASTQ Quality Control       bMe(leent)       scattrol_8       1935564       19355645       1935646       19355717       1936346       19355717       1936346       1936277		bMelGeo1	scaffold_8	19362774	19363404	scattold_8_ORF.56074	1 -	+	cdkn2a v2 10	
Send Data       Deviceeri scartord_B       19360864       19360864       19360864       20360286       scartord_B_Q_RRF.212212       2       +         Collection Operations       bMelCeot       scartfold_B       19360386       scartfold_Q_Q_RRF.212249       2       +         GENERAL TEXT TOOLS       bMelCeot       scartfold_B       19360379       scartfold_B_Q.RF.212250       2       +       176 : Collapse Collection on dat       >       >       176 ; Collapse Collectin canot       >       176	Get Data	bMelGeo1	scaffold_8	19350421	19350598	scattold_8_ORF.212211	2 -	+		
DMR/Level         ScatfordB         19305/84	Send Data	bMelGeo1	scattold_8	19350604	19350988	scattold_8_ORF.212212	2 -	+	■770 kB 922 > 154 €	
Outcourse       Description       Station of Section of Secti	Collection Operations	DMelGeo1	scatfold_8	19355146	19355653	scarroid_s_ORF.212226	2 -	+		
General LEXT FOULS       Dumber of scarfiold, a fight 2008       To the collapse Collection on dit • > 1         Join, Subtract and Group       bMelGeo1       scaffold, 8       19362200       scaffold, 8, ORF.368650       3       +       784 regions         GENOMIC FLE MANIPULATION       bMelGeo1       scaffold, 8       19362204       scaffold, 8, ORF.569030       -1       -       B @ 0 C u ift ?       Create charts for each Genome/Chromosome combination       for group in data(data) genome / nnigue():		bMelGeo1	scattoid_8	19362709	10262196	scarfold_8_ORF.212249	2 .		•	
Text Manipulation       Defected       stantor 0       19300000       1       1       784 regions         Join, Subtract and Group       bMelGeo1       scaffold_8       19362264       scaffold_8, ORF.368562       3       +       784 regions         GENOMIC FILE MANIPULATION       bMelGeo1       scaffold_8       19362264       scaffold_8, ORF.569003       -1       -       -       B @ @ C @ # # ?       1       -         FASTA/FASTQ       bMelGeo1       scaffold_8       19350046       19350465       scaffold_8, ORF.724791       -2       -       bMelGeo1       scaffold_8       19350857       scaffold_8, ORF.724824       -2       -       bMelGeo1       scaffold_8       19350875       scaffold_8, ORF.724824       -2       -       bMelGeo1       scaffold_8       19350302       19355687       scaffold_8, ORF.724824       -2       -       <	GENERAL TEXT TOOLS	bMolGoo1	scarfold 9	10262190	10262570	scattold_0_ORF.212250	2 .		176 : Collapse Collection on dat 🐵 🌶 🍵	
Filter and Sort       bMelGeol       scarfold_B       19355041	Text Manipulation	bMelGeo1	scaffold 8	19350560	19350905	scaffold_8_ORF368524	3	+	a 175, data 174, and others	
Join, Subtract and Group         Joins, Subtract and Group         Joi	Filter and Sort	bMelGeo1	scaffold 8	19355207	19355411	scaffold 8 ORF 368538	3		•	
Joint Subtract and order       bbielGeof       scaffold_8       19362569       scaffold_8, 0RF.368562       3 + +         Datamash       bbielGeof       scaffold_8       19362569       scaffold_8, 0RF.368562       3 + +         GENOMIC FLE MANIPULATION       bbielGeof       scaffold_8       19362264       scaffold_8, 0RF.568030       -1       -         FASTA/FASTQ       bbielGeof       scaffold_8       19355045       19355045       scaffold_8, 0RF.569035       -1       -         FASTA/FASTQ       bbielGeof       scaffold_8       19355045       19355045       scaffold_8, 0RF.569068       -1       -       bbielGeof       scaffold_8       1935041       19350457       scaffold_8, 1935971       scaffold_8, 1935971       scaffold_8, 1935971       scaffold_8, 1935972       scaffold_8, 1935979       scaffold_8, 1	Ioin Subtract and Crown	bMelGeo1	scaffold 8	19362740	19362902	scaffold 8 ORF.368561	3	+	784 regions	
