





27/05/2024



Initiation

Loraine Guéguen

Erwan Corre

Credits to Gildas Le Corguillé, Galaxy Training Network

v3.2





This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. [link]





- Connect to local computer
- Get slides and datasets: <u>https://abims.sb-roscoff.fr/training/courses</u>
- Login on Galaxy France: <u>https://usegalaxy.fr</u>
- Visit https://usegalaxy.fr/join-training/init2024/
 - reserved computing resources



- Schedule
 - 09:00 12:00
 - 13:30 17:00

Short round table



- Learning objectives:
 - Familiarize yourself with the basics of Galaxy
 - Learn how to import data
 - Learn how to run tools
 - Learn how histories work
 - Learn how to create a workflow
 - Learn how to share your work
 - Understand and master dataset collections



Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

- Introduction
- Data import
- Tools
- Dataset
- History
- Workflow
- Share







Sequence files manipulation





INTRODUCTION / PROBLEMATIC



- In biomedical research, high-throughput technologies produce large datasets.
- How to perform analyses of these data **without bioinformatics skills** ?
 - Assemble transcript sequences de novo
 - Determine the gene expression
 - Build a phylogenetic tree
 - Predict subcellular targeting for proteins
 - Identify and quantify metabolites detected by LC-MS
 - 0 ...

Station Biologique Roscoff Introduction / problematic

Command line tools represent the majority of scientific tools

- + good parameters completeness
- + can be executed on high performance computers
- + automation, workflows ...
- minimum linux knowledge is required
- cruel lack of ergonomics

Roscoff Introduction / problematic

Command line tools represent the majority of scientific tools

- + good parameters completeness
- + can be executed on high performance computers
- + automation, workflows ...
- minimum linux knowledge is required
- cruel lack of ergonomics

```
login@sbr4-p023:~$ ssh -Y bioinfo.sb-roscoff.fr
```

[...]

login@slurm0:~\$ cd /shared/projects/my_project/finalresult

```
login@slurm0:~$ head -n 2 ../input/query.fasta
```

>Ec-00_000010.1 Ferric reductase, NAD binding (657) ;mRNA; r:150-6731 MGSQCWQGVLNHGNHFHDIYMVCPDDPTQWCNQFSFESGGPTPNQYRLRMIGIFMGLVCSSHFAIILVPVSRD

```
login@slurm0:finalresult$ module load blast/2.9.0
```

```
login@slurm0:finalresult$ srun --mem 100G --cpus-per-task 4 blastp -query
../input/query.fasta -db /shared/bank/uniref90/current/blast/uniref90 -out query.out -outfmt
6 -evalue 1e-6
```

Station Biologique Introduction / problematic

Command line tools represent the majority of scientific tools

- + good parameters completeness
- + can be executed on high performance computers
- + automation, workflows ...

Roscoff

- minimum linux knowledge is required
- cruel lack of ergonomics



login@slurm0:finalresult\$ srun --mem 100G --cpus-per-task 4 blastp -query ../input/query.fasta -db /shared/bank/uniref90/current/blast/uniref90 -out query.out -outfmt 6 -evalue 1e-6

Station Biologique Roscoff ABMS Introduction / problematic

MR.GEEK





login@sbr4-p023:~\$ ssh -Y bioinfo.sb-roscoff.fr

[...]

.ogin@slurm0:~\$ cd /shared/projects/my_project/finalresult

ogin@slurm0:~\$ head -n 2 ../input/query.fasta

Ec-00_000010.1 Ferric reductase, NAD binding (657) ;mRNA; :150-6731

GSQCWQGVLNHGNHFHDIYMVCPDDPTQWCNQFSFESGGPTPNQYRLRMIGIFMGLVCSSH: LLVPVSRD

.ogin@slurm0:finalresult\$ module load blast/2.9.0

Login@slurm0:finalresult\$ srun --mem 100G --cpus-per-task 4
plastp -query ../input/query.fasta -db
/shared/bank/uniref90/current/blast/uniref90 -out query.out
-outfmt 6 -evalue 1e-6











INTRODUCTION / GALAXY





https://training.galaxyproject.org/training-material/topics/introduction/slides/introduction.html

usegalaxy.fr GALAXY FRANCE





Galaxy ABiMS --> Galaxy France



https://usegalaxy.fr/

Station Biologique Roscoff

ARMS



🖬 Galaxy France	Worklow Visualize Données partagées Aide Utilisateur In Itilianteur In Itilian			
Tools f	-	History	+ =	•
search tools 😵	Galaxy France	Rechercher des données	*	×
🏦 Upload Data	The homenage of the Erench Calaus community	vHPO		1
Get Data	The nomepage of the French Galaxy community	-		
Send Data		8 /52 KB	V 0	
Collection Operations	Galaxy is an open-source platform for FAIR data analysis that enables users to:	2 52		4
GENERAL TEXT TOOLS	use tools from various domains (that can be plugged into workflows) through its graphical web	6: blastp prot vs 'uniprot_20 23_02'	0/1	
Text Manipulation	interface. FRANCE			5
Filter and Sort	run code in interactive environments (RStudio, Jupyter) along with other tools or workflows.	5: blastp prot vs 'uniprot_ 2023_02'	0/1	1
Join, Subtract and Group	manage data by sharing and publishing results, workflows, and visualizations.	a survey of some of	2000	5
GENOMIC FILE MANIPULATION	ensure reproducibility by capturing the necessary information to repeat and understand data	2023_02' 1 5	0/1	
Convert Formats	anaryses.		0.45	

and the same and satisfies allows

Galaxy France: <u>https://usegalaxy.fr</u>

Station Biologique Roscoff

1172



📮 Galaxy France	i Workflow Visualize Données partagées - Admin Aide - Utilisateur - 📂 🗘 🏢	Using	12%
From the 4th to 7th of April, usegalaxy fr will be shut	down for maintenance		
Tools ☆ ≔		History C+ [•
search tools		Rechercher des données	0
1 Upload Data		Unnamed history (empty)	•
Transcriptomics			
Biodiversity COMPUTATIONAL BIOLOGY	Welcome to usegalaxy.fr	Cet historique est vide. You can Charger vos propres données	or
Probes and primers	By using this Galaxy instance, we assume that you have read and accept the Term Of Use	Charger des données depuis	
"Nucleic acid sites, features and motifs"	For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy	une source externe	
"Sequence composition, complexity and repeats"	• 13/01/2022: usegalaxy.fr is now running the release 21.09 of Galaxy. Please check the 21.09 user release notes.		
Sequence assembly			
Mapping	Domain specific subdomains:		
Computational biology	The same instance but with only the dedicated tools in order to focus on the domain		
COMPUTER SCIENCE	The same instance but with only the deutcated tools in order to focus on the domain		
Data visualisation	Workflow4Metabolomics 200 Covid10 Metaborcoding		
LABORATORY TECHNIQUES	4 Workhow4wictabolomics 4 Workhow4wictabolomics 4 Workhow4wictabolomics 4 Workhow4wictabolomics 4 Workhow4wictabolomics 4 Workhow4wictabolomics 4 Workhow4wictabolomics 4 EPOGS Office Mothur Obitools		
RNA-Seq	analysis and annotation for communuty approuved workflows DADA2, PICRUSt		
MATHEMATICS	Notdestandards Metabolomics community and datasets.		
Statistics and probability	ProteoREAsk the GalaxyCat		
OMICS	Functional analysis and exploration		
Genomics	of proteomics and transcriptomics Why not search on the IFB Galaxy		
Comparative genomics	data in biomedical research. Catalog ?		
Uncategorized			
WORKFLOWS	Hosted by:		
All workflows		JII.	>

Station Biologique Roscoff Galaxy France

立 ·

× ×

Galaxy France

Collection Operations GENERAL TEXT TOOLS

Join, Subtract and Group

COMMON GENOMICS TOOLS

GENOMICS ANALYSIS

Operate on Genomic Intervals

Fetch Alignments/Sequences

GENOMIC FILE MANIPULATION

Text Manipulation

Filter and Sort

Convert Formats

FASTA/FASTQ FASTQ Quality Control

SAM/BAM BED

VCF/BCF

Nanopore

Annotation Assembly

Mapping Variant Calling

RNA-Seq

Peak Calling

Epigenetics Phylogenetics Phenotype Association

Single-cell GENOMICS TOOLKITS

Picard

1. Upload Data

Tools

search tools

Get Data Send Data



Galaxy France

The homepage of the French Galaxy community

Galaxy is an open-source platform for FAIR data analysis that enables users to:

- use tools from various domains (that can be plugged into workflows) through its graphical web interface
- run code in interactive environments (RStudio, Jupyter...) along with other tools or workflows.
- manage data by sharing and publishing results, workflows, and visualizations.
- ensure reproducibility by capturing the necessary information to repeat and understand data analyses.

The Galaxy Community is actively involved in helping the ecosystem improve and sharing scientific discoveries.

The French Galaxy server

The French Galaxy server UseGalaxy.fr is maintained by IFB NNCR Cluster Task force. Please check our Terms of Use and data retention police before using the server. We offer thousands of tools, increased quota on temporary basis, and compute infrastructure for trainers through Training Infrastructure as a Service (TlaaS).

The following regional platforms, members of IFB, are involved in the maintenance and development of UseGalaxy.fr:

- · GenOuest, in Rennes
- · ABIMS, in Roscoff



	Us	ing 4	15%
History	+	=	ŀ
Rechercher des données		*	×
vHPO			1
8 752 kB		0	6
2 35			
6: blastp prot vs 'uniprot_20 23_02'	0	1	*
▲ 5: blastp prot vs 'uniprot_ 2023_02'	0	1	
A 4: blastp prot vs 'uniprot_ 2023_02'	0	1	
3: blastp prot vs 'uniprot_swi ssprot_2020_03'	0	1	
▲ 2: blastp prot vs 'uniprot_ 2023_02'	0	1	
1: prot	0	1	

= Galaxy

eli

FRANCE



Get help: <u>https://community.france-bioinformatique.fr/c/galaxy/8</u>

Community Support		S'inscrire	onnecter	ର
► Welcome message				×
Galaxy ▶ toutes les étiquettes ▶ tout ◄ Récents Top				
Sujet		Réponses	Vues	Activité
FTP usegalaxy.fr Bonjour, J'ai eu une demande en local (support.abims) pour savoir comment se connecter au serveur ftp de usegalaxy en ligne de commande. Je copie ici la solution proposée afin d'en faire bénéficier le plus grand nombre … lire la suite		1	88	sept. '21
 README: before request a Galaxy quota extension In order to effectively free space and optimize the storage capacity of our servers, please do not forget to clean regularly your datasets and histories by clicking on the small gear at the top of your history ("Pu lire la suite 		2	180	janv. '21
A propos de la catégorie Galaxy For any questions, issues and request regarding usegalaxy.fr and its subdomains *.usegalaxy.fr	3	2	469	oct. '20
☑ FROGS - job was terminated because it used more memory than it was allocated		21	54	11 h
FROGS - bug à l'affichage des sorties html	AL	4	17	^{14 h} 18





Hands-on DATA IMPORT & FIRST TOOL



Hands-on

Station Biologique

Roscoff

https://training.galaxyproject.org /training-material/topics/introduc tion/tutorials/galaxy-intro-101/tu torial.html

Galaxy Basics for genomics

Galaxy Training! Introduction to Galaxy



Learning

Pathways

Help

Settings

Overview

⑦ Questions: \odot \odot · Which coding exon has the highest number of single nucleotide polymorphisms (SNPs) on human chromosome 22? Objectives: · Familiarize yourself with the basics of Galaxy · Learn how to obtain data from external sources · Learn how to run tools · Learn how histories work · Learn how to create a workflow · Learn how to share your work Time estimation: 1 hour C Level: Introductory C III III C Supporting Materials: Datasets Korkflows ⑦ FAQs Recordings • Available on these Galaxies • Published: Dec 19, 2016 Last modification: May 3, 2024 4 License: Tutorial Content is licensed under Creative Commons Attribution 4.0 International License. The GTN Framework is licensed under MIT PURL: https://gxy.io/GTN:T00186 Rating: 4.1 (10 recent ratings, 141 all time) -O- Revision: 131

Stop after step "Find exons with the most SNPs"

Setting the stage: Exons and SNPs	This tutorial aims to familiarize you with the Galaxy user interface. It will teach you how to perform basic tasks such as importing data, running tools, working with histories, creating workflows, and sharing your work.
Get your workspace ready	Decomment: Results may vary
Histories and workflows: A brief	Your results may be slightly different from the ones presented in this tutorial due to differing versions of tools, reference data, external databases, or because of stochastic processes in the algorithms.
introduction Share your work	20

Agenda

20

Search Tutorials



TOOLS & HISTORY









DATA IMPORT





https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/get-data/sli des.html



Tips & tricks DATA IMPORT



STEP 1: CHOOSE A FTP CLIENT



Data import using FTP

Example with WinSCP for usegalaxy.fr



Command line: https://community.france-bioinformatique.fr/t/ftp-usegalaxy-fr/1453



For HUGE public resources: genome, databank ...

