Travaux pratiques – Galaxy* (partie 3)

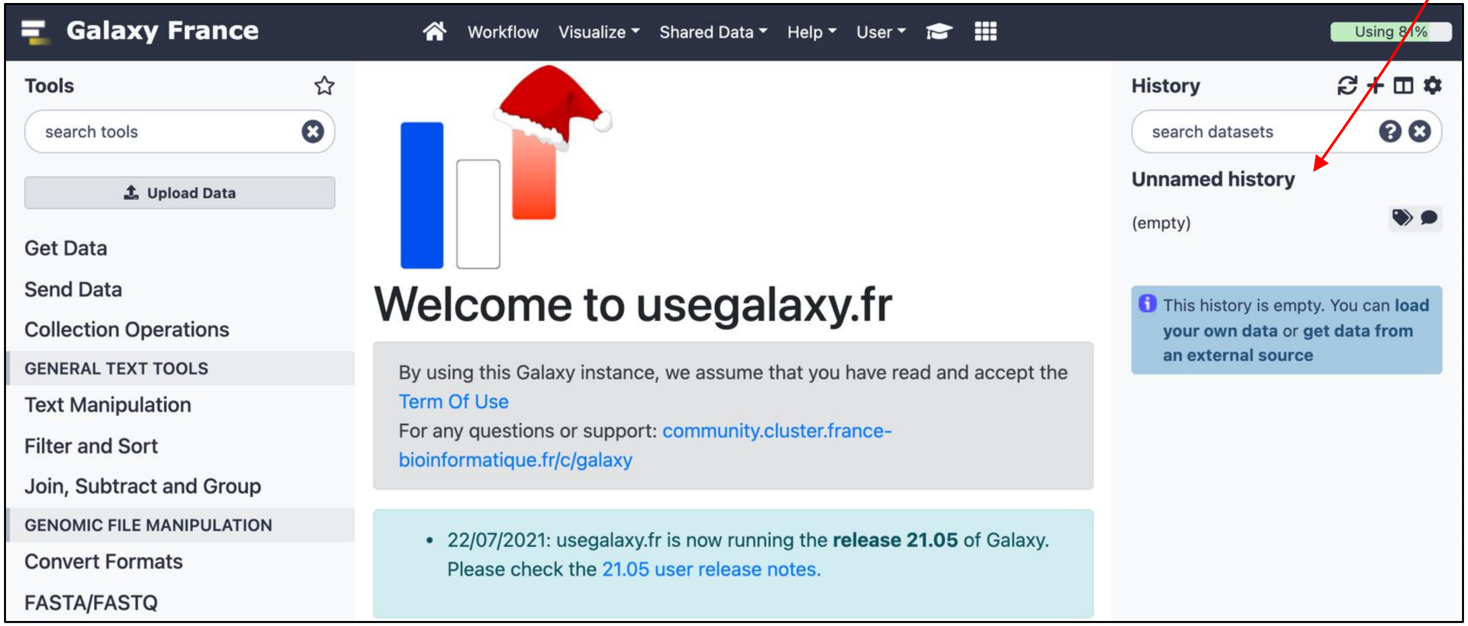
Gaëlle Lelandais et Fabrice Confalonieri

gaelle.lelandais@universite-paris-saclay.fr ; fabrice.confalonieri@universite-paris-saclay.fr

* Les captures d’écran ont été réalisées en 2022, il est possible que l’interface aujourd’hui soit légèrement différente.

Etape 1 : Connexion à Galaxy

- Se connecter à l’instance Galaxy « France » : https://usegalaxy.fr/
- Créer un nouvel historique de travail



Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and accept the Term Of Use
For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy

- 22/07/2021: usegalaxy.fr is now running the **release 21.05** of Galaxy. Please check the 21.05 user release notes.