Datamash         bMelGeol         scaffold_8         19363329         19363464         scaffold_8,ORF.569030         -1         -           GENOMIC FILE MANIPULATION         bMelGeol         scaffold_8         1936277         19363264         scaffold_8,ORF.569031         -1         -           FASTA/FASTQ         bMelGeol         scaffold_8         19350355         scaffold_8,ORF.569053         -1         -           FASTA/FASTQ         bMelGeol         scaffold_8         19350464         scaffold_8,ORF.569053         -1         -           FASTQ Quality Control         bMelGeol         scaffold_8         19350455         scaffold_8,ORF.724791         -2         -         bMelGeol         scaffold_8         193504576         scaffold_8         19350471         10368376         scaffold_8         10350676         scaffold_8         1935041         1035077         10363444         scaffold_8         1935041         1035077         10363444         scaffold_8         1035077         10364457         scaffold_8         1035077         10364457         scaffold_8         1935041         1035077         10364457         scaffold_8         1035077         1036444         scaffold_8         1035077         1036444         scaffold_8         10350771         103677         1036444	Join, Subtract and Group	bMelGeo1	scaffold 8	19362908	19363589	scaffold 8 ORF.368562	3 -	+	format bed, o tabase ?	
GENOMIC FILE MANIPULATION         bMelGeo1         scaffold_B         1936277         19363254         scaffold_B_ORF.569031         -1         -         B @ O C u h ?         Careas charts         Careas charts <th< td=""><td>Datamash</td><td>bMelGeo1</td><td>scaffold 8</td><td>19363329</td><td>19363464</td><td>scaffold 8 ORF.569030</td><td>-1 -</td><td>-</td><td></td><td></td></th<>	Datamash	bMelGeo1	scaffold 8	19363329	19363464	scaffold 8 ORF.569030	-1 -	-		
FASTA/FASTQ         bMelGeo1         scaffold_8         19355085         19355649         scaffold_8_ORF.569053         -1         -         1. Chron 2. Start 3. End 4         5           FASTQ Quality Control         bMelGeo1         scaffold_8         19350405         scaffold_8, ORF.569058         -1         -         bMelGeo1         scaffold_8         19350407         scaffold_8, ORF.569058         -1         -         bMelGeo1         scaffold_8         19350407         scaffold_8, ORF.724791         -2         -         bMelGeo1         scaffold_8         19350507         scaffold_8, ORF.724791         -2         -         bMelGeo1         scaffold_8         19350507         scaffold_8, ORF.724791         -2         -         bMelGeo1         scaffold_8         19350507         scaffold_8, ORF.724809         -2         -         bMelGeo1         scaffold_8         19350507         scaffold_8, ORF.724809         -2         -         bMelGeo1         scaffold_8         19355077         scaffold_8, 19355277         sc	GENOMIC FILE MANIPULATION	bMelGeo1	scaffold_8	19362777	19363254	scaffold_8_ORF.569031	-1 -	-	B & O C ≝ ♣ ?	<pre>charts for each Genome/Chromosome combination charts = []</pre>