--> Make a request to the support team

📮 Galaxy France	👫 Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 📰	Using 3%
Tools ☆	NCDI DI ACT i blante Constante	Ĥistory 🕄 🕇 🖽 🌣
search tools	database with protein query sequence(s) (Galaxy Version 0.3.3)	search datasets 2 🛛
🏦 Upload Data	Protein query sequence(s)	blast hands-on 2022
Get Data	□ □ 2: Drosophila_melanogaster.BDGP6	66.36 MB
Send Data	(-query)	
Collection Operations	Subject database/sequences	13: NCBI BLAST+ blastn X
GENERAL TEXT TOOLS	Locally installed BLAST database	a list with 5 items
Text Manipulation	Protein BLAST database	12: queries X
Filter and Sort	Select/Unselect all	a list with 5 items
Join, Subtract and		11: blastx query5. 💿 🖋 🗙
Group		fa vs 'protein BLA
GENOMIC FILE MANIPULATION	NCBI NR (non redundant) 2021-01-15	10: megablast que 💿 🖋 🗙
Convert Formats	uniprot_swissprot	ry5.fa vs 'nucleoti
FASTA/FASTQ	uniprot	de BLAST database from data
FASTQ Quality Control	See help text for default parameter values for each bLAST type. (-task)	9: nucleotide BLA 💿 🖋 🗙
SAM/BAM	Set expectation value cutoff	ST database from data 1
BED	0.001	
< > ×	(-evalue)	database from da
< Ⅲ	Output format	v III >



Icons & buttons DATASET DATASET C LL A: ?



Dataset display : text, tabular, pdf, picture, html ...

= Galaxy France	প Workflow	Visualize Données partagées 🔻 Aide 👻 Utilisateur 👻 💼	€ 1	Using 4
Tools 🗘 🔹	Column 1	Column 2	History	+ ≓
search tools 😵 🗙		TRINITY DN2 c0 o1 i1	Rechercher des données	*
🍰 Upload Data	TRINITY_DN1_c0_g1	TRINITY_DN1_c0_g1_i1	Trinity example	
Cat Data	TRINITY_DN5_c1_g1	TRINITY_DN5_c1_g1_i1		
	TRINITY_DN0_c0_g1	TRINITY_DN0_c0_g1_i1	S 41.3 kB	9
end Data	TRINITY_DN0_c0_g2	TRINITY_DN0_c0_g2_i1	2 0	
Collection Operations	TRINITY_DN0_c0_g3	TRINITY_DN0_c0_g3_i1	4: Trinity on data 2 and da	ata 💿 🌶
ENERAL TEXT TOOLS	TRINITY_DN3_c0_g1	TRINITY_DN3_c0_g1_i1	1: Gene to transcripts map	, <u> </u>
			Add Tags 📎 🕒	
Iter and Sort			7 lines format tabular , génome de	référence ?
oin, Subtract and Group				
ENOMIC FILE MANIPULATION			886⊂≞∴?	
onvert Formats			1 2	c9 c1 i1
ASTA/FASTQ			TRINITY_DN1_c0_g1 TRINITY_DN1	_c0_g1_i1
ASTQ Quality Control			TRINITY_DN5_c1_g1 TRINITY_DN5	_c1_g1_i1
AM/BAM			TRINITY_DN0_C0_g1 TRINITY_DN0 TRINITY DN0 C0 g2 TRINITY DN0	_c0_g1_i1
D				
F/BCF			3: Trinity on data 2 and da	ita 🎯 🥖
anopore			I: Assembled Transcripts	
MMON GENOMICS TOOLS			2: reads.right.fq	0 /
perate on Genomic Intervals			1: reads.left.fq	0 /
tch Alignments/Sequences				
NOMICS ANALYSIS				
notation				31



Rename, annotate, change datatype...

= Galaxy France	Workflow Visualize Données partagées 🛪 Aide 🛪 Utilisateur 🛪 💼 📢 🏢		Using 45	5%
Tools 😒 🔹	Modifier les attributs du jeu de données	History	+ ≓	•
search tools 😵 🗙		Rechercher des données	*	×
1 Upload Data	Attributes Datatypes Permissions	Trinity example		-
Get Data	Trinity on data 2 and data 1: Gene to transcripts map	🛢 41.3 kB	04	4 2
Send Data	Info			•
Collection Operations			0 4	
GENERAL TEXT TOOLS		4: Trinity on data 2 and data 1: Gene to transcripts map		
Text Manipulation	Annotation - optional	Add Tags 🍑		
Filter and Sort		7 lines		
Join, Subtract and Group		format tabular , génome de ré	férence ?	
GENOMIC FILE MANIPULATION	Add an annotation or notes to a dataset: annotations are available when a history is viewed.	₽00Сш∴?		
Convert Formats	Database/Build - optional	1 2		Т
FASTA/FASTQ		TRINITY_DN2_c0_g1 TRINITY_DN2_c0 TRINITY_DN1_c0_g1 TRINITY_DN1_c0	_g1_i1 _g1_i1	
FASTQ Quality Control		TRINITY_DN5_c1_g1 TRINITY_DN5_c1	_g1_i1	
SAM/BAM	Number of comment lines	TRINITY_DN0_c0_g1 TRINITY_DN0_c0 TRINITY_DN0_c0_g2 TRINITY_DN0_c0	_g1_i1 _g2_i1	
BED	0			
VCF/BCF	B Save C Auto-detect	3: Trinity on data 2 and data 1: Assembled Transcripts	• /	1
Nanopore				
COMMON GENOMICS TOOLS		2: reads.right.fq	•	
Operate on Genomic Intervals		1: reads.left.fq	0/	Î
Fetch Alignments/Sequences				
GENOMICS ANALYSIS			-	
Annotation			3	1



Rename, annotate, change datatype...





Download ; link to data





Informations

fools 🏠 -	Dataset Informatio	on		History	+ ≓	•
search tools 😵 🗙	Number	4		Rechercher des données	×	x
	Name	Trinity on data 2 and data 1: Gene to transcripts map				
🏦 Upload Data	Created	Tuesday Jan 11th 10:25:01 2022 GMT+1		Trinity example		1
	Filesize	273 b				
et Data	Dbkey	?		🛢 41.3 kB	0	4
end Data	Format	tabular				
ollection Operations	File contents	contents		M 45		
ollection operations	History Content API ID	9ba7911219ddd43c		4: Trinity on data 2 and data	01	T
ENERAL TEXT TOOLS	History API ID	055d73e92eaa7468		1: Gene to transcripts map		
ext Manipulation	UUID	a479ce18-2fa8-4d27-89fb-e744e47d19d0		Add Tags 🌑		
ilter and Sort	Originally Created From a File Named	Trinity.fasta.gene_trans_map		7 lines		
oin, Subtract and Group	Tool Parameters			format tabular , génome de réf	érence ?	
ENOMIC FILE MANIPULATION	Input Parameter	Value		₿₿₿₿Сш#?		
onvert Formats		Vas		1 2		
ASTA/FASTO	datasets?			TRINITY_DN2_c0_g1 TRINITY_DN2_c0_	g1_i1	
ASTATASTQ	Paired or Single-end	paired		TRINITY_DN1_c0_g1 TRINITY_DN1_c0_g1_i1		
ASTQ Quality Control	data?			TRINITY_DN5_c1_g1 TRINITY_DN5_c1_	g1_i1	
AM/BAM	Left/Forward strand reads	1: reads.left.fq	• / 1	TRINITY_DN0_c0_g1 TRINITY_DN0_c0_g1_i1 TRINITY_DN0_c0_g2 TRINITY_DN0_c0_g2_i1		
ED	Right/Reverse strand					
CE/BCE	reads	2: reads.right.fq	• / 1	3: Trinity on data 2 and data	0/	Î
	Strand specific data	false		1: Assembled Transcripts		
anopore	Jaccard Clip options	False		2: reads.right.fg	01	-
OMMON GENOMICS TOOLS	Run in silico normalization	True				
perate on Genomic Intervals	of reads			1: reads.left.fq	0/	Î
	additional_params			-		
etch Alignments/Sequences	Minimum Contig Length	200				
ENOMICS ANALYSIS	Use the genome guided	no				
nnotation	mode?					
	Error-corrected or circular					
	conconcue (CCC) nac bio					



Re-run a job

Galaxy France	👫 Workflow Visualize Données partagées 🛪 Aide 🛪 Utilisateur 🛪 🕋 🗰		sing 45%	
Tools 🌣 🔹	🖌 Trinity de novo assembly of RNA-Seg data (Galaxy Version 2.9.1+galaxy2)	+	≓ ,	
search tools 🛛 😽 🗙	- Tool Parameters	ner des données	* ×	
🍰 Upload Data	Are you pooling sequence datasets?	example	1	,
Get Data				
Send Data	Yes 👻 🗧 41.3 kB		Q 4	2
Collection Operations	Paired or Single-end data?			\$
GENERAL TEXT TOOLS	4: Trinity 1: Gene to	on data 2 and data) / T	
Text Manipulation	Paired-end Add Tags	 Image: A set of the set of the		
Filter and Sort	Left/Forward strand reads * 7 lines			
Join, Subtract and Group	🕐 🗅 3: Trinity on data 2 and data 1: Assembled Transcripts * 🔹 🕞 format tab	ular, génome de référen	ice ?	
GENOMIC FILE MANIPULATION	1: (unavailable) reads.left.rq	C 🖮 🏦 ?		
Convert Formats		2		
FASTA/FASTQ		<pre>¿_c0_g1 TRINITY_DN2_c0_g1_i 1 c0 g1 TRINITY DN1 c0 g1 i</pre>	1	
FASTQ Quality Control	(left)	5_c1_g1 TRINITY_DN5_c1_g1_i	.1	
SAM/BAM	Right/Reverse strand reads * TRINITY_ON	<pre>3_c0_g1 TRINITY_DN0_c0_g1_i 8 c0 g2 TRINITY DN0 c0 g2 i</pre>	1	
BED	C 3: Trinity on data 2 and data 1: Assembled Transcripts			
VCF/BCF	2: (unavailable) reads.right.fq 3: Trinity	on data 2 and data) / T	
Nanopore		teu manacripta		
COMMON GENOMICS TOOLS	2: reads.ri	ght.fq @) / 1	
Operate on Genomic Intervals	(right) 1: reads.let	ft.fq 🧉		
Fetch Alignments/Sequences	Strand specific data			
GENOMICS ANALYSIS	No			
Annotation				
Assembly			3:	С



Visualisation tools






Visualisations in Galaxy



Helena Rasche) 🦃 Saskia Hiltemann) (🖧 Add Contributions!

Updated: Jul 9, 2021 Or view the plain-text slides without JS Tip: press P to view the presenter notes

https://training.galaxyproject.org/training-material/topics/visualisation/slides/introduction.html



Show related items

= Galaxy France	👫 Workflow Visualize Données par	rtagées 🔹 Aide 👻 Utilisateur 💌 💼 📢 🏢		Using 45%
Tools 🔂 -			History	+ = •
search tools × ×	сомминату нив		related:4	× ×
1 Upload Data			Trinity example	1
Get Data	Galaxy France			
Send Data			2 41.3 kB	Q 4 C
Collection Operations	The homepage of the French Ga	laxy community	2 35	•
GENERAL TEXT TOOLS			4: Trinity on data 2 and a 1: Gene to transcripts n	d dat 🐵 🖍 👕 🏾
Text Manipulation	Galaxy is an open-		Add Tags 🌑	
Filter and Sort	source platform for		7 lines	
Join, Subtract and Group	FAIR data analysis		format tabular , génome de	e référence ?
GENOMIC FILE MANIPULATION	that enables users	FRANCE	B & O C	
Convert Formats	to:	NUEST/	1 2	
EASTA/EASTQ	• use tools from		TRINITY_DN2_c0_g1 TRINITY_DN	2_c0_g1_i1
	various CFRANCE		TRINITY_DN1_c0_g1 TRINITY_DN	1_c0_g1_i1
FAST& Quality Control	domains (that		TRINITY_DNS_c1_g1 TRINITY_DN TRINITY_DN0_c0_g1 TRINITY_DN	5_c1_g1_11 10_c0_g1_i1
SAM/BAM	can be plugged into workflows)	through its graphical web	TRINITY_DN0_c0_g2 TRINITY_DN	0_c0_g2_i1
BED	interface.			
VCF/BCF	• run code in interactive environn	nents (RStudio, Jupyter)	2: reads.right.fq	• / 1
Nanopore	along with other tools or workflow		🕥 1: reads.left.fq	0/1
COMMON GENOMICS TOOLS	 manage data by sharing and public 	olishing results, workflows, and		
Operate on Genomic Intervals	visualizations.			
Fetch Alignments/Sequences	 ensure reproducibility by capture 	ing the necessary information		
GENOMICS ANALYSIS	to repeat and understand data an			
Appotation	The Galaxy Community is actively invo	olved in helping the ecosystem		38

