











# **Galaxy**

Initiation

Loraine Guéguen

Erwan Corre

Credits to Gildas Le Corguillé, Galaxy Training Network

v3.0





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





- Local computer:
  - login: guest
  - pwd: guest29
- Slides:
  - google slides: <u>https://tinyurl.com/glx-init-slides</u>
  - pdf: <u>https://tinyurl.com/glx-init-pdf-slides</u>
- Datasets: <u>https://tinyurl.com/glx-init-datasets</u>
- Galaxy France <u>https://usegalaxy.fr/</u>
  - registration ok ?



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

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



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



## 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 Introduction / problematic

- Graphical interface click-button tools within windows
  - + very ergonomic
  - too ergonomic  $\rightarrow$  lack of flexibility
  - few

Roscoff

- paying for it!
- Tools available on the **internet** 
  - + very ergonomic
  - too ergonomic  $\rightarrow$  lack of flexibility Come
  - A small part of the available tools
  - the submission size /storage is often limited
  - must not be paranoid





## Introduction / problematic

#### library(xcms)

#### loaddata() Command line tools polar

#### noise=250000

xset <- xcmsSet(cdffiles,ppm=ppm, mzdiff=mzwid, peakwidth=peakwidth,\_noise=noise,\_snthresh=snth, method="centWave", fitgauss=TRUE, nSlaves=8)</pre> xset2<-retco+ represent almost the majority of scientific tools dev.copv2pdf(devi "), paper="a4", height=9, width=14) xset3<-group(xset2, minfrac = 0.2, bw = bw, minsamp = 1, mzwid = mzwid, max = 50, sleep = 0)</pre> xset5<-fillPeak

### + good parameters completeness

reporttab <- diffreport(xset5, filebase =paste(pathResult,"/Rapport\_",expe,"\_",polar, sep=""), mzdec=4, eicmax=5000, metlin = metlin, classeic=levels(xset5@phenoDa

### + can be executed on high performance computers

dir.create(paste(pathResult,"/Rapport\_", expe,"\_", polar,"\_diffreport/", sep=""), showWarnings = FALSE) write.table(reporttab,paste(patpResult,"/Rapport\_",expe,"\_",polar,"\_diffreport/resultat\_",expe,"\_",polar,".xls", sep=""),sep="\t")
+ automatable, workflowsable, ...

library(CAMERA)

### and anotate (x minimum linux knowledge is required and maxiso and minimum linux knowledge is required and maxiso and minimum linux knowledge is required and maxiso and maxiso and minimum linux knowledge is required and minimum linux knowledge is require

## diffreport1<-getPeaklist(an) - cruel lack of ergonomics

#diffreport <- annotateDiffreport(xsg,pval\_th=0.05,fc=0.1, nSlaves=8, calcIso=TRUE, calcCaS=FALSE, maxcharge=3, maxiso=4, minfrac=0.5, # polarity=polarity, sortpval=FALSE) diffreport<-cbind(reporttab,diffreport1[,c("isotopes", "adduct","pcgroup")])</pre> write.table(diffreport, file=paste(pathResult,"/result\_",expe,"\_",polar,"\_CAMERA\_diffreport-fast.xls",sep=""), row.names=FALSE, sep="\t")

```
library(FactoMineR)
pca3<-PCA(t(matacp), axes=c(1,2))</pre>
pca3<-PCA(t(matacp), axes=c(1,3))</pre>
pca3<-PCA(t(matacp), axes=c(2,3))</pre>
pca4<-PCA(t(matacplog2))</pre>
```

```
png("percentage_of_variance.png", width =800, height = 400);
barplot(resPCA$eig$per,xlab="Components",ylab="percentage of variance");
dev.off()
```

```
png("eigenvalue.png", width =800, height = 400);
barplot(resPCA$eig$eig,xlab="Components",ylab="eigenvalue");
dev.off()
```

```
library(ctc)
if (normalization) {
    data=t(scale(t(data)))
```

3

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err
tophat2 panda v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq
-GTF ../input/panda v121029.gtf --b2-sensitive -r 100
-num-threads 8
[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
[login@n0 mapping]$ ls
accepted hits.bam junctions.bed qsub.err unmapped.bam
deletions.bed logs
                          qsub.out
insertions.bed prep reads.info tmp
[login@n0 mapping]$ cd
```

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err
                                                 t/IllR1-2.fq
tophat2 panda v1
-GTF ../input/
-num-threads
```

[login@n0 mapping]\$ qsub -q long.q -pe thread 8 tophat.qsub Your job 5338969 ("tophat.qsub") has been submitted [login@n0 mapping]\$ ls accepted\_hits.bam junctions.bed qsub.err unmapped.bam deletions.bed logs qsub.out insertions.bed prep\_reads.info tmp [login@n0 mapping]\$ cd

## Introduction / problematic

tation Biologique





## Introduction / problematic







## « I want to know the gene expression »





« I want to map my reads on a reference genome and count them »





« I want to launch the tools tophat2 and cufflinks.

I have fastq files and my genome in fasta and gtf. »





 « I want 1TB for my project. I will launch tophat2 through SSH on the cluster in multi-thread mode.
 Next I want to submit the bam file to my genome with cufflinks."





« I have a bunch of cool tools! But I'm the only one who can launch them.

Comments? »



# **INTRODUCTION / GALAXY**





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



# ABMS Introduction / Galaxy

## **MR.GEEK**



[login@n0 ~]\$ cdprojet
[iogineno iogin]\$ cd is-0/-29-panda/ump/mapping
[login@n0 mapping]\$ cat tophat.qsub
#!/bin/bash
#\$ -S /bin/bash
#\$ -M login@sb-roscoff.fr
#\$ -m bea
#\$ -V
#\$ -cwd
#\$ -o qsub.out
#\$ -e qsub.err
tophat2 panda_v121029/input/IIIR1-1.fq/input/III
-GTF/input/panda v121029.gtfb2-sensitive -r 100

-num-threads 8 [login@n0 mapping]\$ qsub -q long.q -pe thread 8 tophat.qsub Your job 5338969 ("tophat.qsub") has been submitted

R1-2.fq





## Introduction / Galaxy

- ogın@n0 ~]Ş cdpro
- [Ioginenu iogin]\$ cd i3-0/-29-panda/ump/ma
- #!/bin/bash
- \$ -S /bin/b
- \$ -M login@sb-roscoff.fr
- \$ -m be
- #\$ -V
- \$ -cwd
- \$ -o qsub.c
- S -e qsub.er

cophat2 panda\_v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq -GTF ../input/panda\_v121029.gtf --b2-sensitive -r 100 -num-threads 8

[login@n0 mapping]\$ qsub -q long.q -pe thread 8 tophat.qsul Your job 5338969 ("tophat.qsub") has been submitted











## Introduction / Galaxy

[lecorguille@n0 ~]\$ e-PCR --help e-PCR: invalid option -- usage: [-hV] [posix-options] stsfile [fasta ...] [compat-options] where posix-options are: -m ## Margin (default 50) -w ## Wordsize (default 7) Max mismatches allowed (default 0) -n ## Max indels allowed (default 0) -g ## Use ## discontiguos words, slow if -f ## ##>1 -0 ## Set output file -t ## Set output format: 1 - classic, range (pos1..pos2) 2 - classic, midpoint 3 - tabular 4 - tabular with alignment in comments (slow) Set default size range -d##-## (default 100-350) Turn hits postprocess on/off -p +--v ## Verbosity flags Use presize alignmens (only if -a alf gaps>0), slow a - Allways or f - as Fallback Use 5'-end lowercase masking of -x +primers (default -) Uppercase all primers (default -) -u +-

#### Galaxy / ABiMS

#### e-PCR (version 1.0.0)

#### STS file:

100: (as tabular) Trinity on data 9..Transcripts

#### Fasta file:

100: Trinity on data 9.. Transcripts

#### format : fasta

Wordsize (W):

```
7
```

Set word size for primers hash (nucleotide positions). Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near 'inner' boundary of primers unless one uses discontiguous words, and no gaps are ever allowed in that region.

#### Use ## discontinuos words (F):

#### 1

Set discontiguous word count for primers hash (1 means 'use contiguous words'). Discontiguous words increase number of hash tables and decrease 'effective' word size (thus increasing hash collision rate), so make search significantly slower, but increase sencitivity by allowing mismatches within word size. Reasonable values are 1 (contiguous words) and 3.

#### Margin (M):

50

Set maximal allowed deviation of hit product size from expected STS size.

#### Set default sts lower size (D):

100

Set ddefault STS size range - values used for STSs that have no size associated in file.

#### Set default sts higher size (D):

400

Set ddefault STS size range - values used for STSs that have no size associated in file.

#### Max mismatches allowed (N):

0

Set maximal number of mismatches allowed in primer-to-sequence alignment (per primer!).

#### Max indels allowed (G):

Set maximal number of gaps allowed in primer-to-sequence alignment (per primer!).

#### Set output format (T):

tabular

Execute

0

Output formats

[...]

# usegalaxy.fr GALAXY FRANCE





### Station Biologique Roscoff Galaxy France ARMS

## FB 6 INSTITUT FRANÇAIS DE BIOINFORMATIQUE

### https://usegalaxy.fr

1872

📮 Galaxy France	🏠 Workflow Visualize 🕶 Shared Data 🕶 Help 👻 User 🕶 📰		Using 3%
Tools		History	2+0 <b>¢</b>
search tools		search datasets	00
1 Upload Data		<b>blast hands-on 202</b> 8 shown, 10 hidden	22
Get Data		66.36 MB	2 🌑 🗩
Send Data	Welcome to usedalaxy fr		_
Collection Operations	Welconne to useguluxy.n	13: NCBI BLAST+ blas	tn across c X
GENERAL TEXT TOOLS	By using this Galaxy instance, we assume that you have read and accept the Term Of Use	a list with 5 items	
Text Manipulation	For any questions or support: community.cluster.france-bioinformatique.fr/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 v	rsʻpr ⊛∦X
GENOMIC FILE MANIPULATION	the 21.05 user release notes.	otein BLAST database m data 2'	fro
Convert Formats	·	10: magablast support	5 0 A Y
FASTA/FASTQ		vs 'nucleotide BLAST	data
FASTQ Quality Control		base from data 1'	
SAM/BAM	Ack the	9: nucleotide BLAST d	lata 💿 🥒 🗙
BED	ASK LITE	base from data 1	
VCF/BCF	GalavyCat	8: protein BLAST data e from data 2	ibas 💿 🖋 🗙
Nanopore	GalaxyCat	2: Drosophila melanc	gast 💿 🖋 🗙
COMMON GENOMICS TOOLS	Can't find a tool on	er.BDGP6.22.pep.all.f	a.gz
Operate on Genomic Intervals	usegalaxy.fr?	1: Drosophila_melano	gast <sup>26</sup> ∉ ×

## Galaxy ABiMS --> Galaxy France

### https://galaxy.sb-roscoff.fr/

Station Biologique Roscoff

ARMS



🗧 Galaxy / ABiMs	5		Analyze Data Workflow Visualize Shared Data Admin Help User -		Using 34%
Tools search tools	☆ ± 8	Velcome	to galaxy.sb-roscoff.fr	History     search datasets	2+⊞≎
Get Data Send Data Collection Operations COMMON TOOLS	~	14-10-2020 Dear users, we plan So, if you start a ne This instance is mor If tools are missing,	to eventually close this https://galaxy.sb-roscoff.fr instance in favour of https://usegalaxy.fr w analysis, please consider migrating to the new national instance https://usegalaxy.fr, re modern, supported by the IFB Core Cluster, maintained in part by ABiMS and the IFB NNCR Cluster TaskForce. do not hesitate to request them on https://community.france-bioinformatique.fr/.	Galaxy initiation traini tools hands-on 11 shown, 4 deleted, 7 hidden 11.73 MB	ing 2019 - 🕑 📡 🗩



## https://usegalaxy.fr/





### Station Biologique Roscoff Galaxy France

### Get help

🗧 Galaxy France	👚 Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🖛 📻 🏢		Using 3%
Tools		History	2+0 <b>¢</b>
search tools		search datasets	88
🏂 Upload Data		<b>blast hands-on 2022</b> 8 shown, 10 hidden	
Get Data		66.36 MB	
Send Data	Welcome to usegalaxy fr		_
Collection Operations	Weicome to abegulaxy.m	13: NCBI BLAST+ blastr ollection 12	across c X
GENERAL TEXT TOOLS	By using this Galaxy instance, we assume that you have read and accept the Term Of Upe	a list with 5 items	
Text Manipulation	For any questions or support: community.cluster.france-bioinformatique.fr/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	'pr 🕑 🖋 🗙
GENOMIC FILE MANIPULATION	the 21.05 user release notes.	otein BLAST database f m data 2'	ro
Convert Formats	×	10. manablast manuf. 6	
FASTA/FASTQ		vs 'nucleotide BLAST da	a 🕑 🌮 🗡
FASTQ Quality Control		base from data 1'	
SAM/BAM	Ack the	9: nucleotide BLAST dat	ta 🕑 🖋 🗙
BED	ASK LITE	base from data 1	
VCF/BCF	GalavyCat	8: protein BLAST datab e from data 2	as 💿 🖋 🗙
Nanopore	GalaxyCat	2: Drosophila_melanog	ast 💿 🥒 🗙
COMMON GENOMICS TOOLS	Can't find a tool on	er.BDGP6.22.pep.all.fa.	gz
Operate on Genomic Intervals	usegalaxy.fr?	1: Drosophila_melanog	



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

Community Support	S'inscrire	Se connecter	ର
► Welcome message			×
Galaxy ▶ toutes les étiquettes ▶ tout ◄ Récents Top			
Sujet	Répo	nses 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	٩ 1	88	sept. '21
<ul> <li>README: before request a Galaxy quota extension</li> <li>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</li> </ul>	2	2 180	janv. '21
A propos de la catégorie Galaxy For any questions, issues and request regarding usegalaxy.fr and its subdomains *.usegalaxy.fr	2	469	oct. '20
☑ FROGS - job was terminated because it used more memory than it was allocated	2	1 54	11 h
FROGS - bug à l'affichage des sorties html		17	<sup>14 h</sup> 29



# DATA IMPORT





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



# Step 1: Choose a FTP Client DATA IMPORT USING FTP



# STEP 1: CHOOSE A FTP CLIENT





# STEP 1: CHOOSE A FTP CLIENT



# Step 2: Easy! DATA IMPORT USING FTP




# Data import using FTP

# Example with WinSCP for usegalaxy.fr





# **DATA IMPORT**



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	ි History සි 🕇 🗖 🌣
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 >



# Hands-on DATA IMPORT







1. Fetch the files with your internet browser: <u>https://tinyurl.com/glx-init-datasets</u>

2. Upload files into Galaxy (exons.bed, snps.bed, repeats.bed)

- a. From disk
- b. Using FTP



# TOOLS

# Station Biologique Roscoff ARSS TOOLS - panel

💶 Galaxy France	👚 Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🖛 💼			Using 3%
Tools     ☆       search tools     ⊗       Lupload Data		~	History search datasets blast hands-on 2022	2+0¢ 00
Get Data			66.36 MB	
Collection Operations GENERAL TEXT TOOLS	Welcome to usegalaxy.tr By using this Galaxy instance, we assume that you have read and accept the Term Of Use		13: NCBI BLAST+ blastn ollection 12 a list with 5 items	across c X
Text Manipulation Filter and Sort	For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy		<b>12: queries</b> a list with 5 items	×
Join, Subtract and Group GENOMIC FILE MANIPULATION	<ul> <li>22/07/2021: usegalaxy.fr is now running the release 21.05 of Galaxy. Please check the 21.05 user release notes.</li> </ul>		11: blastx query5.fa vs ' otein BLAST database fr m data 2'	pr 💿 🥒 🗙
Convert Formats FASTA/FASTQ FASTO Quality Control			10: megablast query5.fa vs 'nucleotide BLAST da base from data 1'	ta 🕑 🖉 🗙
SAM/BAM BED	Ask the		9: nucleotide BLAST dat base from data 1	a 💿 🖉 🗙
VCF/BCF	GalaxvCat		8: protein BLAST databa e from data 2	15 🖲 🥒 X
COMMON GENOMICS TOOLS	Can't find a tool on		2: Drosophila_melanoga er.BDGP6.22.pep.all.fa.g	ist 💿 🥒 X jz
Operate on Genomic Intervals Fetch Alignments/Sequences	usegalaxy.fr? Why not search on the IFB		1: Drosophila_melanoga er.BDGP6.22.cds.all.fa.g	ast ④ A X Z
GENOMICS ANALYSIS	Galaxy Catalog ?			