FASTQ Quality Control         bMelGeo1         scaffold_B         19350486         19351017         scaffold_B_ORF.569068         -1         -         bMelGeo1         scaffold_B         19350470         (juitable 1)         (juitable	FASTA/FASTQ	bMelGeo1	scaffold_8	19355085	19355493	scaffold_8_ORF.569053	-1 -	-	1.Chrom 2.Start 3.End 4 5	for energies in data cant unline thus ( michael ) ( improve ) unique ( ).
PASID Quality Control         bMelGeo1         scaffold_8         19362827         19363436         scaffold_8_0RF.724791         -2         -         bMelGeo1         scaffold_8         19362827         19383436         scaffold_8_0RF.724791         -2         -         bMelGeo1         scaffold_8         1937507         19383879         19383849         cold         col	FACTO Constant	bMelGeo1	scaffold_8	19350486	19351017	scaffold_8_ORF.569068	-1 -	-	bMelGeo1 scaffold_8 19350141 19350576 sca	for group in data[data['genome']==species]['genome_chr'].unique():
SAM/BAM         bMelGeo1         scaffold_8         19355657         scaffold_8/05567         scaffold_8/05667         scaffold_8/05667         scaffold_8/05667         scaffold_8/05667         scaffold_8/05667         scaffold_8/05667         scaffold_8/0567         scaffold_8/056767         scaffold_8/05676	FASTQ Quality Control	bMelGeo1	scaffold_8	19362827	19363436	scaffold_8_ORF.724791	-2 -	-	bMelGeo1 scaffold_8 19350579 19350858 sca	<pre>src = data[data['genome_chr']==group] charts append(chart(erc))</pre>
BED         bMelGeol         scaffold_8         19350302         19350857         scaffold_8_0RF.724824         -2         bMelGeol         scaffold_8         1936274         1936244         scaffold_8         19362790         19362450         scaffold_8_0RF.724824         -2         bMelGeol         scaffold_8         19362470         19362480         scaffold_8_0RF.724824         -2         -         bMelGeol         scaffold_8         1936274         19362444         scaffold_8         1936240         scaffold_8         1936444         scaffold_8         19364	SAM/BAM	bMelGeo1	scaffold_8	19355081	19355657	scaffold_8_ORF.724809	-2 -	-	bMelGeo1 scaffold_8 19355247 19355622 sca	
VCF/BCF         bMelGeo1         scaffold_8         19362790         19363480         scaffold_8_00971         -3         -           bMelGeo1         scaffold_8         1935570         19355536         scaffold_8_00FR.880991         -3         -         <	BED	bMelGeo1	scaffold_8	19350302	19350857	scaffold_8_ORF.724824	-2	-	bMelGeo1 scaffold_8 19362774 19363404 sca	alt.vconcat(*charts).resolve_scale( v='shared'
VCF/BCF         bMelGeo1         scaffold_8         19355536         scaffold_8_ORF.880991         -3         -         fontsize=10, anchor*istart',            bMelGeo1         scaffold_8         19350316         scaffold_8_ORF.880991         -3         -         anchor*istart',	NOFIDOE	bMelGeo1	scaffold_8	19362790	19363480	scaffold_8_ORF.880971	-3 -	-	DWergeot scallora"s 13320451 13320238 2CG.	).configure_title(
International ensitient & 10350310 10350310 10350310 2030310 2 - 2 - 2 - 2 - 2 - 2 - 2 - 2 - 2 - 2	VCF/BCF	bMelGeo1	scaffold_8	19355170	19355536	scaffold_8_ORF.880991	-3 -	-		<pre>fontSize=10, anchor='start',</pre>
	<	hMalGao1	scaffold 8	10350310	10350016	coaffold & OPE 881005	_? .	-	<b>`````````````````````````````````````</b>	)
										3 +++
										-2-
										-1
										1 xbp1 ex1
										xbp-lu-pt
										2-
										3