Station Biologique Roscoff ARMS Dataset Station Biologique Roscoff

Help

= Galaxy France	😤 Workflow Visualize Données partagées 🛪 Aide 🛪 Utilisateur 🛪 📻 📢 🏢		Using 45%
Tools 🔅 -	Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.9.1+galaxy2)	History	+ = -
search tools 😵 🗙		related:4	* ×
1 Upload Data	Defaults to max. read coverage of 50. (no_normalize_reads)	Trinity example	1
Get Data	Additional Options		
Send Data	Additional Options	S 41.3 kB	Q4 2
Collection Operations	Email notification		0
ENERAL TEXT TOOLS		4: Trinity on data 2 and 1. Compare to the provide the second	d dat 💿 🖍 🔋
ext Manipulation	Send an email notification when the job completes.	a I: Gene to transcripts m	ар
ilter and Sort		7 lines	
Join, Subtract and Group	Run Tool	format tabular , génome de	e référence ?
ENOMIC FILE MANIPULATION	- Help	₽800ш	
Convert Formats	Trinity assembles transcript sequences from Illumina RNA-Seq data.	1 2	
ASTA/FASTQ	Tutorials	TRINITY_DN2_c0_g1 TRINITY_DN2	2_c0_g1_i1
-ASTQ Quality Control	There is 1 tutorial available which uses this tool. View all tutorials referencing this tool.	TRINITY_DN1_c0_g1 TRINITY_DN1 TRINITY_DN5_c1_g1 TRINITY_DN5	1_c0_g1_i1 5_c1_g1_i1
SAM/BAM		TRINITY_DN0_c0_g1 TRINITY_DN0	0_c0_g1_i1
	Tutonats available in T category *	TRINITY_DN0_c0_g2 TRINITY_DN0	ð_c0_g2_i1
/CF/BCF	Citations:	🗿 2: reads.right.fq	• / 1
lanopore	- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L.,	1: reads.left.fg	0/1
OMMON GENOMICS TOOLS	Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.		
perate on Genomic Intervals	data without a reference genome. <i>Nature Biotechnology</i> , <i>29</i> (7), 644–652. https://doi.org/10.1038/nbt.1883		
etch Alianments/Sequences	Visit Citation 🗹		
ENOMICS ANALYSIS	Requirements: See details		
nnotation	- coreutils (Version 8.32)	1	20
Accombly	- rsync (Version 3.2.3)		55

Galaxy raining Notwork

History & datasets

Station Biologique

Roscoff

@ 0 · How do Galaxy histories work? Objectives: · Gain understanding on navigating and manipulating histories **X** Time estimation: 30 minutes C Supporting Materials: ⑦ FAQs Recordings - Available on these Galaxies -Published: Feb 20, 2017 Last modification: Mar 27, 2023 Dicense: Tutorial Content is licensed under Creative Commons Attribution 4.0 International License. The GTN Framework is licensed under MIT @ PURL: https://gxy.io/GTN:T00150 A Rating: 4.8 (0 recent ratings, 4 all time) -O- Revision: 12 The History A Warning: Compatible Versions of Galaxy Basic Searching This tutorial has been tested to work with 23.0 Galaxy's Interface may be different to the Galaxy where you are following this tutorial Frequently Asked • All tutorial steps will still be able to be followed (potentially with minor differences for moved buttons or changed icons.) Question Tools will all still work Feedback Citing this Tutorial When data is uploaded from your computer or analysis is done on existing data using Galaxy, each output from those steps generates a dataset. These datasets (and the output datasets from later analysis on them) are stored by Galaxy in Histories. The History All users have one 'current' history, which can be thought of as a workspace or a current working directory in bioinformatics terms. Your current history is displayed in the right hand side of the main 'Analyze Data' Galaxy page in what is called the history panel.

🌛 Galaxy Training! 🗅 Using Galaxy and Managing your Data 🎓 Learning Pathways 🔞 Help 👻 🌣 Settings 🔹

Cristóbal Gallardo

Helena Rasche

Understanding Galaxy history system

Martin Čech

Björn Grüning

Author(s) Anton Nekrutenko

Ekaterina Polkh Bérénice Batut

Editor(s)

Overview ⑦ Questions: Q Search Tutorials

Saskia Hiltemann

https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/history/tutorial.html

Supplementary material on deleting data:

https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/download-delete-data/tutori al.html#deleting-data-from-galaxy



Tips & tricks Handle tool errors

ABMS Tools - Handle errors

Station Biologique Roscoff

🗧 Galaxy France	😤 Wo	orkflow Visualize - Shared Data - Help - User - 💼 📰			Using 3%
Tools ☆	NCBI Accession	Download	<u>́н</u>	story 🖌	3 + 🗆 🕯
search tools	Dataset Informatio	on		search datasets	00
1 Upload Data	Number Name	2 NCBI Accession Download on : Log	Er	ror example	
Get Data	Created Filesize	Monday Jan 10th 4:13:29 2022 UTC 109 b	77	2 b	
Send Data Collection Operations	Format File contents	txt contents	4: ad	NCBI Accession Downlo on : Log	• / >
Text Manipulation	History API ID History API ID UUID	11e6e18a7189130t 21c4852849066ede 6882ed13-dee0-4ccc-9b85-b645f5d911f7	3: : D a li	NCBI Accession Downloa ownloaded Files st with 1 item	ad on 🗙
	Tool Parameters	Value	E	2: NCBI Accession Do nload on : Log	• / ×
Convert Formats	Select source for IDs ID List	accession_list AAG19274.1	er A	ror n error occurred with this o	dataset:
FASTA/FASTQ FASTQ Quality Control	Molecule Type File Format	nucleotide FASTA	F	ICBI Entrez returned error ailed to download file with	code 400, a id AAG192
SAM/BAM	failures	Abort with error on first failure		80002?	•
BED VCF/BCF	Job Information Galaxy Tool ID:	toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download/ncbi_acc_download /0.2.7 +galaxy0 [〕	S on al	View details 1: NCBI Accession Down : Downloaded Files	nload 🗙
COMMON GENOMICS TOOLS	Command Line Tool Standard Output	empty NCBI Entrez returned error code 400, are ID(s) AAG19274.1 valid?			
Operate on Genomic Intervals Fetch Alignments/Sequences	Tool Standard Error	Failed to download file with id AAG19274.1 from NCBI			
GENOMICS ANALYSIS	~				

🗧 Galaxy France	🕋 Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 💼		Using 3%
Tools රූ	Downloading accession number: AAG19274.1 Failed to download file with id AAG19274.1 from NCBI	History	S+0¢
search tools	failed.	search datasets	00
🏝 Upload Data		Error example 4 shown, 1 hidden	
Get Data		772 b	Solution
Send Data			
Collection Operations		4: NCBI Accession Do	wnlo 💿 🖋 🗙
GENERAL TEXT TOOLS		uu on . tog	
Text Manipulation		3: NCBI Accession Do : Downloaded Files	wnload on X
Filter and Sort		a list with 1 item	
Join, Subtract and Group		2: NCBI Accession	Do 💿 🖉 🗙
GENOMIC FILE MANIPULATION		wnload on : Log	View data
Convert Formats		2 1: NCBI Accession	Download X
FASTA/FASTQ		a list	,
FASTQ Quality Control			
SAM/BAM			
BED			
VCF/BCF			
Nanopore			
COMMON GENOMICS TOOLS			
Operate on Genomic Intervals			
Fetch Alignments/Sequences			
GENOMICS ANALYSIS			

Tools - Handle errors

Galaxy France

Station Biologique

Roscoff

Tools

search too

🏦 Upload Data

- Get Data
- Send Data
- **Collection Operations**
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- GENOMIC FILE MANIPULATION
- **Convert Formats**
- FASTA/FASTQ
- FASTQ Quality Control
- SAM/BAM
- BED
- VCF/BCF
- Nanopore
- COMMON GENOMICS TOOLS
- Operate on Genomic Interval
- Fetch Alignments/Sequences

*	Workflow Visualize - Shared Data - Help - User - 🖻 🏢	
NCBI Accessi	on Download	History
Dataset Informa	ation	search o
Number	2	Error over
Name	NCBI Accession Download on : Log	Error exa
Created	Monday Jan 10th 4:13:29 2022 UTC	4 shown, 1 h
Filesize	109 b	772 b
Dbkey	?	
Format	txt	4. NCBLA
File contents	contents	ad on : Lo
History Content API II	D 11e6e18a7189130f	
History API ID	21c4852849066ede	3: NCBI A
UUID	6882ed13-dee0-4ccc-9b85-b645f5d911f7	: Downloa a list with
Tool Parameter	s	2 : NC
Input Parameter	Value	wnload o
Select source for IDs	accession_list	error
ID List	AAG19274.1	An error o
Molecule Type	nucleotide	NCBI Ent
File Format	FASTA	Failed to

failures

How to handle download

Job Information



Abort with error on first failure

2 + m **a** 88 datasets ample idden ccession Downlo ۲ XX cession Download on × ded Files item **BI Accession Do** • / × n : Log occurred with this dataset: rez returned error code 400, ar download file with id AAG1927 < > * • • • • 7 ? View details 1: NCBI Accession Download × on : Downloaded Files a list

2

1

T I Code

1

= Galaxy France	🚷 Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 🏢		Using 3%
Tools රූ	Dataset Error Report	History	2 + 🗆 🕇
search tools	An error occurred while running the tool toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download	search datasets	00
🍰 Upload Data	Details	Error example	
Get Data	Execution resulted in the following messages:	4 shown, 1 hidden 772 b	
Send Data	Fatal error: Exit code 1 ()		
Collection Operations	Tool apported the following standard error	4: NCBI Accession Downlo ad on : Log	• • • * ×
GENERAL TEXT TOOLS	foor generated the following standard enor.	3: NCBI Accession Downlo	ad on X
Filter and Sort	Failed to download file with id AAG19274.1 from NCBI	: Downloaded Files a list with 1 item	
Join, Subtract and Group	Troubleshooting	2: NCBI Accession Do	• / ×
GENOMIC FILE MANIPULATION	There are a number of helpful resources to self diagnose and correct problems. Start here: My job ended with an error. What can I do?	wnload on : Log	
Convert Formats	Issue Report	An error occurred with this	dataset:
FASTA/FASTQ	Please provide detailed information on the activities leading to this issue:	NCBI Entrez returned error	r code 400, ai
FASTQ Quality Control		Failed to download life with	n id AAG 192.
SAM/BAM		* 8002?	۲
BED		View or report this error	
VCF/BCF		on : Downloaded Files	nioad X
Nanopore	🕱 Report	a list	
COMMON GENOMICS TOOLS			
Operate on Genomic Intervals			
Fetch Alignments/Sequences			
GENOMICS ANALYSIS			

= Galaxy France	Workflow Visualize - Shared Data - Help - User - 🞓 🏢		Using 3%
Tools රූ	Dataset Error Report	History	3 + 🗆 🗢
search tools	An error occurred while running the tool toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download	search datasets	00
🏦 Upload Data	Details	Error example 4 shown, 1 hidden	
Get Data	Execution resulted in the following messages:	772 b	🗹 📎 🗩
Send Data Collection Operations	Fatal error: Exit code 1 ()	4: NCBI Accession Downlo ad on : Log	● # ×
GENERAL TEXT TOOLS	Tool generated the following standard error:	3: NCBI Accession Downlo	ad on X
Text Manipulation Filter and Sort	Failed to download file with id AAG19274.1 from NCBI	: Downloaded Files a list with 1 item	
Join, Subtract and Group GENOMIC FILE MANIPULATION	Iroubleshooting https://galaxyproject.org/support/tool-error/ There are a number of helpful resources to self diagnose and correct problems. Start here: My job ended with an error. What can I do?	2: NCBI Accession Do wnload on : Log	④ ∦ ×
Convert Formats FASTA/FASTQ FASTQ Quality Control SAM/BAM BED	Issue Report Please provide detailed information on the activities leading to this issue:	An error occurred with this NCBI Entrez returned error Failed to download file with R 🖸 🔗 🕄 💭 ? View or report this error	dataset: code 400, ar i d AAG1927
VCF/BCF Nanopore	Report	on : Downloaded Files a list	nload X
COMMON GENOMICS TOOLS Operate on Genomic Intervals Fetch Alignments/Sequences			
GENOMICS ANALYSIS			

Galaxy France	👫 Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 📰		Using 3%
Tools	Dataset Error Report	History	3 + 🗆 🕈
search tools	An error occurred while running the tool toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download	search datasets	00
🏝 Upload Data	Details	Error example	
Get Data	 Execution resulted in the following messages: 	772 b	
Send Data	Fatal error: Exit code 1 ()	4: NCBI Accession Downlo	• # X
GENERAL TEXT TOOLS	Tool generated the following standard error:	ad on : Log	
Text Manipulation	Failed to download file with id AAG19274.1 from NCBI	3: NCBI Accession Downloa : Downloaded Files	ad on X
Filter and Sort	Travible she at in a	a list with 1 item	
Join, Subtract and Group GENOMIC FILE MANIPULATION	There are a number of helpful resources to self diagnose and correct problems. Start here: My job ended with an error. What can I do?	2: NCBI Accession Do wnload on : Log	● / ×
Convert Formats	Issue Report	error An error occurred with this	dataset:
FASTA/FASTQ	Please provide detailed information on the activities leading to this issue:	NCBI Entrez returned error	code 400, ar
FASTQ Quality Control		Failed to download file with	Id AAG 1927
SAM/BAM		₩ 8 0 0 2 ?	۰
BED		View or report this error	
VCF/BCF		on : Downloaded Files	iload X
Nanopore	The Report	a list	
COMMON GENOMICS TOOLS			
Operate on Genomic Intervals	For usedalaxy fr. create a new request on		
Fetch Alignments/Sequences	https://community.franco.bioinformatique.fr/c/galaxy/		
GENOMICS ANALYSIS			