### Station Biologique Roscoff TOOIS - panel

📱 Galaxy France	Workflow Visualize - Shared Data - Help - User - 🖻 🏢		Using 3%
Tools ☆		History	<b>∂+</b> ∎¢
search tools		search datasets	00
🎝 Upload Data		blast hands-on 2022	
^		8 shown, 10 hidden	
Get Data		66.36 MB	
Send Data	Welcome to usegalaxy fr		_
Collection Operations	Welconne to abegulaxy.m	13: NCBI BLAST+ blastn ollection 12	across c X
GENERAL TEXT TOOLS	By using this Galaxy instance, we assume that you have read and accept the Term Of Use	a list with 5 items	
Text Manipulation	For any questions or support: community.cluster.france-bioinformatique.fr/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 '	pr 👁 🥒 🗙
GENOMIC FILE MANIPULATION	the 21.05 user release notes.	otein BLAST database fr m data 2'	0
Convert Formats			
FASTA/FASTQ	M/hat tools are availa	hlo?	ita
FASTQ Quality Control			
SAM/BAM	Ack the	9: nucleotide BLAST dat	a 🕑 🖉 🗙
BED	ASK LITE	base from data 1	
VCF/BCF	Galavy/Cat	8: protein BLAST databa e from data 2	is 💿 🧬 🗙
Nanopore	GalaxyCat	2: Drosophila melanoga	act OAX
COMMON GENOMICS TOOLS	Can't find a tool on	er.BDGP6.22.pep.all.fa.g	jz
Operate on Genomic Intervals	usegalaxy.fr?	1: Drosophila_melanoga	ast 💿 🥒 🗙
Fetch Alignments/Sequences	Why not search on the IFB	er.BDGP6.22.cds.all.fa.g	z
GENOMICS ANALYSIS	Galaxy Catalog ?		
<			2



## 130+ platforms for using Galaxy: <a href="https://galaxyproject.org/use">https://galaxyproject.org/use</a>

[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

## **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:**





#### Free "app" store: Galaxy Tool Shed

- Thousands of tools already available
- Most software can be integrated
  - If a tool is not available, ask the Galaxy community for help!
- Only a Galaxy admin can install tools

32 valid tools on Dec 04, 2018	Repositorie	es by Category	
earch Search for valid tools	search repository na	me, description Q	
Search for workflows	Name	Description	Repositories
lid Galaxy Utilities	Assembly	Tools for working with assemblies	128
<u>Custom datatypes</u>	ChIP-seq	Tools for analyzing and manipulating ChIP- seq data.	65
Tool dependency definitions	Combinatorial Selections	Tools for combinatorial selection	10
Repositories Browse by category	Computational chemistry	Tools for use in computational chemistry	76
ailable Actions Login to create a repository	Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	12
	Convert Formats	Tools for converting data formats	114
		Tools for exporting data to various	1.2

### Station Biologique Roscoff ARMS TOOLS - panel

🚍 Galaxy France	倄 Workflow Visuali	ze 🕶 Shared Data 🕶 Help 👻 User 👻 📰			Using 3%
Tools ☆			^	History	<b>₽+</b> ∎¢
trinity				search datasets	00
🏂 Upload Data				Trinity example	
				2 shown	
@ Show Sections				2.5 MID	
Generate gene to transcript map for Trinity assembly	Welcome to us	segalaxy.fr		2: reads-rv.fastq	⊙ / ×
Generate SuperTranscripts from a Trinity assembly	By using this Galaxy instance, we	assume that you have read and accept the T	erm Of Use	1: reads-fw.fastq	⊛ ∲ ×
Filter low expression transcripts from a Trinity assembly	For any questions or support: co	mmunity.cluster.france-bioinformatique.fr/c/	galaxy		
Extract and cluster differentially expressed transcripts from a Trinity assembly	• 22/07/2021: usegalaxy.fr is the 21.05 user release note	now running the <b>release 21.05</b> of Galaxy. Ple s.	ease check		
Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly			~		
Differential expression analysis using a Trinity assembly	Pirty				
<b>Trinity</b> de novo assembly of RNA-Seq data	P 13	Ask the			
Partition genes into expression clusters after differential expression analysis using a Trinity assembly		GalaxyCat			
<b>Build expression matrix</b> for a de novo assembly of RNA-Seq data by Trinity		Can't find a tool on usegalaxy.fr?			
RNASeq samples quality check for transcript quantification		Why not search on the IFB Galaxy Catalog ?			

# Tools - form

#### **Galaxy France**

Station Biologique

Roscoff

Tools

trinity

assembly

data

🐣 Workflow Visualize 🕶 Shared Data 🕶 Help 👻 User 🕶 💼 History 2 + m 🌣 Trinity de novo assembly of RNA-Seq data (Galaxy ☆ Favorite & Versions Options 8 00 Version 2.9.1+galaxy2) Are you pooling sequence datasets? **Trinity example** 1. Upload Data Yes • Show Sections Paired or Single-end data? Generate gene to transcript map for Single-end -O / X 2: reads-rv.fastq Trinity assembly Single-end reads Generate SuperTranscripts from a O / X 1: reads-fw.fastq Trinity assembly C 2: reads-rv.fastg £ B 1: reads-fw.fastq Filter low expression transcripts from a Trinity assembly Extract and cluster differentially expressed transcripts from a Trinity (--single) Compute contig Ex90N50 statistic Strand specific data and Ex90 transcript count from a Trinity assembly No **Differential expression analysis** Run in silico normalization of reads using a Trinity assembly Yes Trinity de novo assembly of RNA-Seq Defaults to max. read coverage of 50. (--no\_normalize\_reads) Partition genes into expression Additional Options 20 clusters after differential expression Email notification analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

Trinity assembles transcript sequences from Illumina RNA-Seq data.

Send an email notification when the job completes.

No

Execute

# Tools - form

#### **Galaxy France**

Station Biologique

Roscoff

data

Trinity

RNASeq samples quality check for

transcript quantification

🐣 Workflow Visualize 🕶 Shared Data 🕶 Help 👻 User 🕶 💼 2 + m 🌣 Tools History Trinity de novo assembly of RNA-Seq data (Galaxy ☆ Favorite & Versions Options 8 00 trinity Version 2.9.1+galaxy2) Are you pooling sequence datasets? **Trinity example** 1. Upload Data Yes • Show Sections Paired or Single-end data? Generate gene to transcript map for Single-end -O / X 2: reads-rv.fastq Trinity assembly Single-end reads Generate SuperTranscripts from a O & X 1: reads-fw.fastg Trinity assembly C 2: reads-rv.fastg £ B 1: reads-fw.fastq Filter low expression transcripts from a Trinity assembly Extract and cluster differentially expressed transcripts from a Trinity assembly (--single) Compute contig Ex90N50 statistic Strand specific data and Ex90 transcript count from a Trinity assembly No **Differential expression analysis** Run in silico normalization of reads using a Trinity assembly Yes Trinity de novo assembly of RNA-Seq Defaults to max. read coverage of 50. (--no\_normalize\_reads) Partition genes into expression Additional Options R clusters after differential expression Email notification analysis using a Trinity assembly No Build expression matrix for a de novo assembly of RNA-Seq data by Send an email notification when the job completes.



Trinity assembles transcript sequences from Illumina RNA-Seq data.

## Station Biologique Roscoff Tools - form

#### **Galaxy France**

#### Tools 8 trinity

1 Upload Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

Extract and cluster differentially expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

**Differential expression analysis** using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

<

. .

efaults to m	nax. read coverage of 50. (no_normalize_reads)
Additional (	<u>Options</u>
Minimum	Contig Length
200	
All contigs <b>Use the ge</b>	shorter than this will be discarded (min_contig_length) nome guided mode?
No	
lf you alrea	dy mapped the reads to the genome, Trinity can use this information
Error-corr	ected or circular consensus (CCS) pac bio reads
Error-corr	ected or circular consensus (CCS) pac bio reads           Image: Description of the set of the se
Error-corr Experiment (long_rea Minimum	ected or circular consensus (CCS) pac bio reads          Image: Description of the state and the state orientation as short reads if they are strand specific ds)         count for K-mers to be assembled
Error-corr Experiment (long_rea Minimum 1 (min_kme	ected or circular consensus (CCS) pac bio reads          Image: No fasta dataset available.       Image:
Error-corr Experiment (long_rea Minimum 1 (min_kme mail notific No end an ema	ected or circular consensus (CCS) pac bio reads   Image: No fasta dataset available.     al feature! Long reads must be in the same orientation as short reads if they are strand specific ds)   count for K-mers to be assembled   er_cov)   In notification when the job completes.

History	2 <b>+ 🗆 </b>
search datasets	00
Trinity example	
2 shown	
2.5 MB	Solution
2: reads-rv.fastq	● / ×
1: reads-fw.fastq	③

# Tools - form

#### **Galaxy France**

R

Station Biologique Roscoff

ols	t
trinity	8
📩 Upload I	Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

**Extract and cluster differentially** expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

**Differential expression analysis** using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

<

....

Use the genome guided mode?	^ History	S+0:
No	Caparah da	
If you already mapped the reads to the genome, Trinity can use this information	search da	
Error-corrected or circular consensus (CCS) pac bio reads	Trinity ex	ample
🗅 🗘 🗅 No fasta dataset available. 🔹 🗈	2 shown	
	2.5 MB	
Experimental feature! Long reads must be in the same orientation as short reads if they are strand specific		
ong_reads) Minimum count for K-mers to be assembled	2: reads-rv	.fastq 💿 🖉
	1: reads-fw	v.fastq 💿 🥒
nail notification No nd an email notification when the job completes.		
nail notification No Ind an email notification when the job completes. Execute inity assembles transcript sequences from Illumina RNA-Seq data.		
nail notification No Ind an email notification when the job completes. Execute Inity assembles transcript sequences from Illumina RNA-Seq data. Citations: []		
nail notification No Ind an email notification when the job completes. Execute Inity assembles transcript sequences from Illumina RNA-Seq data. Citations:  Citations:  Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011).		
<ul> <li>nail notification</li> <li>No</li> <li>nd an email notification when the job completes.</li> <li>Execute</li> <li>Inity assembles transcript sequences from Illumina RNA-Seq data.</li> <li>Citations:  <ul> <li>Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011).</li> <li>Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature</li> <li>Rictarban logy, 20(7), 644, 652, https://doi.org/10.1028/pht1822</li> </ul></li></ul>		
<ul> <li>nail notification         <ul> <li>No</li> <li>Ind an email notification when the job completes.</li> </ul> </li> <li>Execute         <ul> <li>Execute</li> </ul> </li> <li>Inity assembles transcript sequences from Illumina RNA-Seq data.</li> <li>Citations:              <ul> <li>Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011).</li> <li>Full-length transcriptome assembly from RNA-Seq data without a reference genome. <i>Nature Biotechnology, 29</i>(7), 644–652. https://doi.org/10.1038/nbt.1883                  </li> </ul> </li> </ul>		
<ul> <li>nail notification</li> <li>No</li> <li>nd an email notification when the job completes.</li> <li>Execute</li> <li>Inity assembles transcript sequences from Illumina RNA-Seq data.</li> <li>Citations: </li> <li>Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011).</li> <li>Full-length transcriptome assembly from RNA-Seq data without a reference genome. <i>Nature Biotechnology</i>, 29(7), 644–652. https://doi.org/10.1038/nbt.1883 C</li> <li>Requirements: ?</li> <li>trinity (Version 2.9.1)</li> </ul>		
<ul> <li>nail notification</li> <li>No</li> <li>nd an email notification when the job completes.</li> <li>Execute</li> <li>Inity assembles transcript sequences from Illumina RNA-Seq data.</li> <li>Citations:  <ul> <li>Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011).</li> <li>Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644–652. https://doi.org/10.1038/nbt.1883 C</li> <li>Requirements: ?</li> <li>trinity (Version 2.9.1)</li> <li>coreutils (Version 8.32)</li> </ul> </li> </ul>		
<ul> <li>No</li> <li>No</li> <li>Ind an email notification when the job completes.</li> <li>Execute</li> <li>Inity assembles transcript sequences from Illumina RNA-Seq data.</li> <li>Citations:  <ul> <li>Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011).</li> <li>Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644–652. https://doi.org/10.1038/nbt.1883 </li> <li>Requirements: ?</li> <li>trinity (Version 2.9.1)</li> <li>coreutils (Version 8.32)</li> <li>rsync (Version 3.2.3)</li> </ul> </li> </ul>		
<ul> <li>nail notification</li> <li>No</li> <li>nd an email notification when the job completes.</li> <li>✓ Execute</li> <li>Inity assembles transcript sequences from Illumina RNA-Seq data.</li> <li>Citations: ■</li> <li>Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. <i>Nature Biotechnology</i>, <i>29</i>(7), 644–652. https://doi.org/10.1038/nbt.1883 C</li> <li>Requirements: ?</li> <li>trinity (Version 2.9.1)</li> <li>coreutils (Version 8.32)</li> <li>rsync (Version 3.2.3)</li> <li>References:</li> </ul>		