## Dynamically define reusable tools

		-			
Interactive JupyTool and notebook (Galaxy Version 0.0.1)	☆ •				
o you already have a notebook?					
Start with a fresh notebook		тт 1	C 1	D	
not, no problem we will provide you with a default one. ser inputs		User-d	enned	Re-runnabl	e and usable
1: User inputs		inputs		in workflow	/S
Name for parameter		mpace			
dataset1				,	
Required. ASCII letters and numbers only. Additional optional description			/		
Choose the input type		🂢 File Edit View Pan	Kernel Tabs Settings Help		
Dataset	•		± C E Launcher	X e ipython_galaxy_noteboo	Python 3
Select value		O Mame A	Last Modified		.,
D   D   2: Pasted Entry	• •	<ul> <li>galaxy_inputs</li> <li>outputs</li> <li>ipython_galaxy_not</li> </ul>	4 minutes ago a minute ago 2 minutes ago You can ad open (get	ome to the interactive Galaxy IP	ython Notebook. can access dataset number 42 with handle = wi(filename txt). The dataset will then be
2: User inputs	ā	4	available	in your galaxy history.	
Name for parameter			in R, Julia,	ng a non-Python kernel, get and put are available as command- , and Ruby. For example, to read dataset number 42 into R, you can	<pre>write handle &lt;- file(system('get -i 42',</pre>
someint			intern =	TRUE)). To save data in R, write the data to a file and then call s	<pre>/stem('put -p filename.txt') . Notebooks can be hop interface</pre>
Required. ASCII letters and numbers only.			More help	and informations can be found on the project website.	in interface.
Additional optional description			[1]: GALAXY_IN	IPUTS	
Choose the input type			<pre>[1]: {'dataset     'metad     'someint</pre>	:1': [{'path': 'galaxy_inputs/dataset1/Pasted Entry.txt',  ata_files': []}], :': 999}	
Integer	•		[2]: with open	<pre>('outputs/new_file.txt', 'w') as fh:</pre>	
Select value			fh.wr for i in	<pre>rite('new file with user provided integer input: %i' % (GALA) range(2):</pre>	Y_INPUTS['someint']))
999			with	<pre>open('outputs/collection/%s.txt' % (i), 'w') as fh: fh.write('new collection file: %s with user provided integer</pre>	<pre>input: %i' % (i, GALAXY_INPUTS['someint']))</pre>
+ Insert User inputs			t 1: [		
✓ Execute		0 🖪 1 ④ Python 3   Idle			Mode: Edit 🐼 Ln 1, Col 1 ipython galaxy notebook.ipyn
					÷ 17 _5 7_ 17

•





## $Data \leftrightarrow Compute$



https://www.ncbi.nlm.nih.gov/sra/docs/sragrowth/






- Remote workflow execution
- Tiered storage





- Global Pulsar cloud
- Bring your own storage (BYOS)

# Training & Outreach

### A Global Hub: A home on the web for every Galaxy community



### The Latest GTN News

Keep an eve on this page for the latest news around the GTN. New tutorials, GTN features, upcoming training events, and much much more!

### New Tutorial: Workflow Reports

Vor	kflow Invocations				
	Warkflow	History	Invoked 0	Updated 0	State
^	GTN Training: Galaxy 101 Far Everyone (Importe	Weddlow Reports Tet.	34 minutes ago	34 minutes ego	scheduled
W	ew Report			hrecati	ox 4824149473464
	10.0	to steps accessfully scheduled.			
		9 af 9 piles complete.			
D	wishaad BieCompute Object				
1.1	Deputs				
1.2	Trans.				

Authors: new tutorial Galaxy Tips & Tricks

Workflow reports are a great Galaxy feature that enable you to define summary reports for any workflow to display the most important results in a single view. This tutorial will guide you through the basics of viewing, customizing, and sharing such workflow reports!

Full Story

### **GTN Metadata Schemas**





Within the GTN we have had schemas for the various metadata we keep for a long time. It helps us automatically check community contributions to ensure they conform to all of the GTN standards, and will work correctly in our environment.

### New Tutorial Feature: Choose Your Own Tutorial



### New Tutorial: VGP assembly pipeline



#### Authors:

new tutorial assembly pacbio

We are proud to announce that, as result of the collaboration with the Vertebrate Genomes Project (VGP), a new training describing the VGP assembly pipeline (fig. 1) is now available in the Galaxy Training Network. The Vertebrate Genomes Project aims to generate high-quality, near-error-free, gap-free, chromosome-level, haplotype-phased, annotated reference genome assemblies for every vertebrate species.