Hands-on **TOOLS**







https://training.galaxyproject.org /training-material/topics/introduc tion/tutorials/galaxy-intro-101/tu torial.html

Galaxy Training! Analyses	Learning Pathways	Help •	Settings +	Search Tutorials
alaxy Basics for gend	mics			
uthorn:				
🔟 🕤 Saskia Hiltemann 🕅 🕞	Nicola Soranzo	问 📫 Clen	nens Blank	
			Holono Boo	cho
Anton Nekrutenko	Anne Paj		Helena Raso	che
and the second				
Overview				
② Questions:				
Which coding exon has the highest number of sing	le nucleotide polymo	orphisms (SNPs) or	human chron	nosome 22?
© Objectives:				
Familiarize yourself with the basics of Galaxy				
Learn how to obtain data from external sources				
Learn how to run tools				
Learn how histories work				
 Learn how to create a workflow 				
Learn how to share your work				
▼ Time estimation 1 bour				
Level: Introductory				
C Supporting Materials:				
Datasets K Workflows @ FAQs	cordinas - 🌐 Ava	ilable on these Ga	laxies •	
Dublished: Dec 19, 2016			action of the second	
Last modification: May 3, 2024				
Dicense: Tutorial Content is licensed under Creative	Commons Attribution	n 4.0 International	License. The (GTN Framework is lic
under MIT				
PURL: https://gxy.io/GTN:T00186				
Rating: 4.1 (10 recent ratings, 141 all time)				

Pur novt ctone

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	in next steps
Setting the stage: Exons and SNPs Get your workspace	This tutorial aims to familiarize you with the Galaxy user interface. It will teach you how to perform basic tasks such as importing data, running tools, working with histories, creating workflows, and sharing your work.
ready	Comment: Results may vary
Analysis Histories and workflows: A brief	Your results may be slightly different from the ones presented in this tutorial due to differing versions of tools, reference data, external databases, or because of stochastic processes in the algorithms.
Share your work	49

Agenda



# Datatypes: a short overview **DATASET**



- Every Galaxy dataset is associated with a datatype.
- Datasets produced by a tool have their datatype assigned by the tool



7. exc	ms.beu		
,410	regions		
ormat:	<b>bed</b> , gé	nome d	e référence: ?
ipload	ded bed	file	
3 <i>o</i>	<b>6</b> 📖	?	۵ 🔊
isplay	IN IGB V	lew	
isplay isplay .Chrom	in IGB V with IG\ 2.Start	lew / local 3.End	4.Name
isplay isplay .Chrom hr22	in IGB V with IGV 2.Start 16258185	(Iew / local 3.End 16258303	4.Name uc002zlh.1_cds_1_0_chr22_16
isplay isplay Chrom hr22 hr22	in IGB V with IGV 2.Start 16258185 16266928	(Iew / Iocal 3.End 16258303 16267095	4.Name uc002zlh.1_cds_1_0_chr22_16 uc002zlh.1_cds_2_0_chr22_16
Isplay isplay .Chrom hr22 hr22 hr22	in IGB V with IGV 2.Start 16258185 16266928 16268136	/Iew / local 3.End 16258303 16267095 16268181	4.Name uc002zlh.1_cds_1_0_chr22_16 uc002zlh.1_cds_2_0_chr22_16 uc002zlh.1_cds_3_0_chr22_16
Isplay .Chrom hr22 hr22 hr22 hr22	in IGB V with IGV 2.Start 16258185 16266928 16268136 16269872	/Iew / local 3.End 16258303 16267095 16268181 16269943	4.Name uc002zlh.1_cds_1_0_chr22_16 uc002zlh.1_cds_2_0_chr22_16 uc002zlh.1_cds_3_0_chr22_16 uc002zlh.1_cds_4_0_chr22_16
Isplay .Chrom hr22 hr22 hr22 hr22 hr22	in IGB V with IGV 2.Start 16258185 16266928 16268136 16269872 16275206	<pre>/lew /local 3.End 16258303 16267095 16268181 16269943 16275277</pre>	4.Name uc002zlh.1_cds_1_0_chr22_16 uc002zlh.1_cds_2_0_chr22_16 uc002zlh.1_cds_3_0_chr22_16 uc002zlh.1_cds_4_0_chr22_16 uc002zlh.1_cds_5_0_chr22_16

4: reads-fw.fastq 💿 🖋 🗙	
1.2 MB	1: Drosophila_melanogaster.BD 💿 🖋 🗙
format: <b>fastqsanger</b> , génome de référence: <b>dm3</b>	GP6.22.cds.all.fa.gz 30.504 sequences format: <b>fasta.gz</b> , génome de référence: <b>?</b>
uploaded fastqsanger file	uploaded fasta.gz file
© 𝔄 🖩 ? 🗣 ♥	B & B III ?
@SRR031714.1 HWI-EAS299_130MNEAAXX:2:1:785:5	display with IGV local
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	>FBtr0070129 cds chromosome:BDGP6.22:X:656673:657899:1 ATGCCCAGCATCAAGTTGCAATCTTCGGATGAGGAGATCTTTGACACGGATATC GCCAAGTGCTCCGGCACTATCAAGACCATGCTGGAGGACTGCGGCATGGAGGAC
111111111111111111111111111111111111111	GCTCACTACCACAAGGACGACCCCCCAGCCAACGAGGATGATGAGAGGAGGAGGAGGAGGAGGAGGACGACCCCCCAGCCAACGGAGGA
@SRR031714.2 HWI-EAS299_130MNEAAXX:2:1:1703:	< >>
< >>	



Common text formats:

- *txt*: plain text ('.txt')
- *tabular*: tab delimited ('.tab', '.txt', etc.)

wt_37_2 wt_37_3 wt_37_1 TR24|c0_g1_i1 90.00 67.00 85.00 TR2779|c0_g1_i1 186.00 137.00 217.00 TR127|c1_g1_i1 9.0023.00 16.00

• csv: comma-separated values ('.csv')

Year,Make,Model 1997,Ford,E350 2000,Mercury,Couga

r

• *html*: standard language for web pages

```
<!DOCTYPE html>
<html>
<head>
<title>This is a
title</title>
</head>
<body>
Hello world!
</body>
</html>
```

# ABMS Dataset - Datatypes

Station Biologique

📮 Galaxy France	🕋 Workflow Visualize - Shared Data - Help - User - 📻 🏢		Using 3%
Tools		^ History #	2 <b>+ 0 ¢</b>
search tools		search datasets	88
🏝 Upload Data		Trinity example	
Get Data		40.34 KB	
Send Data Collection Operations	Welcome to usegalaxy.fr	4: Trinity on data 2 and d	⊛ # ×
GENERAL TEXT TOOLS	By using this Galaxy instance, we assume that you have read and accept the Term Of Use	ata 1: Gene to transcripts map	
Text Manipulation	For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy	3: Trinity on data 2 and d	⊙ / ×
Filter and Sort	^	ata 1: Assembled Transcri	
Join, Subtract and Group	• 22/07/2021: usegalaxy.fr is now running the <b>release 21.05</b> of Galaxy. Please check	po	
GENOMIC FILE MANIPULATION	the 21.05 user release notes.	2: reads.right.fq	• / ×
Convert Formats	~	1: reads.left.fq	④ ♂ ×
FASTA/FASTQ			
FASTQ Quality Control	D-1		
SAM/BAM	Ack the		
BED	ASK LITC		
VCF/BCF	GalavyCat		
Nanopore	GalaxyCat		
COMMON GENOMICS TOOLS	Can't find a tool on		
Operate on Genomic Intervals	usegalaxy.fr?		
Fetch Alignments/Sequences	Why not search on the IFB		
GENOMICS ANALYSIS	Galaxy Catalog ?		- 2
(			CC CC



Common binary formats:

- data: generic binary format
- *zip, tar*: archives
- *pdf, png, jpg, bmp, tiff, gif*: images
- *rdata*: statistical computing program R
- *bam*, wig, bigwig: sequence alignment

```
Station Biologique
Roscoff
Dataset - Datatypes
```

Sequence file formats:

 fasta: a single-line description with '>', followed by lines of sequence data ('.fasta', '.fas')

```
>sequence1
  atgcgtttgcgtgcatgcgtttgcgtgcatgcgtttgcgtgcatgcgtttgcgtg
  atgcgtttgcgtgc
  >sequence2
  tttcgtgcgtatagtttcgtgcgtatagtttcgtgcgtatagtttcgtgcgtata
  q
  tggcgcggt
fastq: sequence + quality score ('.fastq', '.fq')
  @SEQ ID
  GATTTGGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTT
  Т
  !''*(((((***+)))%%%++)(%%%%).1***-+*''))**55CCF>>>>>CCCCCCC6
  5
  @SEQ ID2
  GATTTGGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTT
  Т
  !''*(((((***+)))%%%++)(%%%%).1***-+*''))**55CCF>>>>CCCCCCC6
```

# ABMS Dataset - Datatypes

Station Biologique Roscoff

📮 Galaxy France	🕋 Workflow Visualize - Shared Data - Help - User - 💼 🏢			Using 3%
Tools		^	History	3 <b>+ 0 \$</b>
search tools			search datasets	00
1 Upload Data			Trinity example	
Get Data			40.34 KB	
Send Data	Welcome to usedalaxy fr			
Collection Operations	Welconne to usegalaxy.n		4: Trinity on data 2 and d	⊙ / ×
GENERAL TEXT TOOLS	By using this Galaxy instance, we assume that you have read and accept the Term Of Use		map	
Text Manipulation	For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy		3: Trinity on data 2 and d	⊙ / X
Filter and Sort		^	ata 1: Assembled Transcri	
Join, Subtract and Group	• 22/07/2021: usegalaxy.fr is now running the <b>release 21.05</b> of Galaxy. Please check		pts	
GENOMIC FILE MANIPULATION	the 21.05 user release notes.		2: reads.right.fq	• / ×
Convert Formats		~	1: reads.left.fq	⊙ / ×
FASTA/FASTQ				
FASTQ Quality Control				
SAM/BAM	Ack the			
BED	ASK LITE			
VCF/BCF	GalavyCat			
Nanopore	GalaxyCat			
COMMON GENOMICS TOOLS	Can't find a tool on			
Operate on Genomic Intervals	usegalaxy.fr?			
Fetch Alignments/Sequences	Why not search on the IFB			
GENOMICS ANALYSIS	Galaxy Catalog ?			
<		~		56 >



Sequence file formats:

#### • gff3, bed, genbank: sequence + annotations

```
track name=pairedReads description="Clone Paired Reads"
bed
       useScore=1
        chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
        chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
        ##gff-version 3
        ctg123 . exon 1300 1500 . + .
        ID=exon00001
        ctg123 . exon 1050 1500 . + .
        ID=exon00002
gttd
                       3000 3902 . + .
        ctq123 . exon
        ID=exon00003
        ##FASTA
        >ctq123
        cttctgggcgtacccgattctcggagaacttgccgcaccattccgcctt
        q
  https://genome.ucsc.edu/FAQ/FAQformat.html#format3
```



• On upload, datatype can be detected or user specified.



58



• Change the datatype of a dataset

📮 Galaxy France	🕋 Workflow Visualize - Shared Data - Help - User - 📧 🏢		Using 3%	
Tools ☆	Edit dataset attributes	History	S+0 🕈	
search tools		search datasets	00	
🏝 Upload Data	Change datatype	Trinity example 4 shown		
Get Data Send Data	New Type	40.34 KB		
Collection Operations	tabular -	4: Trinity on data 2 ar d data 1: Gene to tra		
GENERAL TEXT TOOLS	trk	eripts map	-	
Filter and Sort	trr	7 lines format: <b>tabular</b> , databa	ase: ?	
Join, Subtract and Group	ttl			
GENOMIC FILE MANIPULATION Convert Formats	twobit txt	D)   _	1	
FASTA/FASTQ	uniprotxml 🗸	₿ & ● 2	۰	
FASTQ Quality Control		1 2 TRINITY_DN2_c0_g1 TRINITY_D	12_c0_g1_i1	
BED		TRINITY_DN1_c0_g1 TRINITY_D	1_c0_g1_i1 5_c1_g1_i1	
VCF/BCF		TRINITY_DN0_c0_g1 TRINITY_DN TRINITY_DN0_c0_g2 TRINITY_D	W0_c0_g1_i1 W0_c0_g2_i1	
Nanopore COMMON GENOMICS TOOLS		3: Trinity on data 2 an data 1: Assembled Tra	d <b>⊛</b> ∦×	
Operate on Genomic Intervals	e on Genomic Intervals			



## Dataset - Datatypes

• Many tools will only accept input datasets with the appropriate datatype assigned.