# Station Biologique Roscoff ARMS TOOLS - form

<b>=</b> Galaxy France	😤 Workflow Visualize 🕶 Shared D	ata 🕶 Help 👻 User 🖛 💼		Using 3%
Tools	<b>Trinity</b> de novo assembly of RNA-Seq data (Galaxy Version 2.9.1+galaxy2)	☆ Favorite	History     search datasets	ສ+⊡‡ ຄຄ
1 Upload Data	Are you pooling sequence datasets? Yes	<ul> <li>Switch to 2.9.1+galaxy1</li> <li>Switch to 2.9.1</li> <li>Switch to 2.8.5</li> </ul>	Trinity example	
Show Sections	Paired or Single-end data?	Switch to 2.0.5	2.5 MB	
Generate gene to transcript map for Trinity assembly	Single-end		2: reads-rv.fastq	● / ×
Generate SuperTranscripts from a Trinity assembly	Single-end reads	^ 1	1: reads-fw.fastq	● / ×
Filter low expression transcripts from a Trinity assembly	1: reads-fw.fastq			
Extract and cluster differentially expressed transcripts from a Trinity assembly		-b		
Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly	(single) Strand specific data No			
Differential expression analysis using a Trinity assembly Trinity de novo assembly of RNA-Seg	Run in silico normalization of reads			
data	Defaults to max. read coverage of 50. (no_normalize_	eads)		
Partition genes into expression	Additional Options	6	>	
analysis using a Trinity assembly	Minimum Contig Length			
Build expression matrix for a de	200			
novo assembly of RNA-Seq data by Trinity	All contigs shorter than this will be discarded (min_c	ontig_length)		
RNASeq samples quality check for	Use the genome guided mode?			
transcript quantification	No		•	
<pre></pre>	If vou alreadv mapped the reads to the genome. Trini	tv can use this information		>

## Station Biologique Roscoff Tools - form

#### Cal

assembly

📮 Galaxy France	🕋 Workflow Visualize - Shared Data - Help - User - 扄 🏬			Using 3%
Tools ☆	Run in silico normalization of reads	^	History	₽+□¢
trinity	Ves Defaults to max. read coverage of 50. (no_normalize_reads)		search datasets	00
🏦 Upload Data	Additional Options (*		Trinity example	
Show Sections	Minimum Contig Length	) <sup>2</sup>	2.5 MB	v 🔊 🗩
Generate gene to transcript map for Trinity assembly	All contigs shorter than this will be discarded (min_contig_length)	1	2: reads-rv.fastq	⊙ / ×
Generate SuperTranscripts from a Trinity assembly	Vse the genome guided mode?		1: reads-fw.fastq	● # ×
Filter low expression transcripts from a Trinity assembly	If you already mapped the reads to the genome, Trinity can use this information	8. I.I.		
Extract and cluster differentially expressed transcripts from a Trinity	Error-corrected or circular consensus (CCS) pac bio reads			

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

**Differential expression analysis** using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification



Ci	ta	ti	0	n	s

<

# Station Biologique Roscoff TOOLS - form

Galaxy France	প Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 📰		Using 3%
Tools ☆		History	2+□\$
trinity	Executed Trinity and successfully added 1 Job to the queue.	search datasets	00
Upload Data Show Sections	<ul> <li>• 1: reads-fw.fastq</li> <li>• 2: reads-rv.fastq</li> </ul>	Trinity example 4 shown 2.5 MB	v 🗞 🗩
Generate gene to transcript map for Trinity assembly	<ul> <li>It produces 2 outputs:</li> <li>3: Trinity on data 2 and data 1: Assembled Transcripts</li> <li>4: Trinity on data 2 and data 1: Gene to transcripts map</li> </ul>	• 4: Trinity on da	ta 2 an 💿 🖋 🗙
Generate SuperTranscripts from a Trinity assembly	You can check the status of queued jobs and view the resulting data by refreshing the History	pts map	ranscri
Filter low expression transcripts from a Trinity assembly	panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.	3: Trinity on da d data 1: Assemble	ta 2 an 💿 🖋 🗙 ed Tran
Extract and cluster differentially		scripts	
assembly	Tool recommendation	2: reads-rv.fastq	⊙ # ×
Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly	You have used trinity tool. For further analysis, you could try using the 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	1: reads-fw.fastq	<b>●</b> / ×
Differential expression analysis using a Trinity assembly	ones at the bottom. Please click on one of the following/recommended tools to open its definition.		
<b>Trinity</b> de novo assembly of RNA-Seq data	HISAT2		
Partition genes into expression clusters after differential expression analysis using a Trinity assembly	Map with BWA-MEM Salmon quant		
Build expression matrix for a de	Quast		

FASTA-to-Tabular

Busco

Concatenate datasets

Align reads and esti...

Trinity

novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

<

• \*

Y

10

111

V.





# Job is waiting to run

## = the job is in the scheduler « queue »

Duration time of this status depends on the amount of actual queued jobs and on the requested number of processors





# Job is currently running

= the job is being executed on the computing cluster

Duration time of this status depends on the job's attributes and the computing resources allocated.

Some programs are executed with several processors (using 4, 8 or 16 Gb of RAM).

And others are mono-threaded 🙁





# Job has finished successfully

But warnings or errors can be hidden behind!





# Job has finished but with an error status

= the program sends an error

The error is often explained by the program but sometimes ... not.





# Job has finished but with an error status

= the program sends an error

Possible causes of error :

- Bad usage : input file, format or option
- Bad integration of the program into Galaxy (exceeded memory allocation...)
- Non anticipated crash of the program

<b>Galaxy France</b>	👫 Workflow Visualize - Shared Data - Help - User - 🖻 🏢		Using 3%
Tools ☆		Alistory	<b>₽+</b> □ <b>¢</b>
search tools		search datasets	00
🏝 Upload Data		Error example	
Get Data		4 shown, 1 hidden 772 b	
Send Data Collection Operations	Welcome to usegalaxy.fr	4: NCBI Accession Dov	wnlo 💿 🌶 🗙
GENERAL TEXT TOOLS	By using this Galaxy instance, we assume that you have read and accept the Term Of Use	ad on : Log	
Text Manipulation	For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy	3: NCBI Accession Dov : Downloaded Files	wnload on X
Filter and Sort		a list with 1 item	
Join, Subtract and Group	• 22/07/2021: usegalaxy.fr is now running the <b>release 21.05</b> of Galaxy. Please check	2: NCBI Accession	Dow 💿 🖋 🗙
GENOMIC FILE MANIPULATION	the 21.05 user release notes.	nload on : Log	
Convert Formats		21: NCBI Accession	Download ×
FASTA/FASTQ		a list	•
FASTQ Quality Control		L	
SAM/BAM	A clutho		
BED	ASK LITE		
VCF/BCF	Colorado		
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 ?		

# ABMS Tools - Handle errors

Station Biologique Roscoff

n Galaxy France	😤 Wa	orkflow Visualize - Shared Data - Help - User - 🕋 📰			Using 3%
Tools ☆	NCBI Accession	Download	^	History	2+D 4
search tools	Dataset Informatio	on		search datasets	00
🏝 Upload Data	Number Name	2 NCBI Accession Download on : Log		Error example	
Get Data	Created Filesize	Monday Jan 10th 4:13:29 2022 UTC 109 b		772 b	
Collection Operations	Format File contents	txt contents		4: NCBI Accession Downlo ad on : Log	• • <i>•</i> *
Text Manipulation	History API ID UUID	21c4852849066ede 6882ed13-dee0-4ccc-9b85-b645f5d911f7	_	3: NCBI Accession Downlo : Downloaded Files a list with 1 item	oad on 💙
	Tool Parameters	Value		2: NCBI Accession Do wnload on : Log	• / >
Convert Formats	Select source for IDs ID List	accession_list AAG19274.1		error An error occurred with this	dataset:
FASTA/FASTQ FASTQ Quality Control	File Format How to handle download	FASTA Abort with error on first failure		NCBI Entrez returned erro Failed to download file wit	r code 400, a h id AAG192: ג
SAM/BAM BED	Job Information			Image: Weight of the second	wnload X
VCF/BCF Nanopore	Galaxy Tool ID: Command Line	toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download/ncbi_acc_download /0.2.7+galaxy0 [〕 empty		on : Downloaded Files a list	
COMMON GENOMICS TOOLS Operate on Genomic Intervals	Tool Standard Output	NCBI Entrez returned error code 400, are ID(s) AAG19274.1 valid?	z		
Fetch Alignments/Sequences GENOMICS ANALYSIS	Tool Standard Error	Failed to download file with id AAG19274.1 from NCBI	2		

🗧 Galaxy France	Workflow Visualize - Shared Data - Help - User - 🖻 🎆		Using 3%
Tools කි	Downloading accession number: AAG19274.1 Failed to download file with id AAG19274.1 from NCBI	History	2+0¢
search tools	failed.	search datasets	00
🏝 Upload Data		Error example 4 shown, 1 hidden	
Get Data		772 b	2 🌑 🗩
Send Data			
Collection Operations		4: NCBI Accession Dov	wnlo 💿 🖋 🗙
GENERAL TEXT TOOLS		uu on . Log	
Text Manipulation		3: NCBI Accession Dov : Downloaded Files	vnload 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			

### Station Biologique Roscoff Tools - Handle errors

#### **Galaxy France**

#### Tools

1. Upload Data

57

- **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
- VCF/BCF
- Nanopore
- COMMON GENOMICS TOOLS
- Operate on Genomic Intervals
- Fetch Alignments/Sequences

NCBI Accession Download		History
Dataset Informati	on	search datasets
Number	2	Error evenuele
Vame	NCBI Accession Download on : Log	Error example
Created	Monday Jan 10th 4:13:29 2022 UTC	4 shown, 1 hidden
ilesize	<b>109</b> b	772 b
Dbkey	?	
ormat	txt	4: NCBI Accessio
ile contents	contents	ad on : Log
listory Content API ID	11e6e18a7189130f	
listory API ID	21c4852849066ede	3: NCBI Accession
JUID	6882ed13-dee0-4ccc-9b85-b645f5d911f7	: Downloaded Fil
		a list with 1 item
ool Parameters		2: NCBI Acce

Input Parameter	Value
Select source for IDs	accession_list
ID List	AAG19274.1
Molecule Type	nucleotide
File Format	FASTA
How to handle download failures	Abort with error on first failure

#### Job Information

Galaxy Tool ID: toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download/ncbi_acc_dov /0.2.7+galaxy0 🗘	
Command Line	empty
Tool Standard Output	NCBI Entrez returned error code 400, are ID(s) AAG19274.1 valid?
Tool Standard Error	Failed to download file with id AAG19274.1 from NCBI

#### nple lden ession Downlo ۲ AX ession Download on X ed Files item Accession Do • / × wnload on : Log error An error occurred with this dataset: NCBI Entrez returned error code 400, an Failed to download file with id AAG1927 < > \* • • • • 2 ? View details (2) 1: NCBI Accession Download × on : Downloaded Files a list

2

2+m ¢

00

T I Trit Code

1

n Galaxy France	祄 Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 📰		Using 3%
Tools ක්	Dataset Error Report	History	<b>2+⊡</b> ⊄
search tools	An error occurred while running the tool toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download	search datasets	00
1 Upload Data	/ncbl_acc_download/0.2.7+galaxy0.	Error example	
	Details	4 shown, 1 hidden	
Get Data	Execution resulted in the following messages:	772 b	
Send Data	Fatal error: Exit code 1 ()		
Collection Operations		4: NCBI Accession Downl	• • • ×
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 Downl	oad on X
Filter and Sort		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: <b>My job ended with an error. What can I do?</b>	wnload on : Log	
Convert Formats		error	
FΔ STΔ /FΔ STO	Issue Report	NCBI Entrez returned erro	or code 400. a
FASTO Quality Control	Please provide detailed information on the activities leading to this issue:	Failed to download file wi	th id AAG192
FASTQ Quality Control		<	>
SAM/BAM		₩ 8 0 8 2 ?	۲
BED		View or report this error	wnload X
VCF/BCF		on : Downloaded Files	
Nanopore	Report	a list	
COMMON GENOMICS TOOLS			
Operate on Genomic Intervals			
Fetch Alignments/Sequences			
GENOMICS ANALYSIS			

<b>=</b> Galaxy France	祄 Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 📰		Using 3%
Tools ☆	Dataset Error Report	History	S + 🗆 🕈
search tools	An error occurred while running the tool <b>toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download</b>	search datasets	00
🏦 Upload Data	Details	Error example 4 shown, 1 hidden	
Get Data	Execution resulted in the following messages:	772 b	۵ 🌑 🗹
Send Data Collection Operations	Fatal error: Exit code 1 ()	4: NCBI Accession Downlo ad on : Log	• # ×
GENERAL TEXT TOOLS	lool 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	<b>④ ∦ X</b>
Convert Formats FASTA/FASTQ FASTQ Quality Control SAM/BAM BED VCF/BCF	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 Rel O O O C C View or report this error I: NCBI Accession Down on : Downloaded Files a list	dataset: code 400, ar i d AAG1927 > (nioad X
COMMON GENOMICS TOOLS Operate on Genomic Intervals Fetch Alignments/Sequences GENOMICS ANALYSIS			

📮 Galaxy France	প Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 📰	-	Using 3%
Tools	☆ Dataset Error Report	History 🖌	3 + 🗆 💠
search tools	An error occurred while running the tool <b>toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download</b>	search datasets	00
🏦 Upload Data	Details	Error example 4 shown, 1 hidden	
Get Data	A Execution resulted in the following messages:	772 b	
Send Data Collection Operations	Fatal error: Exit code 1 ()	4: NCBI Accession Downlo	● # ×
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	id on X
Filter and Sort	Troubleshooting	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: <b>My job ended with an error. What can I do?</b>	2: NCBI Accession Do wnload on : Log	<b>④ ∦ X</b>
Convert Formats	Issue Report	error An error occurred with this o NCBI Entrez returned error	lataset: code 400. ar
FASTQ Quality Control	Please provide detailed information on the activities leading to this issue:	Failed to download file with	id AAG1927
SAM/BAM		₩ 8 8 9 2 ?	۲
BED		View or report this error	load X
VCF/BCF		on : Downloaded Files	
Nanopore	Report	a list	
COMMON GENOMICS TOOLS			
Operate on Genomic Intervals	Bug report sent to the local galaxy		
Fetch Alignments/Sequences	administrators		
GENOMICS ANALYSIS			