This history is empty. You can **load your own data** or **get data from an external source**

Données utilisées pour le TP

cells

MDPI

Article

Characterization of the Radiation Desiccation Response Regulon of the Radioresistant Bacterium *Deinococcus radiodurans* by Integrative Genomic Analyses

Nicolas Eugénie, Yvan Zivanovic, Gaelle Lelandais, Geneviève Coste, Claire Bouthier de la Tour, Esma Bentchikou, Pascale Servant † and Fabrice Confalonieri *,†

Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France; nicolas.eugenie@i2bc.paris-saclay.fr (N.E.); yvan.zivanovic@i2bc.paris-saclay.fr (Y.Z.); gaelle.lelandais@i2bc.paris-saclay.fr (G.L.); genevieve.coste@i2bc.paris-saclay.fr (G.C.); claire.bouthier@i2bc.paris-saclay.fr (C.B.d.l.T.); esma.bentchikou@i2bc.paris-saclay.fr (E.B.); pascale.servant@i2bc.paris-saclay.fr (P.S.)

* Correspondence: fabrice.confalonieri@i2bc.paris-saclay.fr

† These authors contributed equally to this work.

Abstract: Numerous genes are overexpressed in the radioresistant bacterium *Deinococcus radiodurans* after exposure to radiation or prolonged desiccation. It was shown that the DdrO and IrrE proteins play a major role in regulating the expression of approximately twenty genes. The transcriptional repressor DdrO blocks the expression of these genes under normal growth conditions. After exposure to genotoxic agents, the IrrE metalloprotease cleaves DdrO and relieves gene repression. At present, many questions remain, such as the number of genes regulated by DdrO. Here, we present the first ChIP-seq analysis performed at the genome level in *Deinococcus* species coupled with RNA-seq, which was achieved in the presence or not of DdrO. We also resequenced our laboratory stock strain of *D. radiodurans* R1 ATCC 13939 to obtain an accurate reference for read alignments and gene expression quantifications. We highlighted genes that are directly under the control of this transcriptional repressor and showed that the DdrO regulon in *D. radiodurans* includes numerous other genes than those previously described, including DNA and RNA metabolism proteins. These results thus pave the way to better understand the radioresistance pathways encoded by this bacterium and to compare the stress-induced responses mediated by this pair of proteins in diverse bacteria.

Keywords: radioresistance/desiccation; transcriptional regulator; *Deinococcus radiodurans*; ChIP-seq; RNA-seq; bioinformatic analyses

Citation: Eugénie, N.; Zivanovic, Y.; Lelandais, G.; Coste, G.; Bouthier de la Tour, C.; Bentchikou, E.; Servant, P.; Confalonieri, F. Characterization of the Radiation Desiccation Response Regulon of the Radioresistant Bacterium *Deinococcus radiodurans* by Integrative Genomic Analyses. *Cells* **2021**, *10*, 2536. https://doi.org/10.3390/cells10102536

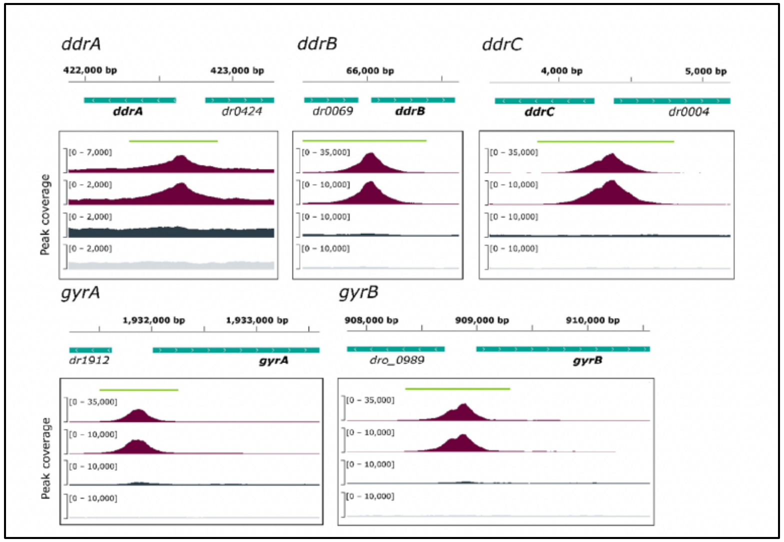
Academic Editors: Bernard S. Lopez and Ivan Matic