📲 Galaxy France	🕋 Workflow Visualize - Shared Data - Help - User - 🞓 🏢		Using 3%
Tools ☆		<b>History</b>	<b>₽+</b> □ <b>\$</b>
	Version 2.9.1+galaxy2)	search datasets	00
	Are you pooling sequence datasets?	Trinity example	
	Yes	2 shown, 2 deleted	
	Paired or Single-end data?	40.34 KB	
	Paired-end		
		2: reads.right.fq	• # ×
	Lett/Forward strand reads	100 sequences format: <b>fastg</b> , database:	?
	No fasta, fastqsanger or fastqsanger.gz dataset available.	uploaded fastosander fi	10
		GE1DEPAATV188284.11188-18494.	3979/2
	-te	CTCAAATGGTTAATTCTCAGGCTGCAAAT	ATTOSTTCAGGATGGAAG
	(left)	+	
	Right/Reverse strand reads	@61DFRA4XX100204:1:100:10497:	13422/2
	🗘 🗅 No fasta, fastqsanger or fastqsanger.gz dataset available. 🏠 🗲	<	>
		1: reads.left.fq	● / ×
		100 sequences	
		format: <b>fastq</b> , database:	?
	(right)	uploaded fastqsanger fil	le
	Strand specific data	800	۰
	No	@61DFRAAXX100204:1:100:10494:	3878/1 60



## Dataset - Datatypes

• Many tools will only accept input datasets with the appropriate datatype assigned.

							History	2+ <b>0</b> \$
		Trinity de nov Version 2.9.1+	/o assembly of RNA-Seq data (Galaxy -galaxy2)	☆ Favorite	& Versions	▼ Options	search dataset	s 00
		Are you pooling	sequence datasets?				Trinity examp	le
		Yes				•	2 shown, 2 deleted	
		Paired or Sing	le-end data?				40.34 KB	See 10 (1998)
		Paired-end				•		
		1.4/	I store and uses the				2: reads.right.fq	• # ×
		Letty Forward	Na fasta fastarra fastarra		1-		100 sequences format: <b>fastq</b> , dat	abase: ?
			INO Tasta, fastqsanger or fastqsanger	.gz dataset avallar	ole,		unloaded fastos:	anger file
ilter and Sort							Lui 2	
oin, Subtract and G	@SEQ_ID				$\sim$ $\lambda$ mmm $\sim$ m		алана Сладина 27 А.С.П.П. (8284:1:10)	00:10494:3070/2
GENOMIC FILE MANIPU	GAIIIGGG(	JIICAAAG	CAGIAICGAICAAAIA	GIAAIC	CAIIIGI	ICAACICAC		SCTGCAAATATTCGTTCAGGATGGAAG
Convert Formats	± +						cccccccc	
ASTA/FASTQ	' !''*((((;	***+))응응	응++) (응응응응) ₋ 1 * * * <b>-</b>	+*''))*	*55CCF>	>>>>>CCCC	CCC6	90:10497:13422/2
ASTQ Quality Cont	5				00001			>
SAM/BAM	@SEQ ID2						eft.fq	⊙ # ×
RED	GATTTGGG	GTTCAAAG	CAGTATCGATCAAATA	GTAAATC	CATTTGT	тсаастсас	CAGTT nces	
	- <del>-</del>		5			-	stq, dat	abase: :
/CF/BCF	1							



## Hands-on TOOLS (sequence files manipulation)







#### Part 1:

You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.





#### Part 1:

You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.



#### Station Biologique Roscoff ARBASS Run Fastqc

💶 Galaxy France	☆ Workflow Visualize Données partagées   Aide   Authentification et Enregistrement	<b>≈</b> ≜ Ⅲ				
Tools •	FastQC Read Quality reports (Galaxy Version 0.73+galaxy0)	& •				
fastqc ×						
1 Upload Data	Please provide a value for this option. Raw read data from your current history					
Show Sections	D D No fastq, fastq.gz, fastq.bz2, bam or sam dataset available.	- 1 0				
FastQC Read Quality reports	Contaminant list					
<b>fastp</b> - fast all-in-one preprocessing for FASTQ files	Image: Description of the sector of the s	• <b>1</b> 🖻				
FROGS Pre-process merging, denoising and dereplication	tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA					
WORKFLOWS	Adapter list					
All workflows	D D No tabular dataset available.	• 1 🖻				
	List of adapters adapter sequences which will be explicity searched against the library. It should be a tab-delimited sequence. (adapters)	I file with 2 columns: name and				
	Submodule and Limit specifing file					
	D D No txt dataset available.	• <b>1</b> 🖻				
	<ul> <li>a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for the each parameter</li> <li>Disable grouping of bases for reads &gt;50bp</li> <li>No</li> <li>Using this option will cause fastqc to crash and burn if you use it on really long reads, and your plots may end up a warned! (nogroup)</li> <li>Lower limit on the length of the sequence to be shown in the report</li> </ul>	ch submodules warning a ridiculous size. You have been				

^

## Part 1:

You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.

- 1. Look into the description of the tool FastQC, what is its purpose ?
- 2. Does it takes compressed files ? ".gz"
- 3. Which encoding of the file (in FastQC results, basic statistics) ? => edit if needed the datatype of your reads files
- 4. Look at mapping tools, map your reads on genome with RNA STAR.
- 5. What is the default parameter of the option "Maximum ratio of mismatches to mapped length" in RNA STAR ?
- 6. How to map on another reference genome or assembly release ?





You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

- 1. Create new history
- 2. Import CDS and peptide sequences databases
- 3. Import query sequences
- 4. Make BLAST databases
- 5. Run BLAST against the CDS database
- 6. Run BLAST against the protein database



You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

- 1. Create new history
- 2. Import CDS and peptide sequences databases
- 3. Import query sequences
- 4. Make BLAST databases Blast --> NCBI BLAST+ makeblastdb
- 5. Run BLAST against the CDS database
- 6. Run BLAST against the protein database



You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

- 1. Create new history
- 2. Import CDS and peptide sequences databases
- 3. Import query sequences
- 4. Make BLAST databases Blast --> NCBI BLAST+ makeblastdb
- 5. Run BLAST against the CDS database Blast --> NCBI BLAST+ blastn
- 6. Run BLAST against the protein database



You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

- 1. Create new history
- 2. Import CDS and peptide sequences databases
- 3. Import query sequences
- 4. Make BLAST databases Blast --> NCBI BLAST+ makeblastdb
- 5. Run BLAST against the CDS database Blast --> NCBI BLAST+ blastn
- 6. Run BLAST against the protein database Blast --> NCBI BLAST+ blastx



# DATASET COLLECTION



• Problematic: you have a large numbers of datasets to send through the same analysis


- Problematic: you have a large numbers of datasets to send through the same analysis
- Solution 1: select multiple datasets as input



# Dataset collection

# Select multiple datasets as input

Galaxy France	👚 Workflow Visualize - Shared Data - Help - User - 🖻 🏢		Using 3%
Tools රා		History	🕄 🕇 🗆 🏟
search tools	Workflow: Find exons with the highest number of features	search datasets	00
1 Upload Data	Exons	Galaxy initiation -	
Get Data	▲ D 1: exons.bed    L ►	3 shown	
Send Data	Features	10.72 MB	Image: Second
Collection Operations	1 1 3: repeats.bed		
GENERAL TEXT TOOLS	2: snps.bed	3: repeats.bed	● / ×
Text Manipulation	exons.bed	2: snps.bed	⊙ / ×
Filter and Sort		1: exons.bed	⊛ # ×
Join, Subtract and Group	This is a batch mode input field. Separate jobs will be triggered for each		
GENOMIC FILE MANIPULATION	dataset selection.		
Convert Formats	Ø &		
FASTA/FASTQ	Expand to full workflow form.		
FASTQ Quality Control			
SAM/BAM			
BED			
VCF/BCF			
Nanopore			
COMMON GENOMICS TOOLS			
Operate on Genomic Intervals			
Fetch Alignments/Sequences			
GENOMICS ANALYSIS			74
/			/



# **ABMS** Dataset collection

## Select multiple datasets as input

Tools		History	8 + 🗆 🕸
search tools	Successfully invoked workflow <b>Find exons with the highest number of features</b> - 2 times.	search datasets	00
🏝 Upload Data	This workflow will generate results in multiple histories. You can observe progress in the <b>history multi-view</b> .	Galaxy initiation - multiple datasets	
Get Data	View Report 1 🖶	5 shown, 2 deleted, 6 hidden	
Send Data	7 of 7 steps successfully scheduled.	11.61 MB	
Collection Operations	5 of 5 jobs complete.	B.	
GENERAL TEXT TOOLS	► Inputs	12: Top 5 exon IDs	⊛ # ×
Text Manipulation	► Steps	7: Top 5 exon IDs	⊙ # ×
Filter and Sort	View Report 2 🖶	3: repeats.bed	• # ×
oin Subtract and Group	7 of 7 steps successfully scheduled. 5 of 5 jobs complete.		
GENOMIC FILE MANIPULATION	Download BioCompute Object	2: snps.bed	• # X
Convert Formats	► Steps	1: exons.bed	• # ×
FASTA/FASTQ			
ASTQ Quality Control			
SAM/BAM			
BED			
/CF/BCF			
lanopore			
OMMON GENOMICS TOOLS			
Operate on Genomic Intervals			
etch Alignments/Sequences			
ENOMICS ANALYSIS	~		75



- Problematic: you have a large numbers of datasets to send through the same analysis
- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection (any number of datasets bundled as a **single entity**, i.e. minimize clutter)



- Problematic: you have a large numbers of datasets to send through the same analysis
- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection (any number of datasets bundled as a **single entity**, i.e. minimize clutter)
  - Dataset list: set of files of the same type
  - Dataset pairs: pairs of read files (forward, reverse)
  - List of dataset pairs



- Problematic: you have a large numbers of datasets to send through the same analysis
- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection (any number of datasets bundled as a **single entity**, i.e. minimize clutter)
  - Dataset list: set of files of the same type
  - Dataset pairs: pairs of read files (forward, reverse)
  - List of dataset pairs
- Galaxy runs the tool automatically on each dataset using the same settings



# **Dataset collection**

# Creating a collection from datasets in your history. Renaming. Tagging.

Click on Select Items at the top of the history panel



- · Check all the datasets in your history you would like to include
- Click n of N selected and choose Build Dataset List
- Enter a name for your collection
- Click Create collection to build your collection
- · Click on the checkmark icon at the top of your history again







# **ABMS** Dataset collection

# Use a collection as input

n Galaxy France	倄 Workflow Visualize - Shared Data - Help - User - 📧 🏢		Using 3%
Tools	NCBI BLAST + blastx Search protein database with translated	History	*+ •• •
blastx	nucleotide query sequence(s) (Galaxy Version 0.3.3)	search datasets	00
Upload Data	Nucleotide <u>query</u> sequence(s)	blast hands-on 2022	
	🖸 🗘 🗀 12: queries 🔹	8 shown, 10 hidden	
Show Sections	Dataset collection his is a batch mode input field. Separate jobs will be triggered for each dataset	66.36 MB	
JBrowse - Data Directory to Standalone upgrades the bare data directory to a full JBrowse instance	(-query) Subject database/sequences	13: NCBI BLAST+ blastn acr ollection 12	ross c X
Diamond makedb Build database from a FASTA file	BLAST database from your history	a list with 5 items	
NCBI BLAST+ tblastx Search translated	Protein BLAST database	a list with 5 items	^
nucleotide database with translated nucleotide query sequence(s)	D       D       8: protein BLAST database from data 2	11: blastx query5.fa vs 'pr	● / ×
NCBI BLAST+ blastx Search protein database with translated nucleotide query	Query genetic code	m data 2'	
sequence(s)	1. Standard 🗸	10: megablast query5.fa	⊙ / ×
JBrowse genome browser	(-query gencode)	vs 'nucleotide BLAST data base from data 1'	
BLAST XML to tabular Convert BLAST	Type of BLAST		0.44
<b>Diamond</b> alignment tool for short	blastx - Traditional BLASTX to compare translated nucleotide query to protein database	base from data 1	• # X
sequences against a protein database	O blastx-fast - Use longer words for seeding, faster but less accurate	8: protein BLAST databas	• / ×
WORKFLOWS	(-task)	e from data 2	
All workflows	Set expectation value cutoff	2: Drosophila_melanogast	• / ×
	0.001	er.BDGP6.22.pep.all.fa.gz	
	(-evalue) Output format	1: Drosophila_melanogast er.BDGP6.22.cds.all.fa.gz	<b>● / ×</b>
	Tabular (extended 25 columns)		80
<		<b>~</b> III	>



# **ABMS** Dataset collection

# Use a collection as input

Fools	☆ 🔽			History	2+0\$
search tools	3	Executed NCBI BLAST+ blastx and successfully added 5 jobs to the queue.		search datasets	00
🏦 Upload Data		The tool uses 2 inputs: • 12: queries • 8: protein BLAST database from data 2		<b>blast hands-on 2022</b> 9 shown, 10 hidden	
Get Data	^	It produces 5 outputs:		66.36 MB	
Collection Operations GENERAL TEXT TOOLS		<ul> <li>20: blastx query5.fa vs 'protein BLAST database from data 2'</li> <li>21: blastx query4.fa vs 'protein BLAST database from data 2'</li> <li>22: blastx query3.fa vs 'protein BLAST database from data 2'</li> </ul>		19: NCBI BLAST+ blastx a s collection 12 a list with 5 items	acros X
Text Manipulation Filter and Sort		<ul> <li>23: blastx query2.fa vs 'protein BLAST database from data 2'</li> <li>24: blastx query1.fa vs 'protein BLAST database from data 2'</li> <li>You can check the status of queued jobs and view the resulting data by refreshing the History</li> </ul>		13: NCBI BLAST+ blastn a s collection 12 a list with 5 items	acros X
GENOMIC FILE MANIPULATION		panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.		<b>12: queries</b> a list with 5 items	×
Convert Formats FASTA/FASTQ FASTO Quality Control		Tool recommendation		11: blastx query5.fa vs 'protein BLAST databas e from data 2'	● / ×
SAM/BAM BED		following/recommended tools. The recommended tools are shown in the decreasing order of their scores predicted using machine learning analysis on workflows. Therefore, tools at the top may be more useful than the ones at the bottom. Please click on one of the following (recommended tools to open its definition		10: megablast query5.f a vs 'nucleotide BLAST database from data 1'	● # ×
VCF/BCF Nanopore		following/recommended tools to open its delinition.		9: nucleotide BLAST da tabase from data 1	⊛ # ×
COMMON GENOMICS TOOLS				8: protein BLAST datab ase from data 2	⊕ # ×
Fetch Alignments/Sequences GENOMICS ANALYSIS		NCBLBLAST+ blasty		2: Drosophila_melanog aster.BDGP6.22.pep.all. fa.gz	• / ×
			~		0.