# ABMS Tools - Handle errors

Station Biologique Roscoff

📮 Galaxy France	🕋 Workflow Visualize - Shared Data - Help - User - 🖻 🏢		Using 3%
Tools	☆ Dataset Error Report	History	C + 🗆 🕻
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	Failed to download file with id AAG19274.1 from NCBI	3: NCBI Accession Downlo : Downloaded Files	ad on X
Filter and Sort Join, Subtract and Group GENOMIC FILE MANIPULATION	Troubleshooting There are a number of helpful resources to self diagnose and correct problems. Start here: <b>My job ended with an error. What can I do?</b>	2: NCBI Accession Do wnload on : Log	● # ×
Convert Formats FASTA/FASTQ FASTQ Quality Control	Issue Report Please provide detailed information on the activities leading to this issue:	error An error occurred with this NCBI Entrez returned error Failed to download file with	dataset: r code 400, a n id AAG192 <b>&gt;</b>
SAM/BAM BED		Image: Second state         Image: Second sta	mload X
VCF/BCF Nanopore	Report	on : Downloaded Files a list	
COMMON GENOMICS TOOLS Operate on Genomic Intervals	For usegalaxy.fr. create a new request on		
Fetch Alignments/Sequences GENOMICS ANALYSIS	https://community.france-bioinformatique.fr/c/galaxy		



# **HISTORY AND DATASETS**




## Current history, drag and drop datasets

🗧 Galaxy France		👫 Workflow Visualize <del>-</del> Shared	Data 🕶 Help 🕶 User 🕶 📰	Using 3%
search histories	8	search all datasets	••••	Create new
Current History		Switch to 🗸	Swit	ch to 🔹 Switch to 🔹
Trinity example 4 shown 40.34 KB	c 📚 🗩	Error example 4 shown, 1 hidden	rnaseq hands-on 2022 14 shown, 3 hidden 9.88 MB	blast hands-on 2022         8 shown, 10 hidden         66.36 MB
search datasets	00	search datasets	search datasets	search datasets
Drag datasets here to copy them to the cur 4: Trinity on data 2 and data 1: Gene to transcripts map	rrent history	4: NCBI Accession Download ③ 🖋 🗙 on : Log	17: RNA STAR on clata 2 and data 1: mapped.b m	A X 13: NCBI BLAST+ blastn across coll X ection 12 a list with 5 items
3: Trinity on data 2 and data 1: Assembled Transcripts	● / ×	3: NCBI Accession Download on : D X ownloaded Files a list with 1 item	16: RNA STAR on data 2 and data 1: splice junctions.bed	X     12: queries     X     a list with 5 items
2: reads.right.fq	⊙ # ×	2: NCBI Accession Downlo 💿 🖋 🗙 ad on : Log	15: RNA STAR on data 2 and ④ data 1: log	
1: reads.left.fq	• / ×	error	14: RNA STAR on data 2 and data 1: mapped.bam	# × a 2'
		NCBI Entrez returned error code 400, are ID( Failed to download file with id AAG19274.11	13: RNA STAR on data 2 and ④ data 1: splice junctions.bed	10: megablast query5.fa vs ' () () () () () () () () () () () () ()
		 ★ B Ø Ø ₽ ?	12: RNA STAR on data 2 and ④ data 1: log	<ul> <li>𝔅 ×</li> <li>𝔅 nucleotide BLAST databas</li> <li>𝔅 𝔅 ×</li> <li>𝔅 from data 1</li> </ul>
		1: NCBI Accession Download on     X     : Downloaded Files	8: MultiQC on data 6 and dat ④ a 4: Webpage	8: protein BLAST database fr () X om data 2
		a list	7: MultiQC on data 6 and data 4: St ats a list with 3 items	× 2: Drosophila_melanogaster. ④ 🖋 × BDGP6.22.pep.all.fa.gz
			6: FastQC on data 2: RawDat ④	1: Drosophila_melanogaster.       Image: Comparison of the second s



Station Biologique Roscoff History

<b>=</b> Galaxy France		🔥 Workflow Visua	lize 👻 Shared D	ata 🕶 Help 🕶 User 🕶 📰			Using 39	%
search histories	8	search all datasets	88.	**			Create	new
Current History	•		Switch to 🔹		Switch to 🔹		Switch to	•
Trinity example 4 shown 40.34 KB	v 🖏 🗩	Error example 4 shown, 1 hidden 772 b	v 🔊 🗩	rnaseq hands-on 2022 14 shown, 3 hidden 9.88 MB	v 🔊 🗩	<b>blast hands-on 2022</b> 8 shown, 10 hidden 66.36 MB	v 📎 🗩	
search datasets	00	search datasets	00	search datasets	00	search datasets	00	)
Drag datasets here to copy them to the c 4: Trinity on data 2 and data	current history	4: NCBI Accession Download on : Log	● / ×	17: RNA STAR on data 2 and data 1: mapped.bam	● # ×	13: NCBI BLAST+ blastn across ection 12 a list with 5 items	coll X	
3: Trinity on data 2 and data 1: Assembled Transcripts	• # ×	3: NCBI Accession Download of ownloaded Files a list with 1 item	on:D X	16: RNA STAR on data 2 and data 1: splice junctions.bed	• # ×	<b>12: queries</b> a list with 5 items	×	
2: reads.right.fq	• # ×	2: NCBI Accession Downlo ad on : Log		data 1: log	• # ×	11: blastx query5.fa vs 'prot ein BLAST database from dat a 2'	● / ×	
1: reads.left.fq	• # ×	error An error occurred with this data: NCBI Entrez returned error code Failed to download file with id A	set: e 400, are ID( AG19274.1 1	data 1: mapped.bam 13: RNA STAR on data 2 and data 1: splice junctions.bed	● # ×	10: megablast query5.fa vs ' nucleotide BLAST database f rom data 1'	⊕ # ×	
		< ∦ ◙ Ø Ø ₽ ?	> •	12: RNA STAR on data 2 and data 1: log	● # ×	9: nucleotide BLAST databas e from data 1	⊛ # ×	
		1: NCBI Accession Downloa     Downloaded Files	ad on 🗙	8: MultiQC on data 6 and dat a 4: Webpage	● # ×	8: protein BLAST database fr om data 2	⊛ # ×	
		a list		7: MultiQC on data 6 and data ats a list with 3 items	4: St 🗙	2: Drosophila_melanogaster. BDGP6.22.pep.all.fa.gz	⊛ # ×	
				6: FastQC on data 2: RawDat a	● / ×	1: Drosophila_melanogaster. BDGP6.22.cds.all.fa.gz	• # ×	74 ~



Station Biologique Roscoff History

search histories	8	search all datasets	00	•			Creater
urrent History	•	Si	witch to 🔹		Switch to 🔹		Switch to
Trinity example 4 shown 40.34 KB	2 🏷 🗩	Error example 4 shown, 1 hidden 772 b	v 🔊 🗩	rnaseq hands-on 2022 14 shown, 3 hidden 9.88 MB		<b>blast hands-on 2022</b> 8 shown, 10 hidden 66.36 MB	v 📎 🗩
search datasets	00	search datasets	00	search datasets	08	search datasets	00
Drag datasets here to copy them to the . 4: Trinity on data 2 and data	current history	4: NCBI Accession Download on : Log	● # ×	17: RNA STAR on data 2 and data 1: mapped.bam	● # ×	13: NCBI BLAST+ blastn across ection 12 a list with 5 items	coll X
1: Gene to transcripts map 3: Trinity on data 2 and data 1: Assembled Transcripts	⊛ # ×	3: NCBI Accession Download on : ownloaded Files a list with 1 item	DX	16: RNA STAR on data 2 and data 1: splice junctions.bed	• # ×	12: queries a list with 5 items	×
2: reads.right.fq	• # ×	2: NCBI Accession Downlo ad on : Log	● / ×	data 1: log 14: RNA STAR on data 2 and	• / ×	11: blastx query5.fa vs 'prot ein BLAST database from dat a 2'	⊛ # X
1: reads.left.fq	• # ×	error An error occurred with this dataset: NCBI Entrez returned error code 40 Failed to download file with id AAG	00, are ID( 19274.1 <del>1</del>	data 1: mapped.bam 13: RNA STAR on data 2 and data 1: splice junctions.bed	● # ×	10: megablast query5.fa vs ' nucleotide BLAST database f rom data 1'	⊕ # ×
		< x ⊠ Ø € 2 ?	> •	12: RNA STAR on data 2 and data 1: log	● # ×	9: nucleotide BLAST databas e from data 1	● / ×
		1: NCBI Accession Download      : Downloaded Files	on ×	8: MultiQC on data 6 and dat a 4: Webpage	● / ×	8: protein BLAST database fr om data 2	⊛ # ×
		a list		7: MultiQC on data 6 and data ats a list with 3 items	4: St 🗙	2: Drosophila_melanogaster. BDGP6.22.pep.all.fa.gz	⊛ # ×
				6: FastQC on data 2: RawDat	● / ×	1: Drosophila_melanogaster. BDGP6.22.cds.all.fa.gz	● # ×



### Search histories and datasets

📮 Galaxy France		😭 Workflow Visu	ialize 🔻 Shared D	Data 🕶 Help 🕶 User 🕶 💼			Using 3%
search histories	8	search all datasets	00	***			Create r
urrent History	*	91	Switch to 🔹		Switch to 🔹		Switch to
<b>Trinity example</b> 4 shown 40.34 KB	v 🔊 🗩	Error example 4 shown, 1 hidden 772 b	v 🖏 🗩	rnaseq hands-on 2022 14 shown, 3 hidden 9.88 MB	2 🏷 🗩	blast hands-on 2022 8 shown, 10 hidden 66.36 MB	2 🌒 🗩
search datasets	00	search datasets	00	search datasets	00	search datasets	00
Drag datasets here to copy them to the of 4: Trinity on data 2 and data 1: Gene to transcripts map	current history	4: NCBI Accession Download on : Log		17: RNA STAR on data 2 and data 1: mapped.bam	● # ×	13: NCBI BLAST+ blastn acros ection 12 a list with 5 items	s coll X
3: Trinity on data 2 and data 1: Assembled Transcripts		3: NCBI Accession Download ownloaded Files a list with 1 item	on:D X	16: RNA STAR on data 2 and data 1: splice junctions.bed	• # ×	<b>12: queries</b> a list with 5 items	×
2: reads.right.fq	● # ×	2: NCBI Accession Downlo ad on : Log	• • • ×	data 1: log		11: blastx query5.fa vs 'prot ein BLAST database from dat	● / ×
1: reads.left.fq	⊛ # ×	error	11.022	14: RNA STAR on data 2 and data 1: mapped.bam	• # ×	a 2'	
		An error occurred with this dat NCBI Entrez returned error co Failed to download file with id	aset: de 400, are ID( AAG19274.1 1	13: RNA STAR on data 2 and data 1: splice junctions.bed	⊛ # ×	10: megablast query5.fa vs ' nucleotide BLAST database f rom data 1'	<b>④</b> ∦ X
		x 8 Ø € ₽ ?	» •	12: RNA STAR on data 2 and data 1: log	● # ×	9: nucleotide BLAST databas e from data 1	⊛ # ×
		1: NCBI Accession Downlo     : Downloaded Files	oad on 🗙	8: MultiQC on data 6 and dat a 4: Webpage	● # ×	8: protein BLAST database fr om data 2	⊛ # ×
		a list		7: MultiQC on data 6 and data ats a list with 3 items	4: St 🗙	2: Drosophila_melanogaster. BDGP6.22.pep.all.fa.gz	● / ×
				6: FastQC on data 2: RawDat	● # ×	1: Drosophila_melanogaster. BDGP6.22.cds.all.fa.gz	● # × 7

>



Station Biologique Roscoff

ARMS History



#### The Current 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.

Frequently Asked Questions

Citing this Tutorial

Feedback

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



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

Tools රු	TRINITY_DN2_c0_g1	TRINITY_DN2_c0_g1_i1	History	្ល+០≮
search tools	TRINITY_DN1_c0_g1	TRINITY_DN1_c0_g1_i1	search datasets	00
	TRINITY_DN5_c1_g1	TRINITY_DN5_c1_g1_i1	Trinity example	
🏦 Upload Data	TRINITY_DN0_c0_g2	TRINITY_DN0_c0_g2_i1	4 shown	
Get Data	TRINITY_DN0_c0_g3	TRINITY_DN0_c0_g3_i1	40.34 KB	
Send Data	TRINITY_DN3_c0_g1	TRINITY_DN3_c0_g1_i1		
Collection Operations			4: Trinity on data 2 a	ndd 💿 🖋 🗙
SENERAL TEXT TOOLS			ata 1: Gene to transc map	View data
Text Manipulation			3: Trinity on data 2 ar	dd 👁 🥒 ک
ilter and Sort			ata 1: Assembled Trai	nscri
oin, Subtract and Group			pts	
ENOMIC FILE MANIPULATION			2: reads.right.fq	• / >
onvert Formats			1: reads.left.fq	• / :
ASTA/FASTQ				
ASTQ Quality Control				
AM/BAM				
ED				
CF/BCF				
lanopore				
OMMON GENOMICS TOOLS				
perate on Genomic Intervals				
etch Alignments/Sequences				
ENOMICS ANALYSIS				70



### Rename, annotate, change datatype...

📮 Galaxy France	👚 Workflow Visualize - Shared Data - Help - User - 🖻 🏢		Using 3%
Tools රු	Edit dataset attributes	History	S+0\$
search tools		search datasets	00
Lupload Data	Edit attributes DAuto-detect Dave	<b>Trinity example</b> 4 shown 40.34 KB	2 🏷 🗩
Collection Operations	Trinity on data 2 and data 1: Gene to transcripts map	4: Trinity on data 2 and	d 👁 🖉 🗙
GENERAL TEXT TOOLS	Info	ata 1: Gene to transcript map	Edit attributes
Text Manipulation		3: Trinity on data 2 and d	
Filter and Sort		ata 1: Assembled Transcr	i
Join, Subtract and Group		2: sonde vield fr	
GENOMIC FILE MANIPULATION	Annotation	2: reads.right.rq	• / ×
Convert Formats		1: reads.left.fq	⊙ # ×
FASTA/FASTQ			
FASTQ Quality Control			
SAM/BAM	Add an annotation or notes to a dataset; annotations are available when a history is viewed.		
BED	Database/Build		
VCF/BCF	Additional Species Are Below		
Nanopore	Number of comment lines		
COMMON GENOMICS TOOLS			
Operate on Genomic Intervals			
Fetch Alignments/Sequences		1	
GENOMICS ANALYSIS			79