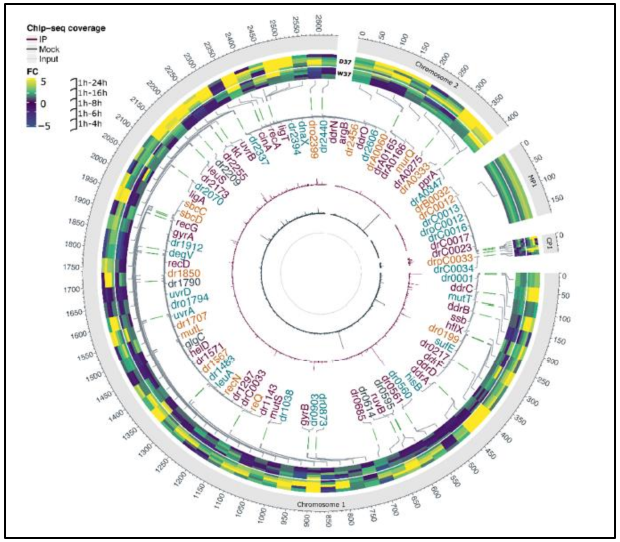
Received: 19 July 2021
Accepted: 10 September 2021
Published: 25 September 2021

1. Introduction

Deinococcus radiodurans is one of the most resistant bacteria to genotoxic agent exposure and desiccation isolated to date [1–4]. Unlike radiosensitive organisms, once exposed to huge γ-ray doses, or after prolonged desiccation, *D. radiodurans* is able to reconstruct an intact genome in a few hours from several hundred DNA fragments [5]. Many factors contribute to the radioresistance of *D. radiodurans*, including efficient DNA repair mechanisms [5–8], a condensed nucleoid limiting the dispersion of genome fragments after irradiation [9,10], and the protection of proteins against oxidative damage [11]. Thus, the exceptional ability of this bacterium to overcome severe DNA damaging conditions is described as a combination of active and passive mechanisms acting in synergy within the cell, enabling survival following these stresses.