# ARMS Dataset collection

## Use dataset collection tools:

📮 Galaxy France	🎢 Workflow Visualize - Shared Data - Help - User - 🕋 📰		Using 3%
Tools		History	2 <b>+ 0 ¢</b>
search tools		search datasets	00
1 Upload Data		blast hands-on 2022	
		8 shown, 10 hidden	
Get Data		66.36 MB	
Send Data	https://training.galaxyproject.org/training_material/tor	vice/galaxy_int	arface
Collection Operations		<u>ncs/galaxy-inte</u>	
GENERAL TEXT TOOLS	/tutorials/collections/tutorial.html#collection-operation	<u> </u>	
Text Manipulation	For any questions of support, community,cluster,mance, bioimormatique.in/c/galaxy	12: queries	×
Filter and Sort		a list with 5 items	
Join, Subtract and Group	• 22/07/2021: usegalaxy.fr is now running the <b>release 21.05</b> of Galaxy. Please check	11: blastx query5.fa vs 'p	• • • ×
GENOMIC FILE MANIPULATION	the 21.05 user release notes.	otein BLAST database fro m data 2'	
Convert Formats	~	10: magphlact quary5 fa	O A Y
FASTA/FASTQ		vs 'nucleotide BLAST data	
FASTQ Quality Control		base from data 1'	
SAM/BAM	Ack the	9: nucleotide BLAST data	⊙ ∥ ×
BED	ASK LITE	base from data 1	
VCF/BCF	GalaxyCat	8: protein BLAST databas e from data 2	④ ∅ ×
Nanopore	GalaxyCat	2: Drosophila melanogas	• • / ×
COMMON GENOMICS TOOLS	Can't find a tool on	er.BDGP6.22.pep.all.fa.gz	- F - 1
Operate on Genomic Interva	Is usegalaxy.fr?	1: Drosophila_melanogas	: • / ×
Fetch Alignments/Sequences	Why not search on the IFB	er.BDGP6.22.cds.all.fa.gz	
GENOMICS ANALYSIS	Galaxy Catalog ?		82
1			\ \



#### Use-case: RNA-seq analysis





#### FYI: Interesting resource on dataset collections



https://training.galaxyproject.org /training-material/topics/galaxy-i nterface/tutorials/collections/tut orial.html

Using dataset collections

Galaxy Training!

Authors: 🚇 Anton Nekrutenko ) 😫 Add Contributions!

#### Overview

#### ⑦ Questions:

• How to manipulate large numbers of datasets at once?

#### Objectives:

- Understand and master dataset collections
- **Time estimation:** 30 minutes
- 🕿 Level: Intermediate 🚘 🞓 📼

#### **Supporting Materials:**

- 🗘 Datasets 🛛 🗖 GTN Video Library 🔻
- Last modification: Nov 16, 2021

Ja License: Tutorial Content is licensed under Creative Commons Attribution 4.0 International License The GTN Framework is licensed under MIT

Getting data Creating a paired

dataset collection Processing data

organized as a collection

Collection

operations Frequently Asked

**Ouestions** Feedback

Citing this Tutorial

Here we will show Galaxy features designed to help with the analysis of large numbers of samples. When you have just a few samples - clicking through them is easy. But once you've got hundreds - it becomes very annoying. In Galaxy we have introduced Dataset collections that allow you to combine numerous datasets in a single entity that can be easily manipulated.

Languages

Help

Extras

Search Tutorials

 $\odot$ 

# Getting data

First, we need to upload datasets. Cut and paste the following URLs to Galaxy upload tool (see a Q **Tip** on how to do this below).

https://zenodo.org/record/5119008/files/M117-bl 1.fq.gz https://zenodo.org/record/5119008/files/M117-bl_2.fq.gz https://zenodo.org/record/5119008/files/M117-ch_1.fq.gz https://zenodo.org/record/5119008/files/M117-ch_2.fq.gz https://zenodo.org/record/5119008/files/M117C1-bl_1.fq.gz https://zonada ang/pacand/E119008/files/M11701_hl 2 fg g



# Hands-on COLLECTION



# Collection – Hands-on

# Part 2:



You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

# Case of Drosophila melanogaster (dm)

- 1. Create new history
- 2. Import CDS and peptide sequences databases
- 3. Import query sequences
- 4. Make BLAST databases
- 5. Run BLAST against the CDS database
- 6. Run BLAST against the protein database
- 7. Create a dataset list with all the query sequences
- 8. Run BLAST against the CDS database on the dataset list



# WORKFLOW



🗅 Exons	¢	<b>&gt;</b>	×	🔑 bedtools	© →	×	🄑 Datamash	Ċ.	→ ×	🔑 Sort	Ċ -	×	🔑 Select firs	st 🗘	÷	×
output (input)	C	<b>→</b>	×	File A to interse File B to interse output (input)	als ect with B ect with A ut, bed)	0	Datamash o dataset(s) (tabi	lataset n input ular)		Sort Query	t)		from out_file1	(input) → >		//
output (input)												Cor aga	mpare ainst out file1 (input)			

# What?

• A sequence of tool operations and parameters



🗅 Exons	¢	<b>&gt;</b>	×	🔑 bedtools	@ →	×	🎤 Datamash	Ċ.	→ ×	8	🖋 Sort	¢ ÷	×		🏓 Select first	¢	→ ×
output (input)			0	Intersect interval         File A to inters         File B to inters	als ect with B ect with A		<ul> <li>Input tabular of</li> <li>Datamash of</li> <li>dataset(s) (tab</li> </ul>	lataset n input ular)			Sort Query	t)	6		from dout_file1 (inp	ut)	
Features	¢	÷	×	output (inpu	ut, bed)								<b>پ</b> two	Com Data	pare 🖸 <del>:</del> asets	×	Ì
														ompai	re		
														out_f	file1 (input)		0

# Why?

- Re-run the same analysis on different input data sets
- Change parameters before re-running a similar analysis
- Make use of the workflow job scheduling (jobs are submitted as soon as their inputs are ready)
- Share workflows for publication and with the community



🗅 Exons	¢	<b>&gt;</b>	×	🔑 bedtools	→	×	1	🔑 Datamash	₫ →	×	- 1	🗲 Sort	0 -	→ ×		🎤 Select first	0 ->	×
output (input)			0	Intersect interval         File A to inters         File B to inters	als ect with B ect with A		0	Input tabular d Datamash o dataset(s) (tabu	ataset n input ılar)			Sort Query			0	from ✓ out_file1 (inpu	ıt)	
Features output (input)	Q	→	×	🗆 output (inpu	ut, bed)	•	J								€ Cor two Da	npare <b>₫ →</b> itasets	×	
														0	Comp	are		
														ľ	out ⊲	file1 (input)		

# How?

- Extracted from a history
- Built manually by adding and configuring tools using the canvas
- Imported using an existing shared workflow



# Our workflow with Galaxy





# From history

search tools	The following list contains each tool that was run to cre you wish to include in the workflow. Tools which cannot be run interactively and thus cannot	ate the	datasets in your current history. Please select thos corporated into a workflow will be shown in gray.	e that	History Actions Copy	bry options
🏦 Upload Data	Workflow name				Share or Publish	
Get Data	Workflow constructed from history 'tuto-galaxy-intro-	101'		1	Show Structure Extract Workflow	1 🌑 🗩
Send Data	Create Workflow Check all Uncheck all				Set Permissions	
Collection Operations	Tool		History items created	_	Make Private	) <i>(</i> * ×
GENERAL TEXT TOOLS			1 Evons hed		Resume Paused Jobs	
Text Manipulation			Treat as input dataset		Dataset Actions	) <b>/ X</b>
Filter and Sort			Exons.bed		Copy Datasets	) 🖉 🗙
Join, Subtract and Group	Data Fetch This tool cannot be used in workflows	Þ		_	Collapse Expanded Datasets	) 🖋 🗙
GENOMIC FILE MANIPULATION	This coor currier be used in workflows		2 SNPs.bed		Unhide Hidden Datasets	X
Convert Formats			Treat as input dataset		Delete Hidden Datasets	5
FASTA/FASTQ			SNPs.bed		Purge Deleted Datasets	) # ×
FASTQ Quality Control	bed ools Intersect intervals		3 bedtools Intersect intervals on data 2 and		Downloads	) <i>A</i> X
SAM/BAM	✓ Include "bedtools Intersect intervals" in workflow		data 1		Export Tool Citations	
BED.					Export History to File	
VCE/BCE	Datamash		4 Datamash on data 3		Beta Features	
Napapara					Use Beta History Panel	
COMMON GENOMICS TOOLS	Sort Gurrent history le "Sort" in workflow		5 Sort on data 4	~	92	>



# From history

search tools  search tools  tools  tools  tools  search tools  tools  tools  worr  Get Data  Get Data  Collection Operations  GENERAL TEXT TOOLS  Text Manipulation  Filter and Sort  Join, Subtract and Group  GENOMIC FILE MANIPULATION	which cannot be run interactively and thus cannot <b>kflow name</b> kflow constructed from history 'tuto-galaxy-intro- tate Workflow Check all Uncheck all	: be inc	orporated into a workflow will be shown in gray.  History items created  1 Exons.bed	Hi C S S E S M R	istory Actions Copy Share or Publish Show Structure Extract Workflow Set Permissions Make Private Sesume Paused Jobs	Pry options
Lupload Data   Get Data   Get Data   Send Data   Collection Operations   GENERAL TEXT TOOLS   Text Manipulation   Filter and Sort   Join, Subtract and Group   GENOMIC FILE MANIPULATION	which cannot be run interactively and thus cannot <b>kflow name</b> kflow constructed from history 'tuto-galaxy-intro- ate Workflow Check all Uncheck all	: be inc 101'	History items created	C S E S M R	Copy Share or Publish Show Structure Extract Workflow Set Permissions Make Private Sesume Paused Jobs	) () ×
Upload Data    Get Data   Send Data   Collection Operations   GENERAL TEXT TOOLS   Text Manipulation   Filter and Sort   Join, Subtract and Group   GENOMIC FILE MANIPULATION	kflow name rkflow constructed from history 'tuto-galaxy-intro- rate Workflow Check all Uncheck all	101'	History items created	S S S S N N R	Share or Publish Show Structure Extract Workflow Set Permissions Make Private Sesume Paused Jobs	) () () ) () () ) () ()
Get DataSend DataCollection OperationsGENERAL TEXT TOOLSText ManipulationFilter and SortJoin, Subtract and GroupGENOMIC FILE MANIPULATION	kflow constructed from history 'tuto-galaxy-intro- ate Workflow Check all Uncheck all	101'	History items created	S – S – R	Show Structure Extract Workflow Set Permissions Make Private Sesume Paused Jobs	
Send Data Collection Operations GENERAL TEXT TOOLS Text Manipulation Filter and Sort Join, Subtract and Group GENOMIC FILE MANIPULATION	ate Workflow Check all Uncheck all		listory items created	E - S - R	Extract Workflow Set Permissions Make Private Resume Paused Jobs	) # X
Collection Operations  GENERAL TEXT TOOLS  Text Manipulation  Filter and Sort Join, Subtract and Group  GENOMIC FILE MANIPULATION			listory items created	S	Set Permissions Make Private Resume Paused Jobs	) @ X
Collection Operations  GENERAL TEXT TOOLS  Text Manipulation  Filter and Sort Join, Subtract and Group  GENOMIC FILE MANIPULATION			listory items created 1 Exons.bed	R	Make Private lesume Paused Jobs	
GENERAL TEXT TOOLS Text Manipulation Filter and Sort Join, Subtract and Group GENOMIC FILE MANIPULATION			1 Exons.bed	- R	Resume Paused Jobs	
Text Manipulation Filter and Sort Join, Subtract and Group GENOMIC FILE MANIPULATION						
Filter and Sort Join, Subtract and Group GENOMIC FILE MANIPULATION			Treat as input dataset	Da	ataset Actions	
Join, Subtract and Group This t GENOMIC FILE MANIPULATION	ta Eatch		Exons.bed	C	Copy Datasets	) <b>(*</b> X
GENOMIC FILE MANIPULATION	cool cannot be used in workflows				Collapse Expanded Datasets	) / X
			2 SNPs.bed	U	Jnhide Hidden Datasets	X
Convert Formats			Treat as input dataset	C	)elete Hidden Datasets	
FASTA/FASTQ			SNPs.bed	Р	urge Deleted Datasets	) / X
FASTQ Quality Control	Itools Intersect intervals		3 bedtools Intersect intervals on data 2 and	D	ownloads	
SAM/BAM	clude "bedtools Intersect intervals" in workflow		data 1	- Е	xport Tool Citations	
RED				E	xport History to File	
Dat	amash		4 Datamash on data 3	Be	eta Features	
VCF/BCF	iclude "Datamash" in workflow			U	Jse Beta History Panel	
Nanopore	t					