1

>

https://usegalaxy.fr/datasets/edit?dataset\_id=9ba7911219dddd3c



### Rename, annotate, change datatype...

ools 2	Edit dataset at	tributes				History	<b>₽+□ \$</b>
search tools	<b>≡</b> Attributes	Convert	Datatypes	Permissions		search datasets	00
🏦 Upload Data	Change dataty	rpe		<b>D</b> Detect datatype	<b>≓</b> Change datatype	Trinity example	
jet Data	New Type					40.34 KB	
end Data							
ollection Operations	tabular					4: Trinity on data 2 an	⊛ / ×
ENERAL TEXT TOOLS					٩	cripts map	
ext Manipulation	trk				^	7 lines	
ter and Sort	trr					format: <b>tabular</b> , database:	?
in, Subtract and Group	ttl						
NOMIC FILE MANIPULATION	twobit					D)   _	
onvert Formats	txt						
ASTA/FASTQ	uniprotxml						
ASTO Quality Control					•		
M/BAM						TRINITY_DN2_c0_g1 TRINITY_DN2_c0	0_g1_i1
						TRINITY_DN1_C0_g1 TRINITY_DN1_c	0_g1_i1
:0						TRINITY_DN5_c1_g1 TRINITY_DN5_c1	1_g1_11 0_g1_i1
CF/BCF						TRINITY_DN0_C0_g2 TRINITY_DN0_C	0_g2_i1
anopore						3: Trinity on data 2 and	• / ×
OMMON GENOMICS TOOLS						data 1: Assembled Tran	
perate on Genomic Intervals						scripts	
etch Alignments/Sequences						2: reads.right.fq	● / ×
NOMICS ANALYSIS						100 sequences	80



## Download

Tools ☆			History	C + 🗆 🕈
search tools			search datasets	00
🍰 Upload Data			Trinity example	
Get Data			4 shown 40.34 KB	Image: Second
Send Data		a a a la va v fa		
Collection Operations	welcome to us	segalaxy.tr	4: Trinity on data 2 and	a ⊛∦×
SENERAL TEXT TOOLS	By using this Galaxy instance, we	assume that you have read and accept the Term Of Use	data 1: Gene to transci pts map	ri
Text Manipulation	For any questions or support: co	mmunity.cluster.france-bioinformatique.fr/c/galaxy	7 lines	
Filter and Sort		â	format: <b>tabular</b> , databas	se: ?
oin, Subtract and Group	• 22/07/2021: usegalaxy.fr is	now running the <b>release 21.05</b> of Galaxy. Please check		
GENOMIC FILE MANIPULATION	the 21.05 user release note		D)   _	
Convert Formats		~		
FASTA/FASTQ			₿&₿₽₩?	
ASTQ Quality Control			Download 2	
SAM/BAM	A THE	Ack the	TRINITY_DN2_C0_BI TRINITY_DN2	C0_g1_i1
BED		ASKUTE	TRINITY_DN1_C0_g1 TRINITY_DN1	_ce_g1_11 5_c1_g1_i1
VCF/BCF		CalavyCat	TRINITY_DN0_C0_g1 TRINITY_DN0	)_c0_g1_i1 3 c0_g2_i1
Nanopore		GalaxyCat		
COMMON GENOMICS TOOLS		Can't find a tool on	3: Trinity on data 2 and data 1: Assembled Tra	i ⊛∦× n
Operate on Genomic Intervals		usegalaxy.fr?	scripts	
etch Alignments/Sequences		Why not search on the IFB	2: reads.right.fq	⊙ / ×
GENOMICS ANALYSIS		Galaxy Catalog ?	100 sequences	81



## Link to data

iools රු			History	2+□¢
search tools			search datasets	00
🏝 Upload Data			Trinity example	
Get Data			4 shown 40.34 KB	Image: Second
Send Data	Welcome to us	segalavy fr		
Collection Operations	vvelcome to us	segalaxy.II	4: Trinity on data 2 an	id 💿 🖋 🗙
SENERAL TEXT TOOLS	By using this Galaxy instance, we	assume that you have read and accept the Term Of Use	pts map	
ext Manipulation	For any questions or support: con	mmunity.cluster.france-bioinformatique.fr/c/galaxy	7 lines	
ilter and Sort			format: <b>tabular</b> , databa	ise: ?
oin, Subtract and Group	• 22/07/2021: usegalaxy.fr is	now running the <b>release 21.05</b> of Galaxy. Please check		
ENOMIC FILE MANIPULATION	the 21.05 user release note		D)   _	
Convert Formats		~		1
ASTA/FASTQ				
ASTQ Quality Control			Conv link 2	
SAM/BAM			TRINITY_DN2_c0_g1 TRINITY_DN	12_00_g1_i1
BED	19	ASK LNE	TRINITY_DN1_c0_g1 TRINITY_DN TRINITY_DN5 c1 g1 TRINITY_DN	1_c0_g1_i1 45 c1 g1 i1
/CF/BCF		CalauriCat	TRINITY_DN0_c0_g1 TRINITY_DN	10_c0_g1_i1
Vanopore		GalaxyCat	TRINITY_DN0_c0_g2 TRINITY_DN	0_c0_g2_i1
COMMON GENOMICS TOOLS		Can't find a tool on	3: Trinity on data 2 an	id 👁 🖋 🗙
Operate on Genomic Intervals		usedalavyfr?	scripts	in
etch Alianments/Sequences		Why not search on the IFB	2: reads.right.fg	• / ×
				121100



### Informations

search tools       Dataset Information       search datasets	Tools స్ట్రీ	Trinity		History	2+0\$
Number       4         Aune       Trinity on data 2 and data 1: Gene to transcripts map         Get Data       Filesce       273 b         Send Data       Filesce       273 b         Collection Operations       Filesce       273 b         Get Data       Filesce       273 b         Collection Operations       File contents       addata 1: Gene to transcripts map         Get Data       Dikey       ?         Format       tabular       -         File contents       contents       -         History Content API ID       955/9712920ea7468       -         UUD       a479re18-248-4027-89fb-e744e47d1940       -         Filter and Sort       Tool Parameters       Value         Join, Subtract and Group       Figure Parameter       Value         Gatasets?       Paired or Single-end data? paired       -         Left/Forward strand       • 1: readsleft.fg       -         reads       Strand specific data       faite       -         Strand specific data       faite       -       -         BED       Jacard Lip option faite       -       -         VC/F/GCF       Run in slico normalization or malization or malization or mease       -       -	search tools	Dataset Informatio	n	search datasets	00
1     Upload Data     Name     Trinity on data 2 and data 1: Gene to transcripts map     Trinity example       Get Data     Created     Tuesday Jan 11th 52:501 2022 UTC     40.34 KB     40.34 KB     1       Send Data     Dokey     ?		Number	4		
Created       Tuesday Jan 11th 9:25:01 2022 UTC       4 shown         Set Data       Filesize       273 b         Set Data       Filesize       273 b         Collection Operations       Filesize       273 b         Filesize       273 b       40.34 KB         Collection Operations       File contents       contents         File contents       contents       data 1: Gene to trans         cripts map       Yines       cripts map         repts mat       tabular       data 1: Gene to trans         cripts map       Yines       cripts map         rine and Sort       Tool Parameter       Value         Anstructure       Value       Vines         Activity Control       Are you pooling sequence datasets?       Vines details?         ASTA/FASTQ       Paired or Single-end data? paired       Left/Forward strand       1: reads.ight.fq         reads       Strand specific data       faise       3: Trinity on data 2 and (or data)         ASTA/FASTQ       Strand specific data       faise       3: Trinity on data 2 and (or data)         AASTQ Quality Control       Right/Revers strand       -2: reads.inght.fq       3: Trinity on data 2 and (or data)         Astand specific data       faise       3: Trinity on data 2	🏦 Upload Data	Name	Trinity on data 2 and data 1: Gene to transcripts map	Trinity example	
iet Data       Filesize       273 b       40.34 KB       1         end Data       Dokey       ?       ?       ?         collection Operations       File contents       contents       ?       ?         iet operations       File contents       contents       ?       ?       ?         iet operations       File contents       contents       ?       ?       ?       ?         iet operations       File contents       contents       ? <t< td=""><td></td><td>Created</td><td>Tuesday Jan 11th 9:25:01 2022 UTC</td><td>4 shown</td><td></td></t<>		Created	Tuesday Jan 11th 9:25:01 2022 UTC	4 shown	
Bend Data     Dbkey     ?       Format     tabular       Format     tabular       Format     tabular       File contents     contents       History Content API ID     055d73e92eaa7468       UUD     a479ce18-2fa8-4d27-89fb-e744e47d19d0       Tool Parameters     format:       UUD     a479ce18-2fa8-4d27-89fb-e744e47d19d0       Tool Parameters     format:       History API ID     055d73e92eaa7468       UUD     a479ce18-2fa8-4d27-89fb-e744e47d19d0       Tool Parameters     format:       History Support     Yes       datasts?     format:       ASTAVFASTQ     Paired or Single-end data? paired       AM/BAM     Right/Forward strand       Right/Reverse strand     - 2: reads.right.fq       reads     Jaccard Clip options       AM/BAM     Right/Reverse strand       LED     additional_params       Manopore     datast       OMMON GENOMICS TOOLS     additional_params       Minimum Contig Length     200       Use Holder     Strinity on data 2 and P       It reads.ieft.fg     2: reads.right.fg       Paired or Group     node?       Run in silico normalization     True       OMMON GENOMICS TOOLS     additional_params	Get Data	Filesize	<b>273</b> b	40.34 KB	
Format       tabular         Format       tabular         File contents       contents         File contents       contents         History API ID       95a7911219ddd3c         History API ID       95a7912292ea37468         UUID       a479ce18-2fa8-4d27-89fb-e744e47d19d0         Tool Parameters       Imput Parameter         VUID       a479ce18-2fa8-4d27-89fb-e744e47d19d0         ENOMIC FILE MANIPULATION       Are you pooling sequence datasets?         ASTA/FASTQ       Paired or Single-end data? paired         Left/Forward strand       - 1: reads.left.fq         reads       - 2: reads.right.fq         reads       False         MAMON GENOMICS TOOLS       Additional_params         OMMON GENOMICS ANALYSIS       Error-corrected or dricular mode?	end Data	Dbkey	?		
collection Operations       File contents       contents         ceNERAL TEXT TOOLS       History Content API ID       9ba7911219ddd43c         ext Manipulation       UUD       a479ce18-2fa8-4d27-89fb-e744e47d19d0         UUD       a479ce18-2fa8-4d27-89fb-e744e47d19d0       7 lines         rines       format tabular, database: ?       7         Convert Formats       Mate You pooling sequence datasets?       7         ASTQ Quality Control       Paired or Single-end data? paired       11: reads.left.fq         AM/BAM       reads       + 11: reads.left.fq       11: reads.left.fq         ED       Strand specific data       false       11: reads.left.fq         Itab 1: some guided       reads       false       11: matry.ge.ge.ge.ge.ge.ge.ge.ge.ge.ge.ge.ge.ge.		Format	tabular	4: Trinity on data 2 an	x
ENERAL TEXT TOOLS       History Content API ID       9ba7911219ddd43c       cripts map         ext Manipulation       History API ID       055d73e92eaa7468       7 lines         iller and Sort       UUID       a479ce18-2fa8-4d27-89fb-e744e47d19d0       7 lines         oin, Subtract and Group       Input Parameter       Value       111111111111111111111111111111111111	Collection Operations	File contents	contents	d data 1: Gene to trans	
ext Manipulation     History API ID     055d73e92eaa7468     7 lines       uUID     a479ce18-2fa8-4d27-89fb-e744e47d19d0     7 lines       format: tabular, database: ?     format: tabular, database: ?       iiter and Sort     TOOl Parameters       iiter and Sort     Input Parameter       Value     Are you pooling sequence       convert Formats     Are you pooling sequence       datasets?     Yes       datasets?     Paired or Single-end data? paired       Left/Forward strand     -1: readsleft.fq       reads     -1: readsleft.fq       reads     false       strand specific data     false </td <td>ENERAL TEXT TOOLS</td> <td>History Content API ID</td> <td>9ba7911219ddd43c</td> <td>cripts map</td> <td></td>	ENERAL TEXT TOOLS	History Content API ID	9ba7911219ddd43c	cripts map	
UUID     a479ce18-2fa8-4d27-89fb-e744e47d19d0       iitler and Sort     Tool Parameters       bin, Subtract and Group     Input Parameter       Input Parameter     Value       Are you pooling sequence datasets?     Are you pooling sequence datasets?       Convert Formats     Are you pooling sequence datasets?       ASTA/FASTQ     Paired or Single-end data? paired       ASTQ Quality Control     Left/Forward strand       AM/BAM     Right/Reverse strand       • 2: reads.right.fq     reads       reads     False       Run in silico normalization of reads     True       OMMON GENOMICS TOOLS     Minimum Contig Length       Operate on Genomic Intervals     Vise the genome guided mode?       etch Alignments/Sequences     ro       ENOMICS ANALYSIS     Error-corrected or circular consensity for correctied or circular	ext Manipulation	History API ID	055d73e92eaa7468	7 lines	
Interfact and Group       Tool Parameters         inin, Subtract and Group       Input Parameter       Value         inin, Subtract and Group       Input Parameter       Value         inin, Subtract and Group       Are you pooling sequence datasets?       Ves         convert Formats       Are you pooling sequence datasets?       Ves         ASTA/FASTQ       Paired or Single-end data? paired       Lift/Forward strand         Left/Forward strand       • 1: reads.left.fq       Input Parameter         reads       Right/Reverse strand       • 2: reads.right.fq       Input Parameter         SED       Strand specific data       false       Input Parameter         VCF/BCF       Jaccard Clip options       False       Input Params         Nanopore       additional_params       True       3: Trinity on data 2 and mode?       Imput Parameter         Operate on Genomic Intervals       Use the genome guided mode?       no       2: reads.right.fq       I: reads.left.fg         SED       Use the genome guided mode?       no       2: reads.right.fq       I: tastry.pae.eg.gl: intastry.pae.eg.gl: in	iltor and Sort	UUID	a479ce18-2fa8-4d27-89fb-e744e47d19d0	format: <b>tabular</b> , database	: ?
EENOMIC FILE MANIPULATION       Are you pooling sequence datasets?       Ves         Convert Formats       Are you pooling sequence datasets?       Ves         ASTA/FASTQ       Paired or Single-end data? paired       IIII J.IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	oin, Subtract and Group	Tool Parameters	Value		
ASTA/FASTQ ASTA/FASTQ ASTA/FASTQ ASTA/CONTOL ASTQ Quality Control AM/BAM AED AED AED AED AED AED ACF/BCF Nanopore COMMON GENOMICS TOOLS Deprate on Genomic Intervals COMMON GENOMICS TOOLS Deprate on Genomic Intervals CENOMICS ANALYSIS ENOMICS ANALYSIS ENOMICS ANALYSIS Paired or Single-end data? paired Left/Forward strand • 1: reads.left.fg Paired or Single-end data? paired Deprate on Genomic Intervals EFRO-Corrected or circular SENOMICS ANALYSIS	GENOMIC FILE MANIPULATION	Are you pooling sequence datasets?	Yes	[∥ D)      _	
ASTQ Quality Control   ASTQ Quality Control   AM/BAM   Right/Reverse strand   *2: reads.right.fq   reads   Strand specific data   Jaccard Clip options   False   Run in silico normalization of reads   of mades   MMNON GENOMICS TOOLS   Minimum Contig Length   200   Use the genome guided mode?   no   ETOMICS ANALYSIS		Paired or Single-end data	? paired		
AM/BAM       Right/Reverse strand       2: reads.right.fq         reads       Strand specific data       false         Jaccard Clip options       False         Run in silico normalization of reads       True         AM/ON GENOMICS TOOLS       additional_params         Minimum Contig Length       200         Use the genome guided mode?       no         EFNOMICS ANALYSIS       Error-corrected or circular	ASTQ Quality Control	Left/Forward strand reads	• 1: reads.left.fq	Image: Solution of the second seco	
SED       Strand specific data       false         Jaccard Clip options       False         Janopore       Run in silico normalization of reads       True         SOMMON GENOMICS TOOLS       additional_params         Minimum Contig Length       200         Use the genome guided mode?       no         Error-corrected or circular       Fror-corrected or circular         Error-corrected or circular       I: reads.left.fq	AM/BAM	Right/Reverse strand reads	• 2: reads.right.fq	TRINITY_DN2_c0_g1 TRINITY_DN2_c TRINITY_DN2_c0_g1 TRINITY_DN2_c	.0_g1_i1 :0_g1_i1
Jaccard Clip options       False         Run in silico normalization of reads       True         additional_params       additional_params         Minimum Contig Length       200         Use the genome guided mode?       no         Error-corrected or circular       Error-corrected or circular         ERNOMICS ANALYSIS       Error-corrected or circular	ED	Strand specific data	false	TRINITY_DN5_c1_g1 TRINITY_DN5_c	:1_g1_i1
Run in silico normalization   Nanopore   Nanopore   common of reads   additional_params   additional_params   Minimum Contig Length   Use the genome guided   mode?   Error-corrected or circular   Error-corrected or circular   Consensus (CCS) pac bio	ICE/BCE	Jaccard Clip options	False	TRINITY_DN0_c0_g1 TRINITY_DN0_c	:0_g1_i1
COMMON GENOMICS TOOLS       additional_params       3: Trinity on data 2 and data 1: Assembled Transcripts         Operate on Genomic Intervals       Minimum Contig Length 200       scripts         Use the genome guided mode?       no         Error-corrected or circular       Error-corrected or circular         CENOMICS ANALYSIS       COS pac bio	Vanopore	Run in silico normalization of reads	True	TRINITY_DN0_c0_g2 TRINITY_DN0_c	0_g2_i1
Minimum Contig Length     200     scripts       Deparate on Genomic Intervals     Minimum Contig Length     200       Use the genome guided mode?     no     2: reads.right.fq       Error-corrected or circular     consensus (CCS) pac bio     1: reads.left.fg	COMMON GENOMICS TOOLS	additional_params		3: Trinity on data 2 and	• / ×
etch Alignments/Sequences mode? no 2: reads.right.fq • / Error-corrected or circular consensus (CCS) pac bio	Operate on Genomic Intervals	Minimum Contig Length	200	scripts	
ENOMICS ANALYSIS Error-corrected or circular 1: reads.left.fg	etch Alignments/Sequences	mode?	no	2: reads.right.fq	⊙ # ×
consensus (ccs) pacisio	SENOMICS ANALYSIS	Error-corrected or circular consensus (CCS) pac bio		1: reads.left.fq	● # 83