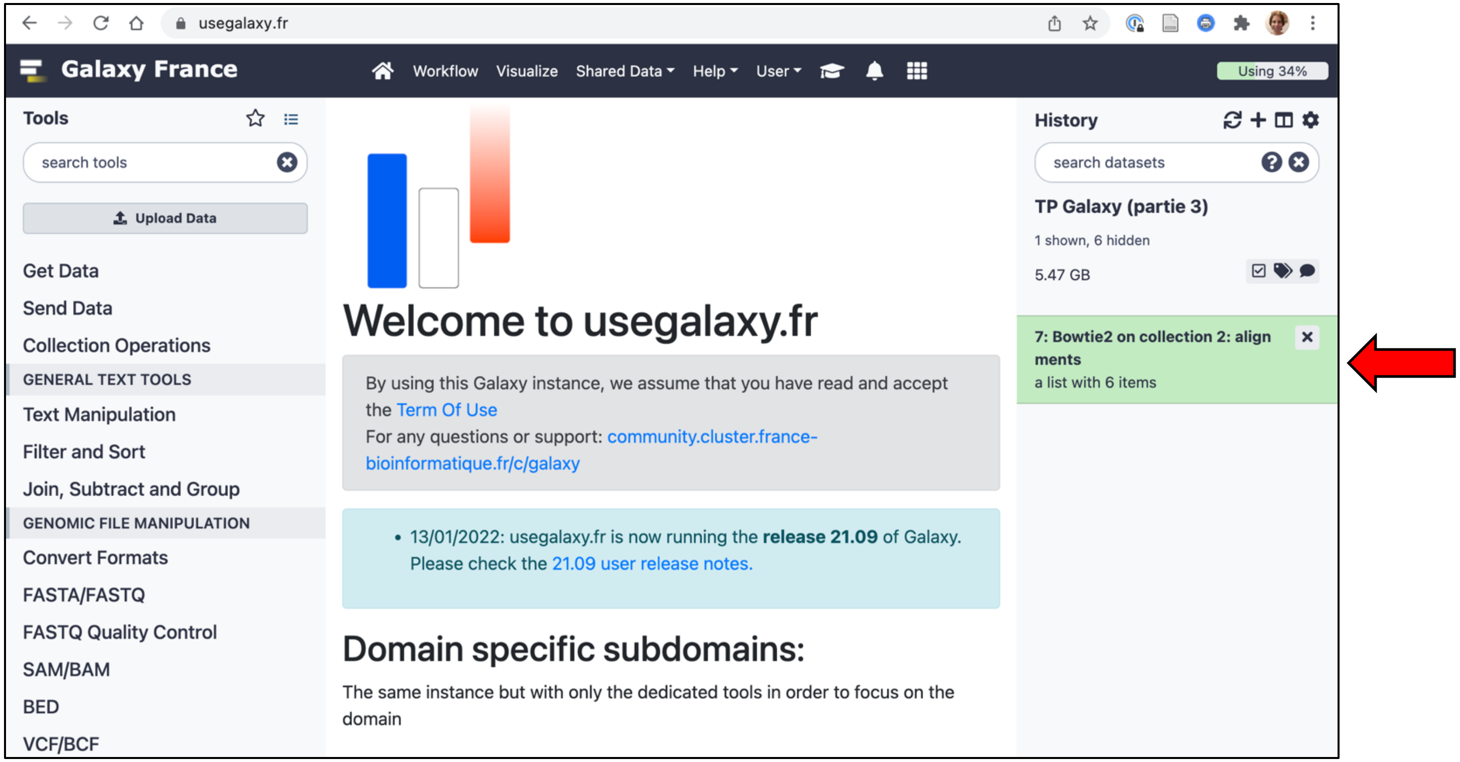
The exposure of *D. radiodurans* to γ-rays, or its recovery from desiccation, results in a rapid upregulation of the expression of numerous genes [12,13], even if constitutively expressed genes are also involved in the mechanisms of radioresistance. In many bacterial species, expression of DNA repair genes is under the control of LexA, the repressor of the well-known SOS response (for review [14]). *D. radiodurans* encodes two LexA homologs





Etape 2 : Importer les résultats des alignements (TP2)

- Récupérer dans votre nouvel historique de travail la collection des fichiers BAM



Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and accept the Term Of Use
For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy

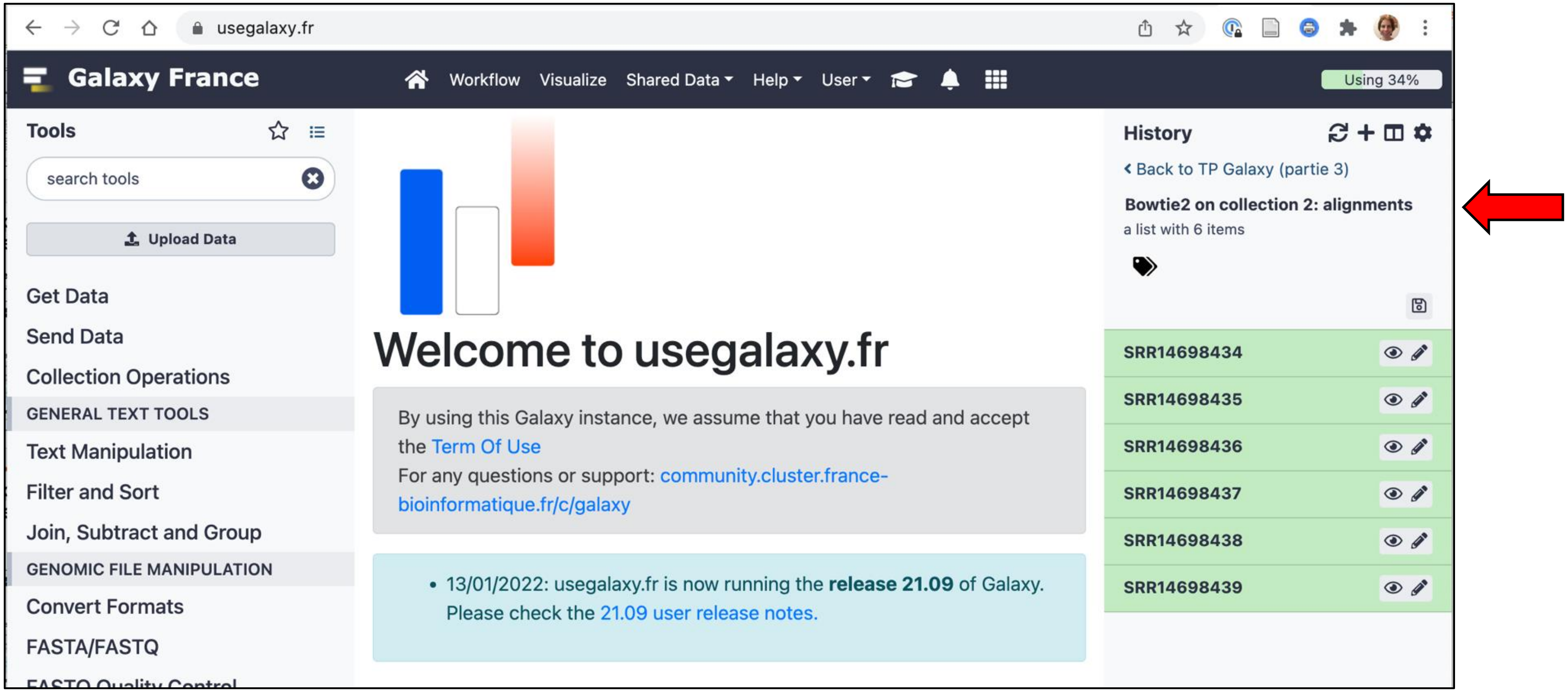
- 13/01/2022: usegalaxy.fr is now running the **release 21.09** of Galaxy. Please check the 21.09 user release notes.

Domain specific subdomains:

The same instance but with only the dedicated tools in order to focus on the domain

Etape 2 : Importer les résultats des alignements (TP2)

- Vérifier le contenu de la collection :