#### New Feature: Automatic RMarkdown

	contains 83 even and 11 sociables.				
INTER-COM					Hale
May is 8 Useful? How Does 1 Compare To sing Base Functions 87	urt = "https://rec.githubusero golati.cov" maleep < read_conturii	ertmet, cons'genomic se lass	/dapleta/was	ter/inst/exidete/e	nieu, s
Data: Manmats Skeep	Ness II Others: II				
Important dplyt Verbo To Innvention	bilaiter "," abr 53) same genes, vere, erier, conservation abr 53) same genes, vere, erier, conservation				
ADV. VERSI IN ADDR	and the subject rangest, rangester, and, ranges super-				
Pipe Operator, Mi-H	I be "specify the relative the full online specification for this data. I Specify the column types or set "show cal tapes = FALSE" to exist this message.				
Educing Columns Using electric)					( white
Servicing Rouis Using ULTAHO	1000				
Seeking Rows Dang 115410 Anarge Of Re-order Rows ing arrange(c)	Table	genes 1077-	VITE	ander 	
Servicing Hous Dang ITTHYC Amange Or Re-order Rows and arrange O Drate New Column Using	nate nate Creater	genus  Acronyx	VIII 	onfer Indian Cerryone	•
Seecing Rose Using Ulter() Anange Or Re-order Rose long annexes Drate New Octomo Using estate()	name school Deathol Did mobility	genus ichtei Achergo Aches	vere -chris	ender 	
Severing House Using UTINO Amonge OF Re-order Rose bing arrangers Oracle View Ocianes Using House Sever Ocianes Using House Severing Of Re-	name color Coesteh Out modey Mourtain biasar	genus octros Acteory Actos Appconta	vera cons one Natio	order -chro Carriven Primars Fodorita	•
Dencing Nove Using USING Average Or Re-order Rows bing arranges Create New Online Using Method Dente summaries of Re	Incomp Index Index Creater Contenter Contenter Content Inter Salet Interner	getsa schre Actoryc Actor Apsonta Biotra	vere const oned helto oned	order -chro Carrivers Prosaes Rodenta Sociooraupha	



Further building on the work in the automatic Jupyter Notebook, we've now re-written the Jupyter export to be faster, and more importantly added support for R and RMarkdown! Check out an example material. Here we take the content of the tutorial, written like normal GTN markdown, and we automatically convert it to Jupyter Notebooks and now RMarkdown documents! Check out the documentation on how to setup your tutorials to support this.



#### New Tutorials: PacBio data QC and Genome Assembly, and Genome Annotation with Funannotate



#### Authors:

new tutorial assembly genome-annotation

We have just finished a new collection of training material for genome assembly and annotation! You now have access to slides and tutorials, focused on the complete analysis of the genome of a fungus species (Mucor mucedo).

Full Story

leve annuation for'. In your history • D'Gwe association for'. Drosopii La\_meLanopacker. 80096-87. gif 'Output/Servert' Serve-ID "\t" read-count (WaltOC/BCSeq2/edgeR/lime-vcom compatible) 'Create game-length file'. Yes . In 'Option: for paired-end read



### GTN Smörgåsbord: 2022 Tapas edition

Welcome & Practical Information

Welcome & Course Information

Get set up for the course Meet & Join the Galaxy Community!

- World wide attendance
- 2500 registrations
- GTN video library

GTN Video Library

80.1

15.3

Admin

6.1

hours

2.5

Proteomics

Galaxy interface

Video Data

Of content in the GTN & Gallantries' Video Library

8.5

hours

4.7

hours

2.3

Transcriptomics

Metagenomics

Sequence analysis

- GTN course builder (beta, feedback welcome)

Home Events Video Library Course Builder GitHub

142

Videos

8.1

hours

4.4

hours

2.1

Epigenetics

Webinar

Statistics

7.6

hours

Dev

2.9

hours

Galaxy

1.8

Data science



#### Circos Visualisation in Galaxy: Intro Circos Visualisation in Galaxy: Hands-on! Video Tutorial Helena Rasche Avans Hogeschoo Watch on Pionia Video Lenath: 50m avans Subtitles By: Beatriz Serrano-Solano, Helena Rasche, Saskia Hiltemann, Alireza Khantevmoori Description: Circos is a popular tool for creating circular graphs to display genomic data. In this tutorial you will create a Circos plot for a cancer dataset Supported Servers • 🔵 UseGalaxy.org.au G UseGalaxy.eu Galaxy.org Supporting Materials • Tutorial: Visualisation with Circos · Video: Watch it on YouTube . EAO Document - Have a question about this training? Check here to see if it has already been answered · Ask an instructor - Have question about the training? Did you run into a problem? Just wanna chat? Ask away on Slack! (Channel: #5-c circos-visualisation ) · Finished the session? - Let us know that you've finished it and what you thought of it! On Slack: (Channel: #5-c circos-visualisation ). Thanks! · Enjoyed it? - Like the video on YouTube. Tweet, and follow the GTN on Twitter! @gxytraining

#### Day 1: Introduction to Galaxy and NGS analysis

Today we start by introducing the Galaxy Platform, and show you how to analyze your NGS data using Galaxy. Today we have a mixture of lecutres, demos, and hands-on tutorials.