## Re-run a job

Fools 🖒						History	\$+□¢
search tools	Version 2.9.1	vo assembly of RNA-Seq data (Galaxy +galaxy2)	☆ Favorite	& Versions	▼ Options	search datasets	00
1. Upload Data	Are you pooling	g sequence datasets?				Trinity example	
	Yes				•	4 shown	
Get Data	Paired or Sing	le-end data?				40.34 KB	Image: Second
Send Data	Paired-end				•		
Collection Operations	Left/Forward	d strand reads				4: Trinity on data 2 and data 1: Gene to transcri	● # ×
SENERAL TEXT TOOLS		3: Trinity on data 2 and data 1: Assem	bled Transcripts		^ <b>t</b> 🕞	pts map	
Fext Manipulation		2: reads.right.fq	iorea manocripto			7 lines	
ilter and Sort		1: reads.left.fq				format: <b>tabular</b> , database	5 <b>7</b>
oin, Subtract and Group							
ENOMIC FILE MANIPULATION	1.1.1.1				4	D)   _	
Convert Formats	(left)	se strand reads					_
ASTA/FASTQ		3: Trinity on data 2 and data 1: Assem	bled Transcrints		^	B & B 2 !!! ?	
ASTQ Quality Control		2: reads.right.fq	iorea manacripta			1 Run this job again	
AM/BAM		1: reads.left.fq				TRINITY_DN2_c0_g1 TRINITY_DN2_(	c0_g1_i1
ED						TRINITY_DN1_c0_g1 TRINITY_DN1_( TRINITY_DN5_c1_g1 TRINITY_DN5_(	:0_g1_i1 c1_g1_i1
CF/BCF	(right)				1	TRINITY_DN0_c0_g1 TRINITY_DN0_c	0_g1_i1
lanopore	Strand speci	fic data				TRINITY_DW8_C8_g2 TRINITY_DW8_C	:0_g2_11
OMMON GENOMICS TOOLS	No					3: Trinity on data 2 and data 1: Assembled Tran	● # ×
perate on Genomic Intervals	Jaccard Clip	options				scripts	
etch Alignments/Sequences	No No	en adremente 6620				2: reads.right.fq	
ENOMICS ANALYSIS	Set if you exp	ect high gene density with UTR overlap (	iaccard clip)			100 sequences	0.4



### Visualisation tools



## Station Biologique Roscoff Dataset

## Help

Tools		History	€+⊡¢
search tools		search datasets	00
🍰 Upload Data		Trinity example	
Set Data		4 shown	
end Data		10.04 KB	
Collection Operations	Welcome to usegalaxy.fr	4: Trinity on data 2 and	• / ×
ENERAL TEXT TOOLS	By using this Galaxy instance, we assume that you have read and accept the Term Of Use	data 1: Gene to transcr pts map	i
ext Manipulation	For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy	7 lines	
ilter and Sort		format: <b>tabular</b> , database	e: <b>?</b>
oin, Subtract and Group	• 22/07/2021: usegalaxy.fr is now running the <b>release 21.05</b> of Galaxy. Please check		
SENOMIC FILE MANIPULATION	the 21.05 user release notes.	D)   _	
Convert Formats			
ASTA/FASTQ		₿002₩?	
ASTQ Quality Control		1 Tool Help	
AM/BAM	Ack the	TRINITY_DN2_c0_g1 TRINITY_DN2_	c0_g1_i1
ED	ASK LITE	TRINITY_DN5_c1_g1 TRINITY_DN5_	c1_g1_i1
/CF/BCF	GalavyCat	TRINITY_DN0_c0_g1 TRINITY_DN0_ TRINITY_DN0_c0_g2 TRINITY_DN0	c0_g1_i1 c0_g2_i1
lanopore	GalaxyCat	Tool help for Trinity	
COMMON GENOMICS TOOLS	Can't find a tool on		
)perate on Genomic Intervals	usegalaxy.fr?	Trinity assembles transcri	ipt sequences lata
etch Alignments/Sequences	Why not search on the IFB	i on indinina tote seq u	ara.
ENOMICS ANALYSIS	Galaxy Catalog ?	3: Trinity on data 2 and	@ # x <sup>86</sup>
ths://usenalavy.fr/#			





# 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



# Hands-on **TOOLS**







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

## Galaxy Training! Introduction to Galaxy Analyses Languages - Help - Extras - Search Tutorials

## Galaxy 101



#### Overview

#### ⑦ 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
- 🕿 Level: Introductory 🕿 📼 📼

#### Supporting Materials:

🗑 Topic Overview slides 🕒 Datasets 🔩 Workflows 🔳 GTN Video Library 👻 🌐 Available on these Galaxies 💌

Last modification: Nov 17, 2021

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

#### Introduction Setting the stage:

Exons and SNPs

Analysis

Galaxy management

Conclusion Frequently Asked

Questions Feedback

Citing this Tutorial

## Introduction

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.

#### D 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.

Agenda



# Deleting DATASET



#### Downloading and Deleting Data in Galaxy

Authors: 🐊 Simon Gladman ) 💮 Helena Rasche ) 😫 Add Contributions!

#### Overview

#### ⑦ Questions:

- How can I efficiently download my data from Galaxy once I've completed my analyses?
- · How do I delete old data to make more room in my guota?

#### Objectives:

- Have a greater knowledge of how Galaxy handles data downloads and deletions
- · Be able to successfully download and delete their own data

#### Time estimation: 20 minutes

🞓 Level: Introductory 🞓 🕋 👘

Last modification: Oct 5, 2020

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

#### Introduction

This tutorial outlines the process to get your data out of Galaxy and to delete it from Galaxy afterwards.

After you have completed your analysis on a Galaxy server, you may need to download the results for use in reports, papers, or other requirements. Galaxy has a myriad of methods for downloading data from individual datasets to a collection of data to an entire history along with all of its metadata. The first section of this tutorial will walk you through all of these methods.

#### Introduction

Downloading Data from Galaxy

Deleting Data from Galaxy

Frequently Asked Questions Feedback

Citing this Tutorial

On some Galaxy servers, user disk quotas have been enabled meaning that users are only able to store a certain amount of data on that particular Galaxy server. Quotas can differ on different servers. User disk quotas for the usegalaxy.\* servers are shown in the table below:

Galaxy Server	Unregistered User	Registered User
usegalaxy.org	5 GB	250 GB
usegalaxy.eu	5 GB	250 GB or 500 GB for Elixir Members
usegalaxy.org.au	5 GB	100 GB or 600GB for Australian Users

Other Galaxy servers may have different quota systems.

As you use a Galaxy server and upload data, perform analyses etc. you will use up your quota. If you have filled your quota and you want to do some new work, you can download your old work and delete it off the Galaxy 91 server to free up quota space. Galaxy uses a two step process to delete user data from disk. The first step is to delete the data , this marks the data as deleted but it can be undeleted for a cartain time 1. The second stan is to



Station Biologique Roscoff

https://training.galaxyproject.org /training-material/topics/galaxy-i nterface/tutorials/download-dele te-data/tutorial.html#deleting-da ta-from-galaxy

 $\odot$   $\odot$ 



# Datatypes DATASET



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

🗧 Galaxy France		📣 Workflow	Visualize 🛪 Shared D	ata <del>▼</del> Help <del>▼</del> User	• 🚖 🏢				Using 3%
Tools	ownload from web or	upload from c	disk					Ŕ	*==
search tools	Regular Composite	Collection	Rule-based					atasets	00
🏦 Upload Data		You added 1	1 file(s) to the queue. Ac	ld more files or click 'St	tart' to proceed.		^	ample	
Get Data	Name	Size	Туре	Genome	Settings	Status			
Send Data	Query1.fa	<b>3.1</b> КВ	Auto-de 🖌 🔍 🦷	Additional 💌	•		Û		
Collection Operations				٩				n data 2 and d	⊙ # ×
GENERAL TEXT TOOLS			Auto-detect	<u>^</u>					
Text Manipulation			ab1					n data 2 and d	⊙ # ×
Filter and Sort			agilentbrukeryep.d.tar	r.				mbled Transcri	
Join, Subtract and Group			arff					aht fa	O A Y
GENOMIC FILE MANIPULATIC			asn1					Jurid	
Convert Formats		Syne (set all):	asn1-binary	× siD:	A		×.	ft.fq	<i>● ∦</i> ×
FASTA/FASTQ		ype (sec an).		any.	Additional				
FASTQ Quality Control	1	Choose local file	s 🖻 Choose remote	files Paste/Feto	h data Start	Pause Reset	t Close		
SAM/BAM			RICHA						
BED			riole	UNL					
VCF/BCF			The thematic fo	r functional					
Nanopore			analysis and exp	oloration of					
COMMON GENOMICS TOOLS			proteomics and						
Operate on Genomic Interva	als		transcriptomics	data in					
Fetch Alignments/Sequence	S		biomedical rese	arch					
GENOMICS ANALYSIS	~		Go to Pr	oteoRE					



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 #	3 <b>+ 🗆 🌣</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	• / ×
FASTA/FASTQ			
FASTQ Quality Control	D-1		
SAM/BAM	Ack the		
BED	ASK LITE		
VCF/BCF	Galaxy/Cat		
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 ?		) E
(			



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 రో	7	^	History	3 <b>+ 0 \$</b>
search tools			search datasets	00
🏝 Upload Data			Trinity example	
Get Data			40.34 KB	
Send Data	Welcome to usedalaxy fr			
Collection Operations	Welconne to usegulaxy.h		4: Trinity on data 2 and d ata 1: Gene to transcripts	⊙ ∥ ×
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	⊙ / ×
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		pis	
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	Galavy/Cat			
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 ?			
<		~		98 >



# 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
gtt3
                       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.