# Workflow manager

💶 Galaxy France	😭 Workflow Visualize 🕶 S	Shared Data 🔻 Help 🔻 User 🔻 💼	Using 3%
Tools ☆	Search Workflows		History 🖯 🕂 🗖 🌣
search tools	Search worknows		search datasets ? 8
🏦 Upload Data	Name 🔶 Tags	🔷 Updated 🔷 Sharing 🖨 Bookmarked 🕯	tuto-galaxy-
Get Data	<ul> <li>Find exons with the highest number of features</li> </ul>	7 days	7 shown
Send Data Collection Operations	<ul> <li>Galaxy initiation training</li> <li>2020 - tools hands-on 1234</li> </ul>	a year ago	
GENERAL TEXT TOOLS Text Manipulation	<ul> <li>▼ Galaxy initiation training</li> <li>2020 - tools hands-on TEST3</li> </ul>	a year ago	Datasets on dat a 6 and data 1
Filter and Sort Join, Subtract and Group	<ul> <li>✓ Galaxy initiation training</li> </ul>	a year ago	6: Select first o 💿 🥒 🗙 n data 5
GENOMIC FILE MANIPULATION	2020 - tools hands-on TEST2 ▼ Galaxy initiation training	a vear ago	5: Sort on data 💿 🤌 🗙 4
FASTA/FASTQ	2020 - tools hands-on TEST		4: Datamash on 💿 🖉 🗙 data 3
FASTQ Quality Control			3: bedtools Inte



# Workflow manager

iools රු	Search Workflows			+ Create	History	8+ <b>0</b> \$
search tools					search datasets	88
🍰 Upload Data	Name	Tags	Updated 🗘 Sha	ring 🖨 🛛 Bookmarked 🛱	tuto-galaxy-	
Get Data	▼ Find exons with the number of features	highest 🌑	7 days ago		7 shown	
Send Data	🕑 Edit	•			8.53 MB	
Collection Operations	🕒 Сору	1g 🔶	a year ago		7: Compare two	
ENERAL TEXT TOOLS	🛓 Download			_	Datasets on dat	
ext Manipulation	≁ Rename	ng 🌑	a year ago		a 6 and data 1	
ilter and Sort	<b>&lt;</b> Share	515			6: Select first o	• / ×
oin, Subtract and Group	View	ng 🌑	a year ago		n data 5	
ENOMIC FILE MANIPULATION	<u> </u> Delete	ST2			5: Sort on data	• 🖉 🗙
Convert Formats	▼ Galaxy initiation trai	ning 🕒	a year ago		4	
Δ STΔ /FΔ STO	2020 - tools hands-on	TEST			4: Datamash on	• / ×
A31A/1A31Q					data 3	



# Edit a workflow: attributes





# Edit a workflow: drag and drop





# Edit a workflow: drag and drop





# Edit a workflow: delete a noodle





# Edit a workflow: add a tool





# Edit a workflow: add a noodle





# Edit a workflow: hide intermediate steps





# Edit a workflow: edit step





# Edit a workflow: rename the outputs





📮 Galaxy France	🚷 Workflov	v Visualize 🔻 Shared Data	a 🕶 Help 🕶 User 🕶 💼		Using 3%
Find exons with the highest ne	umber of features			d 🗹	• •
$\mathcal{F}$ bedtools $\mathbb{C} \rightarrow \times$		🗲 Sort 🗳	→ ×	<ul> <li>Select first lines from a dataset (Galaxy Version 1.0.1)</li> </ul>	
Intersect intervals	Input tabular dataset	Sort Query	from	Label	
File B to intersect with A	Datamash on input	outfile (input)	out_file1 (input)		
output (input, bed)				Add a step label. Step Annotation	
g-			Compare against		
			out_file1 (input)	0	10.
				Add an annotation or not step. Annotations are ava workflow is viewed.	tes to this ilable when a
			=	Select first	
- 100% +				lines	~














#### Save





<u> –</u> Galaxy France	🖀 Workflow V	′isualize <del>▼</del> Shared Da	ita 🔻 Help 🔻 User 👻 ె			Using 3%
Tools     ☆       search tools     👀	Search Workflows			+ Create	History search datasets	\$ + 0 ¢ 8 8
1 Upload Data	Name	Tags	Updated Sharing	Bookmarked 🖨	tuto-galaxy- intro-101	
Get Data	number of features	•	ago		7 shown 8.53 MB	
Collection Operations GENERAL TEXT TOOLS	<ul> <li>Galaxy initiation training</li> <li>2020 - tools hands-on 1234</li> </ul>	۲	a year ago		7: Compare two Datasets on dat	• / ×
Text Manipulation Filter and Sort	<ul> <li>Galaxy initiation training</li> <li>2020 - tools hands-on TEST3</li> </ul>	۲	a year ago		a 6 and data 1 6: Select first o	
Join, Subtract and Group	<ul> <li>Galaxy initiation training</li> <li>2020 - tools hands-on TEST2</li> </ul>	۲	a year ago		n data 5 5: Sort on data	• / ×
Convert Formats FASTA/FASTO	<ul> <li>Galaxy initiation training</li> <li>2020 - tools hands-on TEST</li> </ul>	۲	a year ago		4 4: Datamash on	• / ×
FASTQ Quality Control					3: bedtools Inte	⊙ / × 、



🔁 Galaxy France	Workflow Visualize Données partagées 🗸 Aide 🗸 Utilisateur 🍸 💼 🌲 🏢		Using 40%
UseGalaxy.fr will be undergoing maintenance of	November 28th in the afternoon. Running jobs will not be stopped. Thank you for your understanding.		
Tools ☆		History	S+0 ¢
search tools	Workflow: ABiMS - Galaxy initiation - Find exons with the highest number of features	Rechercher des données	00
1 Upload Data	Exons	ABiMS - Galaxy initiat 2022 - datasets - publ	tion ic
Get Data	1: exons.bed - 1: @	10 shown	
Send Data	Features	29.08 MB	
Collection Operations	□ ① ③ 3: repeats.bed	6: Drosophila, melanoga	OAX
GENERAL TEXT TOOLS	Number of evens	ster.BDGP6.22.cds.all.f	0 0 11
Text Manipulation		a.gz	
Filter and Sort	D Expand to full workflow form	5: reads-rv.fastq	● / ×
Join, Subtract and Group	Expand to full worknow form.	4: reads-fw.fastq	
GENOMIC FILE MANIPULATION		3: repeats.bed	• / ×
Convert Formats			
FASTA/FASTQ		2: snps.bed	• / ×
FASTQ Quality Control	*	1: exons.bed	• / ×
<			>



🔁 Galaxy France	👫 Workflow Visualize Données partagées 🛪 Aide 🛪 Utilisateur 🛪 📄 🛄		Using 40%
UseGalaxy.fr will be undergoing maintenance on No	vember 28th in the afternoon. Running jobs will not be stopped. Thank you for your understanding.		
Tools ☆ ≔		History	S+0 \$
search tools	Workflow: ABiMS - Galaxy initiation - Find exons with the highest number of features	Rechercher des données	00
🏦 Upload Data	Exons	ABiMS - Galaxy initiat 2022 - datasets - publ	ion ic
Get Data	1: exons.bed - 1: @	10 shown	
Send Data	Features	29.08 MB	
Collection Operations	🕒 🗘 3: repeats.bed - 🕹 🗁	6: Drosophila melanoga	• A X
GENERAL TEXT TOOLS	Number of evens	ster.BDGP6.22.cds.all.f	0 y
Text Manipulation		a.gz	
Filter and Sort	5 Expand to full workflow form	5: reads-rv.fastq	⊙ / ×
Join, Subtract and Group	Expand to full worknow form.	4: reads-fw.fastq	• / ×
GENOMIC FILE MANIPULATION		3: repeats.bed	×
Convert Formats			
FASTA/FASTQ		2: snps.bed	• # X
FASTQ Quality Control	*	1: exons.bed	• / ×
<			>



🔁 Galaxy France	🎢 Workflow Visualize Données partagées 🛪 Aide 🛪 Utilisateur 🛪 💼 🌲 🏢		Using 40%
UseGalaxy.fr will be undergoing maintenance or	November 28th in the afternoon. Running jobs will not be stopped. Thank you for your understanding.		
Tools ☆ #		History	3+00
search tools	Workflow: ABiMS - Galaxy initiation - Find exons with the highest number of features	Rechercher des données	00
🏦 Upload Data	Exons	ABiMS - Galaxy initiation 2022 - datasets - public	on S
Get Data	1: exons.bed - 1:	10 shown	
Send Data	Features	29.08 MB	
Collection Operations	C C 3: repeats.bed → L ▷	6: Drosophila melanoga	• A ×
GENERAL TEXT TOOLS		ster.BDGP6.22.cds.all.f	
Text Manipulation		a.gz	
Filter and Sort		5: reads-rv.fastq	● / ×
Join, Subtract and Group	Expand to full worknow form.	4: reads-fw.fastq	• # ×
GENOMIC FILE MANIPULATION		3: repeats.bed	• # X
Convert Formats			
FASTA/FASTQ		2: snps.bed	• # ×
FASTQ Quality Control	•	1: exons.bed	• / ×
<			



🗧 Galaxy France	🕋 Workflow Visualize 🕶 Shared Data 🕶 Help 👻 User 🖛 📰 🏢		Using 3%
Tools		History	3 + 🗉 🕈
search tools	Successfully invoked workflow Find exons with the highest number of features.	search datasets	88
🏦 Upload Data	pane, if this has not already happened automatically.	Galaxy initiation - workflow	
Get Data	Invocation 1	7 shown	
Send Data	7 of 7 steps successfully scheduled.	3.71 MB	
Collection Operations	0 of 5 jobs complete		
GENERAL TEXT TOOLS	► Inputs ► Steps	• 7: Top 5 exons	۷ 🖉
Text Manipulation		6: Top 5 exon IDs	A X
Filter and Sort		<b>0</b> 5: Sort on data 4	(a) a x
Join, Subtract and Group		• 4: Datamash on data 3	(a) 🖉 🗙
GENOMIC FILE MANIPULATION		• 2. hadtaala Interreat int	
Convert Formats		ervals on data 2 and data 1	
FASTA/FASTQ		2: repeats.bed	• / ×
FASTQ Quality Control		1: Exons.bed	• / ×
SAM/BAM			
BED			
VCF/BCF			
Nanopore			
COMMON GENOMICS TOOLS		11	.4



Galaxy France	🎢 Workflow Visualize 🗸 Shared Data 🕇 Help 🕇 User 🕇 💼		Using 3%
Tools		History	£+ <b>⊡</b> ‡
search tools	Successfully invoked workflow Find exons with the highest number of features.	search datasets	00
🏦 Upload Data	You can check the status of queued jobs and view the resulting data by refreshing the History pane, if this has not already happened automatically.	Galaxy initiation - workflow	
Get Data	View Report 1 🖶	4 shown, 3 hidden	
Send Data	7 of 7 steps successfully scheduled.	5.97 MB	
Collection Operations	5 of 5 jobs complete.		
	Download BioCompute Object ► Inputs	7: Top 5 exons	• / ×
Text Manipulation	► Steps	6: Top 5 exon IDs	● # ×
Filter and Sort		2: repeats.bed	• / ×
Join, Subtract and Group		1: Exons.bed	● / ×
GENOMIC FILE MANIPULATION			
Convert Formats			
FASTA/FASTQ			
FASTQ Quality Control			
SAM/BAM			
BED			
VCF/BCF			
Nanopore			
		10	115



Impossible (until now)

Workflow

Station Biologique Roscoff





## Last but not least! SHARE



## Share everything you do in Galaxy - histories, workflows, and visualizations

- Directly using <u>a Galaxy account's email address</u> on the same instance
- Using a <u>web link</u>, with anyone who knows the link
- Using a web link and publishing it to make it <u>accessible to</u> <u>everyone</u> from the Shared Data menu

Sharing your history:

https://training.galaxyproject.org/training-material/faqs/galaxy/histories_sharing.html



## Share everything you do in Galaxy - histories, workflows, and visualizations

- Directly using <u>a Galaxy account's email address</u> on the same instance
- Using a <u>web link</u>, with anyone who knows the link
- Using a web link and publishing it to make it <u>accessible to</u> <u>everyone</u> from the Shared Data menu

Sharing your history:

https://training.galaxyproject.org/training-material/faqs/galaxy/histories_sharing.html