Icebreaker: Introduce yourself in Slack!	~
A Very Short Introduction to Galaxy	^
Your First Galaxy Analysis	^
Demo: NGS Data Logistics using SARS-CoV-2 data	^
Quality Control: Intro	^
Quality Control: Hands-on!	^
Mapping: Introduction	^
Mapping: Hands-on!	^
Genome Assembly: Introduction	^
Genome Assembly: Hands-on!	^
Wrap up & Socialize	^

#### Day 2: Introduction to RNA-Seq analysis with Galaxy and R

Today we dive into (bulk) RNA-seq analysis. We will walk you through an end-to-end analysis and Galaxy, and show you how to perform downstream analysis on the results using Rstudio in Galaxy.



#### ^

# Galaxy Projects Enabling Tremendous Science

### VGP: Assembling reference genomes for all vertebrates in Galaxy https://galaxyproject.org/projects/vgp/

### - Galaxy

#### Tools

Get Data Get ENCODE Data **ENCODE** Tools Text Manipulation Filter and Sort Join, Subtract and Group **Convert Formats** Extract Features Fetch Sequences **Fetch Alignments Get Genomic Scores Operate on Genomic Intervals** Statistics Graph/Display Data EMBOSS HYPHY

Galaxy at	ISMB/ECCB200	7 in Vienna	(July 21-25)	

A

• July 19 10:00am | The Galaxy Framework for Computational Biology Tool Integration July 25 10:15am Room L | GALAXY: a simple web application for the analysis of enormous datasets July 25 11:10am Room L | Effortless integration of tools into simple, scalable, multiuser, pythonic framework

Two Galaxy sessions at the 57th Annual Meeting of the American Society for Human Genetics (October 23-27)! The first session will be designed for biomedical researchers and will concentrate on analyses of genomic and disease association data. The second session will be organized as a hands-on software development workshop for bioinformaticians and computational biologists. For more information click here.

Unsequenced Genomes of the World | July 2007



Giraffe (Giraffa camelopardalis) | Eastern Cape, SAR

Info: report bugs | wiki | screencasts Account: create | login History (options) refresh | collapse all 1 Your history is empty. Click 'Get Data' on the left pane to start

#### Info: report bugs | wiki | screencasts | blog Logged in as james@jamestaylor.org: manage | logout

Tools Get Data Get ENCODE Data **ENCODE Tools Text Manipulation** Filter and Sort Join, Subtract and Group **Convert Formats Extract Features** Pattern-Matching **Fetch Sequences** Fetch Alignments Get Genomic Scores **Operate on Genomic Intervals** Statistics Graph/Display Data **Evolution: HyPhy** EMBOSS

- Galaxy

	History (options)
Visit Galaxy <u>Talk and Posters</u> at the Genome Informatics Meeting in Cold Spring Harbor (November 1 - 5, 2007)	refresh   collapse all
Download Galaxy brochure here.	12: Neighbor Joining Tree of data 10
Unsequenced Genomes of the World   November 2007	
	11: Neighbor Joining Tree of data 10
	10: Stitch MAF blocks on da <u>[5-way multiZ</u> (hg18,panTro2,rheMac2,mr
	9: Compare two Queries on data 7 and data 1
	7: Filter on data 5
	5: Count on data 4
	4: Join on data 2 and data 1
	2: UCSC Main on Human: snp126 (genome)
Spectacied Figing Fox (Pteropus conspiculatus)   Queensland, Australia	1: UCSC Main on Human:

#### Galaxy is for Biologists

Use this site to access popular sources of data like the UCSC Table Browser. Run analyses right on the spot using a variety of integrated tools. Your results are always available and can be easily shared with others. Just watch how.