100



• Change the datatype of a dataset

💶 Galaxy France	isualize - Shared Data - Help - User - 💼 🏢		Using 3%
Tools ☆	Edit dataset attributes	History	S+ 🗆 💠
search tools	Attributes	search datasets	00
🏝 Upload Data	Change datatype	Trinity example	
Get Data Send Data	New Type	40.34 KB	
Collection Operations	tabular -	4: Trinity on data 2 a d data 1: Gene to tra	
GENERAL TEXT TOOLS	trk ^	<b>cripts map</b> 7 lines	
Filter and Sort	trr tsv	format: <b>tabular</b> , datab	ase: ?
Join, Subtract and Group	ttl		
Convert Formats	txt		-1
FASTA/FASTQ	uniprotxml 🗸	₿₽₿₽₩?	•
SAM/BAM		1 2 TRINITY_DN2_C0_g1 TRINITY_D	N2_c0_g1_i1
BED		TRINITY_DNL_C0_g1 TRINITY_D	N5_c1_g1_i1
VCF/BCF		TRINITY_DN0_C0_g2 TRINITY_C	N0_C0_g2_i1
COMMON GENOMICS TOOLS		3: Trinity on data 2 an data 1: Assembled Tra	d 🕑 🇨 🗙
Operate on Genomic Intervals		scripts	101



# 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	• / ×
	Left/Forward strand reads	100 sequences format: <b>fastg</b> , database:	?
	No fasta, fastqsanger or fastqsanger.gz dataset available.	uploaded fastosander fi	•
		G51DEPAAYY188284 -1 -188 - 18494 -1	2070/2
	4	CTCAAATGGTTAATTCTCAGGCTGCAAAT	ATTCGTTCAGGATGGAAG
	(left)	+	
	Right/Reverse strand reads	@61DFRAAXX100204:1:100:10497::	13422/2
	🗘 🗅 No fasta, fastqsanger or fastqsanger.gz dataset available.	<	>
		1: reads.left.fq	⊙ / ×
		100 sequences	
		format: <b>fastq</b> , database:	<b>'</b>
	(right)	uploaded fastqsanger fil	e
	Strand specific data	801	۵
	No	261DFRAAXX100204:1:100:10494:	3878/1 102



# Dataset - Datatypes

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

	Trinity of Version Are you po Yes Paired o Paired-e	le novo assembly of RNA-Seq data (Galaxy 2.9.1+galaxy2) coling sequence datasets? Single-end data?	☆ Favorite	& Versions	<ul><li>✓ Options</li></ul>	search datasets Trinity example 2 shown, 2 deleted	00
	Are you po Yes Paired o Paired-e	<b>Soling sequence datasets?</b> Single-end data?			•	Trinity example 2 shown, 2 deleted	
	Yes Paired o Paired-e	• Single-end data? nd			•	2 shown, 2 deleted	
	Paired o Paired-e	r <b>Single-end data?</b> nd					
	Paired-e	nd				40.34 KB	
	Loff/Eo				-		
		numual saugural unerale				2: reads.right.fq	• # ×
	Certy FO	No facto factoranger er factoranger	az datacat availal		^	format: <b>fastq</b> , database	s <b>?</b>
		No fasta, fastqsanger or fastqsanger.	.gz dataset avalla	ble.		uploaded fastosanger	file
Filter and Sort						Lat 2	
oin, Subtract and C	ъУ_тр папссссанасу у	λααλαπλάαληση	$C = \sqrt{2} \sqrt{2} \sqrt{2}$	ᡣ᠋᠉ᠳᢆ	ͲϹϪϪϹͲϹϪϹ	マカC町町 0204:1:100:10494	4:3070/2
	IIIGGGGIICAA	AGCAGIAICGAICAAAIA	GIAAAIC	CALLIGI	ICAACICAC		TATTOSTTCAGGATGGAAG
Convert Formats							
ASTA/FASTQ	* ( ( ( ( * * * + ) )	888++) (888) .1***-	+*''))*	*55CCF>	>>>>>cccc	CCCC6	13422/2
ASTQ Quality Cont 5			, ,				>
GAM/BAM @SE	EQ ID2					eft.fq	⊕ # ×
GAT	TTTGGGGTTCAA	AGCAGTATCGATCAAATA	GTAAATC	CATTTGT	TCAACTCAC	AGTT nces	. 7
ICE/BCE						erel, database	
https://tratni		piect org/training-m	aterial/to	opics/in	troduction	n/tutorials/ga	lax



# 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.



# 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 ?





## Part 2:

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


## Part 2:

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



## Part 2:

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



## Part 2:

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

Tools ర		History	2 + II <b>Q</b>
search tools	Workflow: Find exons with the highest number of features	search datasets	00
🏦 Upload Data	Exons	Galaxy initiation -	
Get Data	▲ D D 1: exons.bed	3 shown	
Send Data	Features	10.72 MB	
Collection Operations	C C 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	⊛ # ×
loin, 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 X		
FASTA/FASTQ	Expand to full workflow form.		
FASTQ Quality Control			
SAM/BAM			
BED			
/CF/BCF			
Nanopore			
COMMON GENOMICS TOOLS			
Operate on Genomic Intervals			
etch Alignments/Sequences			



### Select multiple datasets as input

Tools		History	S+0 🜣
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.		
SENERAL TEXT TOOLS	Inputs	12: Top 5 exon IDs	⊙ # ×
ext Manipulation	► Steps	7: Top 5 exon IDs	⊛ # ×
ilter and Sort	View Report 2 🖶	3: repeats.bed	• / ×
in. Subtract and Group	7 of 7 steps successfully scheduled. 5 of 5 jobs complete.	2: sums had	
ENOMIC FILE MANIPULATION	Download BioCompute Object	2. snps.beu	
onvert Formats	► Steps	1: exons.bed	• / >
ASTA/FASTQ			
ASTQ Quality Control			
AM/BAM			
ED			
CF/BCF			
anopore			
DMMON GENOMICS TOOLS			
perate on Genomic Intervals			
etch Alignments/Sequences			
ENOMICS ANALYSIS	~		11
		III.	



- 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



### Create a collection from datasets in your history



120



### Renaming



Α.

в.

с.

D.



## Tagging



122



### Use a collection as input

n Galaxy France	倄 Workflow Visualize - Shared Data - Help - User - 💼 🏢	Using	j 3%
Tools	NCBI BLAST+ blastx Search protein database with translated nucleotide query sequence(s) (Galaxy Version 0.3.3)	History 2 + C search datasets	
Upload Data ③ Show Sections	Nucleotide query sequence(s)         Image: Constraint of the sequence of the sequenc	blast hands-on 2022 8 shown, 10 hidden 66.36 MB	» »
JBrowse - Data Directory to Standalone upgrades the bare data directory to a full JBrowse instance Diamond makedb Build database from	selection. (-query) Subject database/sequences	13: NCBI BLAST+ blastn across c ollection 12 a list with 5 items	×
a FASTA file <b>NCBI BLAST+ tblastx</b> Search translated nucleotide database with translated nucleotide query sequence(s)	Protein BLAST database           D         D         B: protein BLAST database from data 2         Image: Comparison of the second	12: queries         a list with 5 items         11: blastx query5.fa vs 'pr ③ a	×
NCBI BLAST+ blastx Search protein database with translated nucleotide query sequence(s) JBrowse genome browser	Query genetic code       1. Standard	otein BLAST database fro m data 2' 10: megablast query5.fa vs 'nucleotide BLAST data base from data 1'	/ X
BLAST XML to tabular Convert BLAST XML output to tabular Diamond alignment tool for short sequences against a protein database	Type of BLAST         Ø blastx - Traditional BLASTX to compare translated nucleotide query to protein database         Ø blastx-fast - Use longer words for seeding, faster but less accurate	9: nucleotide BLAST data ③ a base from data 1	/ ×
WORKFLOWS All workflows	(-task) Set expectation value cutoff	e from data 2 2: Drosophila_melanogast	<i>J</i> ×
	0.001 (-evalue) Output format	er.BDGP6.22.pep.all.fa.gz 1: Drosophila_melanogast er.BDGP6.22.cds.all.fa.gz	/ ×
<	Tabular (extended 25 columns)	<b>•</b> III 1	123



### Use a collection as input

Fools		History	3 + 🗆 🌣
search tools	Executed NCBI BLAST+ blastx and successfully added 5 jobs to the queue.	search datasets	00
	The tool uses 2 inputs:	blast bands on 2022	
🏦 Upload Data	• 12: queries	9 shown 10 hidden	
Get Data	* 8: protein BLAST database from data 2	Shown, to hidden	
Send Data	It produces 5 outputs:	00.30 MB	
	• 20: blastx query5.fa vs 'protein BLAST database from data 2'	19: NCBI BLAST+ blasty ac	cros X
Collection Operations	• 21: blastx query4.fa vs 'protein BLAST database from data 2'	s collection 12	
GENERAL TEXT TOOLS	<ul> <li>22: blastx query3.fa vs 'protein BLAST database from data 2'</li> <li>23: blastx query2 fa vs 'protein BLAST database from data 2'</li> </ul>	a list with 5 items	
Fext Manipulation	<ul> <li>24: blastx query1.fa vs 'protein BLAST database from data 2'</li> </ul>	13: NCBI BLAST+ blastn ac	cros X
ilter and Sort	You can check the status of queued jobs and view the resulting data by refreshing the History	s collection 12	
oin, Subtract and Group	panel. When the job has been run the status will change from 'running' to 'finished' if	a list with 5 items	
GENOMIC FILE MANIPULATION	completed successfully or 'error' if problems were encountered.	12: queries	×
Convert Formats			
FASTA/FASTO	Tool recommendation	11: blastx query5.fa vs 'protein BLAST databas	● / ×
EASTO Quality Control	You have used nchi blasty wrapper tool. For further analysis, you could the using the	e from data 2'	
FASTQ Quality Control	following/recommended tools. The recommended tools are shown in the decreasing order of	10: megablast query5 f	@ # ¥
SAM/BAM	their scores predicted using machine learning analysis on workflows. Therefore, tools at the top	a vs 'nucleotide BLAST	
BED	following/recommended tools to open its definition.	database from data 1'	
/CF/BCF		9: nucleotide BLAST da	⊙ / ×
Vanopore		tabase from data 1	
COMMON GENOMICS TOOLS		8: protein BLAST datab	• # ×
Operate on Genomic Intervals		ase from data 2	
Fetch Alianments/Sequences		2: Drosophila_melanog	⊙ # ×
JENOMICS ANALYSIS		aster.BDGP6.22.pep.all.	12
action of an action	V INCELEDAST HIBSTX	14.92	124



### 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			
GENERAL TEXT TOOLS	/tutorials/collections/tutorial.html#collection-operation	<u>าร</u>	
Text Manipulation	For any questions of support, community,cluster,mance biomormatique.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 'pr	· • / ×
GENOMIC FILE MANIPULATION	the 21.05 user release notes.	otein BLAST database fro m data 2'	
Convert Formats	×	10. manaklast sugar 5 fa	0 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 noni 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	
Operate on Genomic Interva	ls 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 ?		125
1			\ \



### Use-case: RNA-seq analysis





ARMS Dataset

Station Biologique

Roscoff

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

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.

## 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

Using dataset collections

Authors: 🚇 Anton Nekrutenko ) (🐣 Add Contributions! )

#### Overview

• 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

⑦ Questions:

Getting data

collection Collection

operations

**Ouestions** 

Feedback

Frequently Asked

Citing this Tutorial

Creating a paired

dataset collection Processing data organized as a

## Galaxy Training!

Search Tutorials Extras

 $\odot$ 



# 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).

- 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	0 -	×		🎤 Datamash	Ċ -	×	🖋 Sort	₫ →	×	🎤 Select first	₫ →	×
output (input)			0	<ul> <li>Intersect inter</li> <li>File A to inter</li> <li>File B to inter</li> </ul>	vals rsect with rsect with <i>i</i>	B		Input tabular o	dataset on input ular)		Sort Query	out)	of	from	put)	
Features	Q	→	×	output (in	put, bed)		J						🖋 Co two Da	mpare 🗘 ·	× ×	
													Comp	oare st		
													v ou	t_file1 (input)	0	,

- A workflow is a sequence of tool operations and parameters
- A workflow is built to be replayed (more or less strict)



🗅 Exons	¢	>	×	🏓 bedtools	¢ →	×	🔑 Datamash	<b>C</b> -	→ ×		🔑 Sort	¢	<del>)</del>	κ	🎤 Select first	<b>C</b> -	➤
output (input)			0	Intersect interv File A to inter File B to inter	vals sect with B sect with A		<ul> <li>Input tabular o</li> <li>Datamash o</li> <li>dataset(s) (tab</li> </ul>	lataset n input ular)		S	Sort Query	)		of	● from ✓ out_file1 (input	ıt)	
Features output (input)	¢	÷	×	🗆 output (inp	out, bed)	•								م two Da	mpare <b>௴→</b> atasets	×	
													-0	Comp	bare		
													6	again	st		
														v out	t_file1 (input)		>

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



🗅 Exons	¢	<b>&gt;</b>	×	🔑 bedtools	_ ᠿ →	×		🔑 Datamash	Ċ,	→ ×	8	🔑 Sort	0 -	<b>&gt;</b>	¢	🎤 Select first	Ċ -	×
output (input)				Intersect interva File A to inters File B to inters	als ect with B ect with A	0		Datamash o dataset(s) (tabu	ataset n input ular)			Sort Query			of	rom ✓ out_file1 (inp	ut)	
Features output (input)	¢	<b>→</b>	×	output (inpu	ut, bed)		9							ſ	م لwo D∂	mpare ① → atasets	×	
														0	Comp again	oare st		
															🗹 ou	t_file1 (input)	(	\$

## Why would you want to create workflows?

- 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



## Our workflow with Galaxy





## From history

Tools     \$\$\$       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 those that corporated into a workflow will be shown in gray.	t	History Actions Copy	ory options
🏦 Upload Data	Workflow name				Share or Publish	
Get Data	Workflow constructed from history 'tuto-galaxy-intro-           Create Workflow         Check all         Uncheck all	101'			Show Structure Extract Workflow	•
Collection Operations	Tool		History items created		Set Permissions	) / X
GENERAL TEXT TOOLS				1 _	Resume Paused Jobs	
Text Manipulation			Treat as input dataset		Dataset Actions	) / X
Filter and Sort			Exons.bed		Copy Datasets	) / X
Join, Subtract and Group	Data Fetch				Collapse Expanded Datasets	) / X
GENOMIC FILE MANIPULATION	This tool cumot be used in worklows		2 SNPs.bed		Unhide Hidden Datasets	X
Convert Formats			Treat as input dataset		Delete Hidden Datasets	b .
FASTA/FASTQ			SNPs.bed		Purge Deleted Datasets	) / X
FASTQ Quality Control	FA/FASTQ     FQ Quality Control     bedrools Intersect intervals		3 bedtools Intersect intervals on data 2 and		Downloads	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	
Nanonoro					Use Beta History Panel	
COMMON GENOMICS TOOLS	Sort		5 Sort on data 4		135	>