• Sharing tools, scripts and wrappers: toolshed



#### History

Station Biologique Roscoff Share

<u>=</u> Galaxy France	👚 Workflow Visualize 🕶 Shared Data 👻 Help 👻 User 👻 📰		Using 3%
Tools	와 Share or Publish History `Galaxy initiation - workflow`	History	ଟ+ <b>ଅ</b> ¢
search tools	<ul> <li>Make History accessible</li> <li>Make History publicly available in Published Histories</li> <li>This History is currently accessible via link.</li> </ul>	search d Galaxy ir workflov	History Actions Copy Share or Publish
Get Data Send Data	Anyone can view and import this History by visiting the following URL:	4 shown, 3 h 5.97 MB	Show Structure Extract Workflow Set Permissions
Collection Operations GENERAL TEXT TOOLS	Share History with Individual Users	7: Top 5 ex	Make Private Resume Paused Jobs
Text Manipulation Filter and Sort	You have not shared this History with any users.	6: Top 5 ex 2: repeats.	Dataset Actions
Join, Subtract and Group GENOMIC FILE MANIPULATION	Share with a user	1: Exons.bo	Copy Datasets Collapse Expanded Datasets
Convert Formats FASTA/FASTQ			Delete Hidden Datasets Purge Deleted Datasets
FASTO Ouality Control	➤	Ш	Downloads



#### Workflow





#### Workflow: mode





#### • Get shared histories

🗧 Galaxy France	👚 Workflow Visi	ualize 🔻 Shared Data 🔻 Help 🕇	User 🕶 👔 🏭		Using 3%
Tools ☆	Histories shared	with you by others	Logged in as lgueguen@sb-roscoff.fr	tory	2+ <b>0</b> \$
search tools		man you by ounces	Preferences	earch datasets	88
	Name Datasets	Created Last Updated	1 Custom Builds	avy initiation -	
Individual	No items		Logout	rkflow	
Get Data			Datasets	own, 3 hidden	
Send Data			Histories	МВ	
Collection Operations			Histories shared with me		
			Account and saved data	on 5 avons	

🗧 Galaxy France	👚 Workflow Visualize 🔻	Shared Data - Help - I	User 🔻 📂 🏢		Using 3%
Tools ర్లో	Published Histories	Data Libraries	^	History	<b>∂+□</b> ‡
search tools		Histories		search datasets	88
1. Upload Data	search name, annotation, owner, and Advanced Search	Workflows Visualizations		Galaxy initiation -	
Public	Name	Pages	Owner	A shown 3 hidden	
Get Data	Rustenholz - DE Sangiovese		dc244f268e004c7d9977de	4 shown, 5 hidden	
Send Data			dc244f2680004c7d9977dc	5.97 MB	
Collection Operations	Rustenholz - DE Barbera Refosco		ucz4412088004C70997708		



#### • Get shared workflows

📮 Galaxy France	🔗 Workflow Visualiz	ze Shared Data ▼ Help ▼ Us	er 🕶 🚖 🏢			Using 3%
Tools	E Search Workflows			+ Create	History search datasets	2+0¢
🏦 Upload Data	Name	Tags	Updated 🗘 Sharing	🗘 Bookmarked 🛱	Galaxy initiation -	
GetIndividua	▼ Find exons with the highest number of features	۲	3 days ago		4 shown, 3 hidden	
Send Data Collection Operations	▼ Galaxy initiation training 2020 - tools hands-on 1234	•	a year ago		5.97 MB	
GENERAL TEXT TOOLS	▼ COVID-19: variation analysis on ARTIC PE data v0.2	COVID-19 × ARTIC ×	6 months 🌐 <		6: Top 5 exons	
Filter and Sort	on the RNASeq workflow for paired-end data using the same steps for mapping and variant calling, but adds extra logic for trimming ARTIC primer sequences off reads	covid19.galaxyproject.org × emergen ×	ayu 🔛		2: repeats.bed	• # ×

💶 Galaxy France		👚 Workflow Visualize 🔻	Shared Data 🔻 Help	🕶 User 🕶 💼			Using 3%
Tools		Published Workflows	Data Libraries		^	History	<b>₽+□ \$</b>
search tools	8		Histories			search datasets	88
		search name, annotation, owner, and	Workflows				
🏦 Upload Data		Advanced Search	Visualizations	Access published resources		Galaxy initiation -	
<b>PUDIIC</b> Get Data	^	Name	Pages	Annotation		4 shown, 3 hidden	
Send Data		Bordeaux-Global-ESI-pos 🔻				5.97 MB	



#### Import shared



Author

yquitton44

Filter on Acquisition Numbers = *Empty.* Filter on Retention Time = *Empty.* 

Spectra Filters:

125



## Hands-on WORKFLOW AND SHARE





Workflow a

Station Biologique

Roscoff

https://training.galaxyproject.org /training-material/topics/introduc tion/tutorials/galaxy-intro-101/tu torial.html

#### Galaxy Basics for genomics

Galaxy Training! Analyses

Introduction to Galaxy



Learning

Pathways

Exons and SNPs

ready

Analysis

Histories and

workflows: A brief introduction

Share your work

Get your workspace

#### Overview 0 (?) Questions: Which coding exon has the highest number of single nucleotide polymorphisms (SNPs) on human chromosome 22? Objectives: · Familiarize yourself with the basics of Galaxy · Learn how to obtain data from external sources · Learn how to run tools Learn how histories work · Learn how to create a workflow Learn how to share your work Time estimation: 1 hour C Level: Introductory C I I I C Supporting Materials: Datasets Korkflows Available on these Galaxies • ⑦ FAQs Recordings • Published: Dec 19, 2016 Last modification: May 3, 2024 Dicense: Tutorial Content is licensed under Creative Commons Attribution 4.0 International License. The GTN Framework is licensed under MIT PURL: https://gxy.io/GTN:T00186 A Rating: 4.1 (10 recent ratings, 141 all time) -O- Revision: 131 This tutorial aims to familiarize you with the Galaxy user interface. It will teach you how to perform basic tasks Setting the stage:

Search Tutorials

Settings

Help

Comment: Results may vary

Your results may be slightly different from the ones presented in this tutorial due to differing versions of tools, reference data, external databases, or because of stochastic processes in the algorithms.

such as importing data, running tools, working with histories, creating workflows, and sharing your work.

Agenda



## CONCLUSION

## Key points on Galaxy

- Easy-to-use graphical user interface for often complex command-line tools
- Keeps a full record of your analysis in a history
- Workflows enable you to repeat your analysis on different data
- Galaxy can connect to external sources for data import and visualization purposes
- Galaxy provides ways to share your results and methods with others



- Regularly free space with "Purge Deleted Datasets "or" Delete Permanently "
- On usegalaxy.fr, if you need support (issue, request for a tool...), please open a subject on <u>https://community.france-bioinformatique.fr/</u> <u>c/galaxy/8</u>



## Now, choose your favorite Galaxy! FIND A GALAXY SERVER



#### 130+ platforms for using Galaxy: <u>https://galaxyproject.org/use</u>

[Servers, clouds, deployable resources]



#### 130+ platforms for using Galaxy: <u>https://galaxyproject.org/use</u>

[Servers, clouds, deployable resources]

#### **GENERAL PURPOSE:**

<u>usegalaxy.fr</u>, usegalaxy.org, usegalaxy.eu, usegalaxy.org.au



#### 130+ platforms for using Galaxy: <u>https://galaxyproject.org/use</u> [Servers, clouds, deployable resources]

#### **GENERAL PURPOSE:**

<u>Usegalaxy.fr</u>, usegalaxy.org, usegalaxy.eu, usegalaxy.org.au

	Quota		Concurrent jobs		
	registered	unregistered	registered	unregistered	
usegalaxy.fr	100 GB	10 GB	25	2	
usegalaxy.org	250 GB	5 GB	6	1	
usegalaxy.eu	250 GB				

## ABMS Galaxy France: <u>https://usegalaxy.fr</u>

1172

Station Biologique Roscoff

	Tools Search tools	☆ III 3	INSTITUT FRANÇAIS DE BIOINFORMATIQUE		
	1 Upload Data				
<b>_</b> Galaxy	Phenotype Association		😤 Workflow Visualize Données partagées • Admin Aide • Utilisateur • 🖻	¢ 🎟	Using 12%
From the 4th to 7th of	GENOMICS TOOLKITS				
Tools	Picard				History 🕄 🕂 🗖 🌣
search tools	deepTools		フ		Rechercher des données 🛛 🕄
± Up	Gemini		+1600 tools		Unnamed history
	EMBOSS				(empty) 📎 🗩
Transcriptomics	Searching tools			7	
COMPLITATIONAL	RAD-seq	e to	General Text Tools		Cet historique est vide. You can
Probes and prin	Trinity	xy insta			Charger vos propres données or Charger des données depuis
"Nucleic acid sit	METAGENOMICS	or supp	Genomic File Manipulation		une source externe
motifs"	Metagenomic Analysis	•	Common Genomics Tools		
"Sequence com	DADA2	usegala	Genomics Analysis	δ.	
Sequence asse	FROGS		Certornics Analysis		
Mapping	Qiime	ecif	Genomics Toolkits		
Computational I	Obitools		Metagenomics		
COMPUTER SCIER	PICRUSt	at with t			
Data visualisati	Mothur	Work	WORKHOW4Metabolomics	Metabarcoding	
LABORATORY TE	Virology	Metabo	Metabolic Networks	FROGS, Qiime, Mothur, Obitools,	
MATHEMATICS	WORKFLOW4METABOLOMICS	analysi Metabo	Proteomics	DADA2, PICRUSt	
Statistics and p	Get Data - Metabolomics	Droto		Ack the Colour Cat	
OMICS	Preprocessing LCMS	Functio	Galaxy-E	Can't find a tool here or elsewhere?	
Genomics	Prenrocessing FIAMS	of prote	Statistics And Visualisation	Why not search on the IFB Galaxy	-
Comparative ge	nomics	data in	Missellanoous Taola	Catalog ?	
Uncategorized		•	WISCEIIdHEOUS 1001S		
All workflows	Ho	osted by:			
<		<b>K</b> IFB	elijir	$(\mathbf{O})$	

### Galaxy France: <a href="https://sub-domain>.usegalaxy.fr">https://sub-domain>.usegalaxy.fr</a>

#### 7 sub-domains:

Station Biologique

Roscoff



metabarcoding

proteore



#### covid19

workflow4metabolomics



Wm

#### <u>mnhn</u>

<del>الْمَر</del> <u>met4</u>j

#### Focus on your analysis

- Personalized welcome page
- Filtered tool list

<b>= Galaxy</b> France	👚 Workflow Visualize Données pa	rtagées 🔹 Aide 👻 Utilisateur 👻 🔁	Using 45%
Tools 🗠 🔹			
search tools × ×	Communities: dom	aın-centric Galaxy si	ubdomains
📩 Upload Data			
Get Data	Workflow4Metab	ProteoRE	Covid19
Send Data	olomics	Functional analysis and	Variant analysis
Collection Operations		exploration of	consensus using
GENERAL TEXT TOOLS	Data processing, analysis and annotation	transcriptomics data in	workflows and datasets
Text Manipulation	for the metabolomics community	ProteoRE	Covid19
Filter and Sort			
Join, Subtract and Group			
GENOMIC FILE MANIPULATION	VV 111		
Convert Formats	Workflow/instate/omics		
FASTA/FASTQ	Workflow4Metabolomics	ProteoRE	
FASTQ Quality Control			
SAM/BAM			
BED			
VCF/BCF			
Nanopore	Metabarcoding	Met4J	MNHN
COMMON GENOMICS TOOLS	With the following	Open-source Java	Tools from the French
	tools and pipelines : EPOCS Olimo Mothur	library dedicated to the	Muséum National





#### 130+ platforms for using Galaxy: <u>https://galaxyproject.org/use</u> [Servers, clouds, deployable resources]

#### **GENERAL PURPOSE:**

<u>Usegalaxy.fr</u>, usegalaxy.org, usegalaxy.eu, usegalaxy.org.au

#### **DOMAIN SPECIFIC:**

. . .







#### 130+ platforms for using Galaxy: <u>https://galaxyproject.org/use</u> [Servers, clouds, deployable resources]

#### **GENERAL PURPOSE:**

<u>Usegalaxy.fr</u>, usegalaxy.org, usegalaxy.eu, usegalaxy.org.au

#### **DOMAIN SPECIFIC:**









Catalog of French Galaxy tools: <u>http://galaxycat.france-bioinformatique.fr</u>



- On your own:
  - Training materials:



https://galaxyproject.github.io/training-material

– Interactive tours of Galaxy:

https://usegalaxy.fr/tours

• Training courses:

	Training	What ?	Where ?	When ?
Bims	<u>RNAseq analysis</u> <u>with Galaxy</u>	RNAseq	Roscoff, France	?
Galaxy	Galaxy Community Conference (GCC)	General purpose (data-intensive biology and Galaxy)	Brno, Czech Republic	24-29 June 2024
	Workflow4Experimenters	Metabolomics	Archamps, France	April 2024



# Please complete the evaluation questionnaire! **END**



### Training evaluation

# Thank you for completing the training evaluation questionnaire:

http://abims.sb-roscoff.fr/evaluation_formation