History (options)				
refresh   collapse all				
12: Neighbor Joining Tree on (1) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2				
11: Neighbor Joining Tree on ⊕ Ø ☆ data 10				
10: Stitch MAF blocks on data 9 () () () [5-way multiz (hg18,panTro2,rheMac2,mm8,canFam2)]				
9: Compare two Queries on (1) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2				
7: Filter on data 5 🔹 🖉 💥				
5: Count on data 4 🔹 🖉 💥				
4: Join on data 2 and data 1 👁 🖉 💥				
2: UCSC Main on Human:  snp126 (genome)				
1: UCSC Main on Human: knownGene (genome)				

VERTEBRATE GENOMES PROJECT	CONSORTIUM
	$\mathbf{X}$
	C
	-
	U
	ШHL
	Ш
	H
	U
	Ш
	_
	C
	α
	0
	4

Species
Corvus hawaiiensis
Taeniopygia guttata
Scomber japonicus
Sorex araneus
Macropus eugenii
Neoarius graeffei
Acridotheres tristis
Spea bombifrons
Spheniscus humboldti
Monodelphis domestica
Mesoplodon densirostris
Podargus strigoides
Ammospiza nelsoni
Ammodramus caudacutus
Rhea pennata
Rhynochetos jubatus
Cynocephalus volans
Vipera latastei
Neofelis nebulosa
Melospiza georgiana
Gastrophryne carolinensis
Colius striatus
Gavialis gangeticus

Common name
Hawaiian crow
Zebra finch
chub mackerel
Common shrew
Tammar wallaby
Blue salmon catfish
Common myna
Plain's spadefoot toad
Humboldt penguin
Gray short-tailed opossum
Blainville's beak whale
Tawny frogmouth
Nelson's sparrow
Salt marsh sparrow
Darwin's southern rhea
Kagu
Philippine flying lemur
Snub-nosed viper
Clouded leopard
Swamp sparrow
Eastern narrowmouth toac
Speckled mousebird
Gharial

 $\sim$ 

Genome size	Assembler
1,081,864,808	Giulio Formenti
1,009,754,198	Nadolina Brajuka
688,600,335	Linelle Abueg
2,230,886,386	Linelle Abueg
3,312,892,930	Nadolina Brajuka
2,420,249,501	Linelle Abueg
1,354,313,347	Linelle Abueg
959,239,875	Nadolina Brajuka
1,246,817,591	Linelle Abueg
3,216,945,970	Nadolina Brajuka
3,125,258,145	Nadolina Brajuka
1,092,727,389	Linelle Abueg
1,190,976,425	Nivesh Jain
1,225,461,744	Brian O'Toole
1,255,251,577	Linelle Abueg
1,312,755,001	Linelle Abueg
2,715,040,762	Linelle Abueg
1,444,460,287	Marc Palmada
2,395,992,815	Brian O'Toole
1,334,824,158	Tatiana Tilley
<mark>4,095,803,536</mark>	Linelle Abueg
1,130,197,506	Linelle Abueg
2,301,125,167	Nivesh Jain

### AnVIL NHGRI Analysis Visualization and Informatics Labspace

https://anvilproject.org



### 600,000 human genomes and growing!

### When to use Galaxy in AnVIL?

### What does Galaxy bring to AnVIL users?

- All features of Galaxy: accessible, reproducible, integrative science with thousands of tools
- Large, active community of users and contributors
- World-wide training network with materials, educating thousands every year

### What does AnVIL bring to Galaxy users?

- Access to protected NHGRI datasets in FedRAMP-certified ecosystem
- Avoid data downloads
- Use Galaxy without quotas
- Be your own Galaxy administrator
- Connect datasets in novel ways