## From history





## Workflow manager

💶 Galaxy France	🕋 Workflow Visualize 🕶	Shared Data 🔻 Help 🔻 User 🔻 📂 🏬	Using 3%
Tools 🖒	Search Workflows		History 🖯 🕂 🗖 🌣
search tools	Search WORNOWS		search datasets 2 😢
1 Upload Data	Name 🔶 Tags	🗘 Updated 🖨 Sharing 🖨 Bookmarked 荣	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	Galaxy initiation training	a year ago	6: Select first o 🕑 🧳 🗙 n data 5
GENOMIC FILE MANIPULATION	<ul> <li>2020 - tools hands-on TEST2</li> <li>▼ Galaxy initiation training</li> </ul>	a vear ago	5: Sort on data ( ) X 4
FASTA/FASTQ	2020 - tools hands-on TEST		4: Datamash on 💿 🖋 🗙 data 3
FASTQ Quality Control			3: bedtools Inte



## Workflow manager

iools ☆	Search Workflows			+ Create	History	8+ <b>0</b> 4
search tools					search datasets	; <b>8</b> 8
🍰 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	
end 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	💼 Delete	\$12			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: set or release a parameter





#### Edit a workflow: set or release a parameter

📮 Galaxy France 🔗	Workflow Visualize 🕶 Shared Data 🖛 Help 👻 User 🖛 📰 🇱	Using 3%
Find exons with the highest number of features		0 🖻 🗹 🌣 🕨 🔒
$\not$ bedtools $\square \rightarrow \times$ $\not$ Datamash $\square$	$\rightarrow$ X Sort $\square$ $\rightarrow$ X Select first $\square$ $\rightarrow$ X	Select first lines from a dataset (Galaxy Version 1.0.1)
Eile A to intersect with B	t Sort Query from	Label
File B to intersect with A Datamash on inp	ut 💿 outfile (input) 🔽 out_file1 (input)	
output (input, bed)	✓ Compare	Add a step label. Step Annotation
ge-	Compare against	
	☑ out_file1 (input)	
		Add an annotation or notes to this step. Annotations are available when a workflow is viewed.
- 100% +		L ↔ Select first Set at Runtime
>		>



#### Edit a workflow: rename the outputs





#### Save





💶 Galaxy France	👚 Workflow Visualize <del>-</del> Sh	ared Data 🔻 Help 🔻 User 🔻 💼	Using 3%
Tools	Search Workflows	+ Create	History 2+ 🗆 🌣
search tools	Name 🔶 Tags	🗘 Updated 🖨 Sharing 🖨 Bookmarked 🖨	tuto-galaxy-
Get Data	▼ Find exons with the highest number of features	7 days	7 shown           8.53 MB
Collection Operations	<ul> <li>Galaxy initiation training</li> <li>2020 - tools hands-on 1234</li> </ul>	a year ago	► 7: Compare two ④ ♂ ★
Text Manipulation	Galaxy initiation training 2020 - tools hands-on TEST3	a year ago	Datasets on dat a 6 and data 1
Filter and Sort Join, Subtract and Group	<ul> <li>→ Galaxy initiation training</li> <li>2020 - tools hands-on TEST2</li> </ul>	a year ago	6: Select first o () / × n data 5
GENOMIC FILE MANIPULATION Convert Formats	<ul> <li>Galaxy initiation training</li> <li>2020 - tools hands-on TEST</li> </ul>	a year ago	4 4: Datamash on ④ 🖉 🗙
FASTA/FASTQ FASTQ Quality Control			data 3 <u>3: bedtools Inte</u> ③ A ×



<b>T</b> Galaxy France		🕋 Workflow Visualize 🗝 Shared Data 🖣 Help 👻 User 🍷 🗖	> #					Using 3%
Tools	☆						History	€+□¢
search tools	8	Workflow: Find exons with the highest number of features		Run V	Workflow		search datasets	88
🏦 Upload Data		1: Exons			OKIOW		Galaxy initiation - workflow	
Get Data	^	1: Exons.bed	•	£			2 shown	
Send Data							3.71 MB	
Collection Operations		2: Features			۲		2. separts had	
GENERAL TEXT TOOLS		2: repeats.bed	•	t	B		2: repeats.bed	• / *
Text Manipulation							1: Exons.bed	④ ∦ ×
Filter and Sort		3: bedtools Intersect intervals (Galaxy Version 2.30.0)			R			
Join, Subtract and Group								
GENOMIC FILE MANIPULATION					Ø			
Convert Formats		<u>5: Sort (Galaxy Version 1.1.1)</u>			Ø			
FASTA/FASTQ								
FASTQ Quality Control		<u>6: Select first (Galaxy Version 1.0.1)</u>			۲			
SAM/BAM		Select first						
BED		20						
VCF/BCF		lines						
Nanopore		from						
COMMON GENOMICS TOOLS	~	false			~	100		151



🗧 Galaxy France		🕋 Workflow Visualize 🕶 Shared Data 👻 Help 👻 User 🏲 📻						Using 3%
Tools	☆						History	S+0\$
search tools	8	Workflow: Find exons with the highest number of features		Dup V	Vorkflow		search datasets	00
🏦 Upload Data		<u>1: Exons</u>					Galaxy initiation - workflow	
Get Data	^	1: Exons.bed	•	£			2 shown	
Send Data							3.71 MB	
Collection Operations		1 <u>2: Features</u>			۲		2	
GENERAL TEXT TOOLS		2: repeats.bed	•	£	⊳		2: repeats.bed	• # ×
Text Manipulation							1: Exons.bed	• / ×
Filter and Sort		3: bedtools Intersect intervals (Galaxy Version 2.30.0)			Ø			
Join, Subtract and Group								
GENOMIC FILE MANIPULATION		✓ <u>4: Datamash (Galaxy Version 1.1.0)</u>			Ø			
Convert Formats		<u>5: Sort (Galaxy Version 1.1.1)</u>			Ø			
FASTA/FASTQ						1		
FASTQ Quality Control		<u>6: Select first (Galaxy Version 1.0.1)</u>			۲			
SAM/BAM		Select first						
BED		20						
VCF/BCF		lines						
Nanonore		from						
COMMON GENOMICS TOOLS	~	<b>T</b> Dataset has a header			~	1	11	152



<b>=</b> Galaxy France		倄 Workflow Visualize 🕶 Shared Data 🖜 Help 🖜 User 🖛 🞓					Using 3%
Tools	☆					History	S+0\$
search tools	8	Workflow: Find exons with the highest number of features		Run V	Workflow	search datasets	88
🏦 Upload Data		1: Exons		Kun v		Galaxy initiation - workflow	
Get Data	^	L   L     L   Exons.bed	•	£		2 shown	
Send Data						3.71 MB	
Collection Operations		2: Features			۲	2 1 1	
GENERAL TEXT TOOLS		1   1     2: repeats.bed	•	1		2: repeats.bed	• / ×
Text Manipulation						1: Exons.bed	• / ×
Filter and Sort		3: bedtools Intersect intervals (Galaxy Version 2.30.0)			Ø		
Join, Subtract and Group							
GENOMIC FILE MANIPULATION		<u>4: Datamash (Galaxy Version 1.1.0)</u>			R		
Convert Formats		✤ <u>5: Sort (Galaxy Version 1.1.1)</u>			R		
FASTA/FASTQ							
FASTQ Quality Control		<u>6: Select first (Galaxy Version 1.0.1)</u>			۲		
SAM/BAM		Select first					
BED		20					
VCF/BCF		lines					
Nanopore		from					
COMMON GENOMICS TOOLS	~	false				101	153



📮 Galaxy France	Workflow Visualize 🕶 Shared Data 🕶 Help 🕶 User 🕶 📰		Using 3%
Tools		History	3+0\$
search tools	Successfully invoked workflow Find exons with the highest number of features.	search datasets	00
1 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	۵ 🥒 🗙
Text Manipulation		6: Top 5 exon IDs	() () () () () () () () () () () () () (
Filter and Sort		5: Sort on data 4	() () () () () () () () () () () () () (
Join, Subtract and Group		• 4: Datamash on data 3	A 10 10 10 10 10 10 10 10 10 10 10 10 10
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			
		15	,4



<b>Galaxy France</b>	isualize 🗸 Shared Data 🗸 Help 🗸 User 🕇 💼		Using 3%
Tools		History	S+ 🗆 💠
search tools	Successfully invoked workflow Find exons with the highest number of features.	search datasets	88
1 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			
		100	155



Impossible (until now)

Workflow

Station Biologique Roscoff





### Last but not least! SHARE



## biologist $\leftrightarrow$ biologist

- Sharing histories or datasets
  - With or without linked workflow



### bioanalyst ↔ biologist

- Sharing workflows
  - Pre-configured parameters
  - With or without release parameters (set at runtime)
    - According to the user-end knowledge



## bioinformatician $\leftrightarrow$ bioinformatician

- Sharing tools ,scripts and wrappers
  - Toolshed



# ARTS AND Biologique Roscoff Share

#### History

📮 Galaxy France	👫 Workflow Visualize 🕶 Shared Data 🍷 Help 👻 User 🍷 💼		Using 3%
Tools	☆ Share or Publish History `Galaxy initiation - workflow`	History	<b>∂+⊡</b> ¢
search tools	<ul> <li>Make History accessible</li> <li>Make History publicly available in Published Histories</li> </ul>	search d Galaxy ir workflov	History Actions Copy Share or Publish
Get Data Send Data	<ul> <li>This History is currently accessible via link.</li> <li>Anyone can view and import this History by visiting the following URL:</li> <li> <i>O</i> url: https://usegalaxy.fr/u/lgueguen/h/galaxy-initiationworkflow     </li> </ul>	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. Share with a user	2: repeats.	<b>Dataset Actions</b> Copy Datasets
GENOMIC FILE MANIPULATION		1: Exons.b	Collapse Expanded Datasets Unhide Hidden Datasets
Convert Formats FASTA/FASTQ			Delete Hidden Datasets Purge Deleted Datasets
FASTO Ouality Control		Ш	Downloads



#### Workflow





#### Workflow: mode

<b>Galaxy France</b>	/isualize ▼ Shared Data ▼ Help ▼ User ▼ 🞓 🏢 🛛 🗍 Usin	ıg 3%
Go back to Workflows List Workflow ' Find exons with the highest numbe	r of features'	,
Share		
This workflow is currently restricted so that only you and the users listed below ca	n access it.	
Make Workflow Accessible via Link	Restricted community	
Generates a web link that you can share with other people so that they can view a	nd import the workflow.	
Make Workflow Accessible and Publish	→ All the Galaxy server users	
Makes the workflow accessible via link (see above) and publishes the workflow to	Galaxy's Published Workflows section, where it is publicly listed and searchable.	
You have not shared this workflow with any users yet.		
Share with a user	→ Designated community	
Export		
Download workflow as a file so that it can be saved or imported into another	Galaxy server.	
This workflow must be accessible. Please use the option above to "Make Workflow	v Accessible and Publish" before receiving a URL for importing to another Galaxy.	
Create image of workflow in SVG format		
Export to the www.myexperiment.org site.		
myExperiment username:	163	

V



#### • Get shared histories

🗧 Galaxy France	👚 Workflow Visualize ▾ Shared Data ▾ Help ▾	User 🕶 👔 🎆		Using 3%
Tools ☆	Histories shared with you by others	Logged in as lgueguen@sb-roscoff.fr	tory	S+0\$
search tools		Preferences	earch datasets	88
	Name Datasets Created Last Updated	Custom Builds	avy initiation -	
	No items	loqout	rkflow	
Get Data		Datasets	own, 3 hidden	
Send Data		Histories	МВ	
Collection Operations		Histories shared with me		
		<ul> <li>Account and saved data</li> </ul>	on 5 avons	

💶 Galaxy France	👚 Workflow Visualize 🔻	Shared Data 👻 Help 👻 🛛	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 Visua	alize Shared Data ▼ Help ▼	User 🕶 🌲 🏢			Using 3%
Tools     ☆ :=       search tools     Image: Control of the search tool of teol of teol	Search Workflows			+ Create	History search datasets	2+□¢ 00
Lupload Data	Name ▼ Find exons with the highest number of features	Tags	Updated Sharing 3 days ago	Bookmarked	Galaxy initiation - workflow 4 shown, 3 hidden	
Send Data Collection Operations	<ul> <li>Galaxy initiation training 2020 - tools hands-on 123</li> </ul>	34	a year ago		7: Top 5 evons	
GENERAL TEXT TOOLS Text Manipulation Filter and Sort	▼ COVID-19: variation analysis on ARTIC PE data v0 The workflow for Illumina-sequenced ARTIC data builds on the RNASeq workflow for paired-end data using the same steps for mapping and variant calling, but adds	.2 COVID-19 × ARTIC × covid19.galaxyproject.org emergen × •	6 months		6: Top 5 exon IDs 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 -	
Get Data	^	Name	Pages	Annotation		4 shown, 3 hidden	
Send Data		Bordeaux-Global-ESI-pos 🔻				5.97 MB	



#### • Import shared

🗧 Galaxy France	삼 Workflow	Visualize 🔻 Shared Data 🕇	Help 🔻 User 🔻 🔽	▶ Ⅲ		Using 3%
Rustenholz - DE Sangiovese			Listania		About this History	+
5.49 MB			HISTOLIE	es	Author	Import history
search datasets				00		
Dataset				Annotation	dc244f268e004c7d9977d	lea686a1cf4e
					<b>Related Histories</b>	
9: Sangiovese_3_htseq.txt	۲			· · · · · · · · · · · · · · · · · · ·	All published histories	
8: Sanaiovese 2 htsea.txt	۲				Published histories by dc244f268e004c7d9977d	lea686a1cf4e

🔁 Galaxy France		삼 Workflow Visualize 🛪 Shared	Data 🔻 Help 🔻 User 🔻 📂 🏭		Using 3%
Tools search tools	☆ 8	Published Workflows	Workflows	History     search datasets	\$+⊡¢ 00
1 Upload Data		Advanced Search	Annotation	Galaxy initiation - workflow	
Get Data		Rordonux Clobal	Amotation	4 shown, 3 hidden	
Send Data		Run		5.97 MB	
Collection Operations		C_Elegans_Publica Import		7: Top 5 exons	● # ×
GENERAL TEXT TOOLS		W 10 C			166





Level 5

. Share tools and descriptions in the ToolShed

Level 4

- . Launch tools autonomously
- . Use advanced parameters
- Provide workflow for colleagues Level 1-3

Level 3

- Launch tools autonomously
- . Use workflow more or less preset

Level 2

• Use preset workflow

Level 1

• Share his data to colleagues Level 2-5



## Hands-on WORKFLOW AND SHARE





Station Biologique Roscoff

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



#### Introduction

Introduction Setting the stage: Exons and SNPs

Analysis Galaxy

management

Frequently Asked

Citing this Tutorial

Conclusion

Questions

Feedback

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.

#### Besults 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.

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!

#### 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:**





- 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)	Minneapolis, USA	July 2022
4 Wm Nutifierineshients	Workflow4Experimenters	Metabolomics	Europe	2022 (closed), 2023



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