🔍 🔍 💿 Galaxy	× +				
← → C 🔒 leonardo-anvil.dsde-dev.broadinstitute.org:30443/proxy/google/v1/apps/galaxy-dev-env 🖈 🔅 🗄					
📱 🏦 AnVIL		Using 9.6 GB			
AnVIL Tools 🟠 🗘	Callant East	History 😂 🕇 🖽 🌣			
search tools	Ines from a	search datasets 🛛 🕄			
Get Data	Select first	AnVIL RNA-seq			
Collection Operations	10	9 shown			
GENERAL TEXT TOOLS		1.61 GB			
Text Manipulation	lines				
Add column to an existing dataset		9: RefSeq_reference_			
Concatenate datasets tail-to-head					
Condense consecutive characters	9: RefSeq_reference_GTF_(DSv2) -	8: Megakaryocyte_rep			
Convert delimiters to TAB		549358_2)			
Merge Columns together		7: Megakarvocyte rep			
Create single interval as a new dataset	✓ Execute	1_reverse_read_(SRR 549357_2)			
Cut columns from a table		203.5 MB format: fastqsanger, database:			
Change Case of selected columns	What it does				
Paste two files side by side	This tool outputs specified number of lines from the mm10				
Remove beginning of a file	beginning of a dataset	uploaded fastqsanger file			
Select random lines from a file		◙ ◙ ₩ ? ♥ ♥			
Select first lines from a dataset	Example	@SRR549357.100000035/2			
Select last lines from a dataset	Selecting 2 lines from this:	CCCCCACTCCTAAGAGGCGGATGGTCGCGTCCATACCC			
Trim leading or trailing characters	chr7 56632 56652 D17003_CTCF_R6 310 +	+ 7-8			
Line/Word/Character count of a	chr7 56736 56756 D17003_CTCF_R7 354 +	esrr549357.100000062/2			
dataset	chr7 56772 56792 D17003_CTCF_R7 372 +				
Secure Hash / Message Digest on a dataset	chr7 56775 56795 D17003_CTCF_R4 207 + will produce:	2_forward_read_(SRR 549358 1)			
Compute an expression on every row	che7 56622 56652 017002 CTCE 86 200				
annotateMyIDs annotate a generic set of identifiers	chr7 56736 56756 D17003_CTCF_R7 354 +	5: G1E_rep1_reverse_			
melt collapse combinations of variables:values to single lines		4: G1E_rep2_forward_			
Replace parts of text		3: G1E_rep2_reverse_ ④ A X			
cast expand combinations of		read_(SRR549356_2)			
variables:values to columnar format		2: Megakarvocyte rep 💿 🖋 🗙			
JQ process JSON		1_forward_read_(SRR			
Rebase GFF3 features against parent features		549357_1)			
Remove columns by heading		read_(SRR549355_1)			
Sort Column Order by heading					
<					

# Future Pandemic Effort: High quality Galaxy workflows for pathogen variation

https://galaxyproject.org/projects/{covid19|mpxv}



### Future Pandemic Effort

- IWC workflows for variation analysis
- Test data for evaluation
- Documentation and training

# MicroGalaxy: Workflows and training for bacterial and viral data

https://microgalaxy.usegalaxy.eu

### MicroGalaxy

- Microbial genomics is underdeveloped
- Acute need for standardized
   "best practices"
- Documentation and training



Datasets Taxonomy Genome Gene Command-line tools Documentation

### NCBI Datasets BETA

NCBI Datasets is a new resource that lets you easily gather data from across NCBI databases. Find and download gene, transcript, protein and genome sequences, annotation and metadata.

#### Genomes

Browse and download genome data using our species pages. Genome data includes genome, transcript and protein sequences, genome annotation and metadata.

### Find a species

#### **Taxonomy Browser**

View taxonomic relationships and find genome data for closely related species using our interactive species browser.

Browse taxonomy

Feedback

## Beyond Biomedical Data: Climate, Neutron Scattering, ...

### Beyond Biomedical Data:

- Galaxy is NOT just genomics
- Galaxy is NOT life sciences
- Galaxy is EVERYTHING that has tools/data/infrastructure

Fight analytical inequality across disciplines



# Our biggest challenge is making Galaxy features known!

# STOP HERE



### Data $\leftrightarrow$ Compute

- Remote workflow execution
- Tiered storage
- Pulsar cloud
- Bring your own storage (BYOS)



### Data $\leftrightarrow$ Compute

- Remote workflow execution
- Tiered storage
- Pulsar cloud
- Bring your own storage (BYOS)

s pilot on usegalaxy.org usegalaxy.eu pilot on  $\mathbf{c}$ S