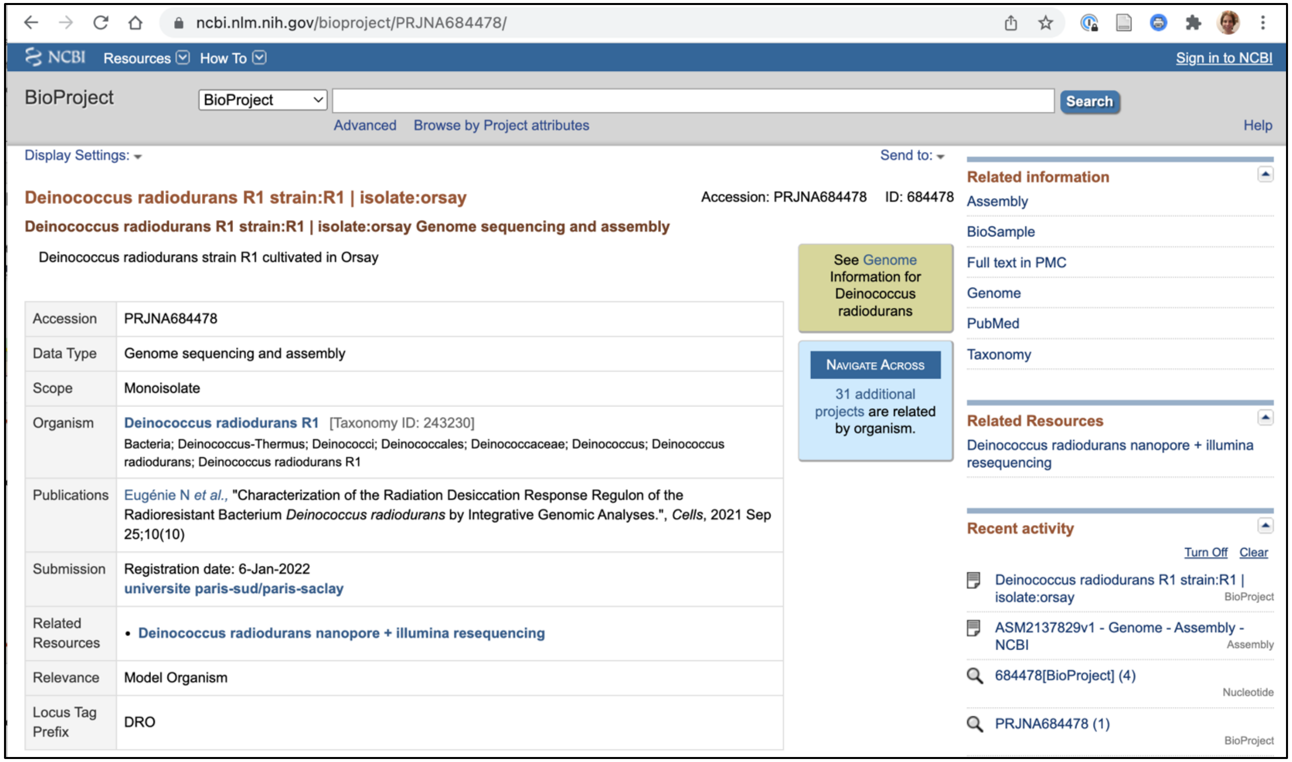
Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and accept the Term Of Use
For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy

- 13/01/2022: usegalaxy.fr is now running the **release 21.09** of Galaxy. Please check the 21.09 user release notes.

Etape 3 : Télécharger le fichier d’annotation du génome (GFF)

- Accéder la page NCBI du génome : https://www.ncbi.nlm.nih.gov/bioproject/PRJNA684478/



Deinococcus radiodurans R1 strain:R1 | isolate:orsay

Accession: PRJNA684478 ID: 684478

Deinococcus radiodurans R1 strain:R1 | isolate:orsay Genome sequencing and assembly

Deinococcus radiodurans strain R1 cultivated in Orsay

Accession	PRJNA684478
Data Type	Genome sequencing and assembly
Scope	Monoisolate
Organism	**Deinococcus radiodurans R1** [Taxonomy ID: 243230] Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales; Deinococcaceae; Deinococcus; Deinococcus radiodurans; Deinococcus radiodurans R1
Publications	Eugénie N *et al.*, "Characterization of the Radiation Desiccation Response Regulon of the Radioresistant Bacterium *Deinococcus radiodurans* by Integrative Genomic Analyses.", *Cells*, 2021 Sep 25;10(10)
Submission	Registration date: 6-Jan-2022 **universite paris-sud/paris-saclay**
Related Resources	• **Deinococcus radiodurans nanopore + illumina resequencing**
Relevance	Model Organism
Locus Tag Prefix	DRO

See Genome Information for Deinococcus radiodurans

NAVIGATE ACROSS

31 additional projects are related by organism.

Related information

- Assembly
- BioSample
- Full text in PMC
- Genome
- PubMed
- Taxonomy

Related Resources

Deinococcus radiodurans nanopore + illumina resequencing

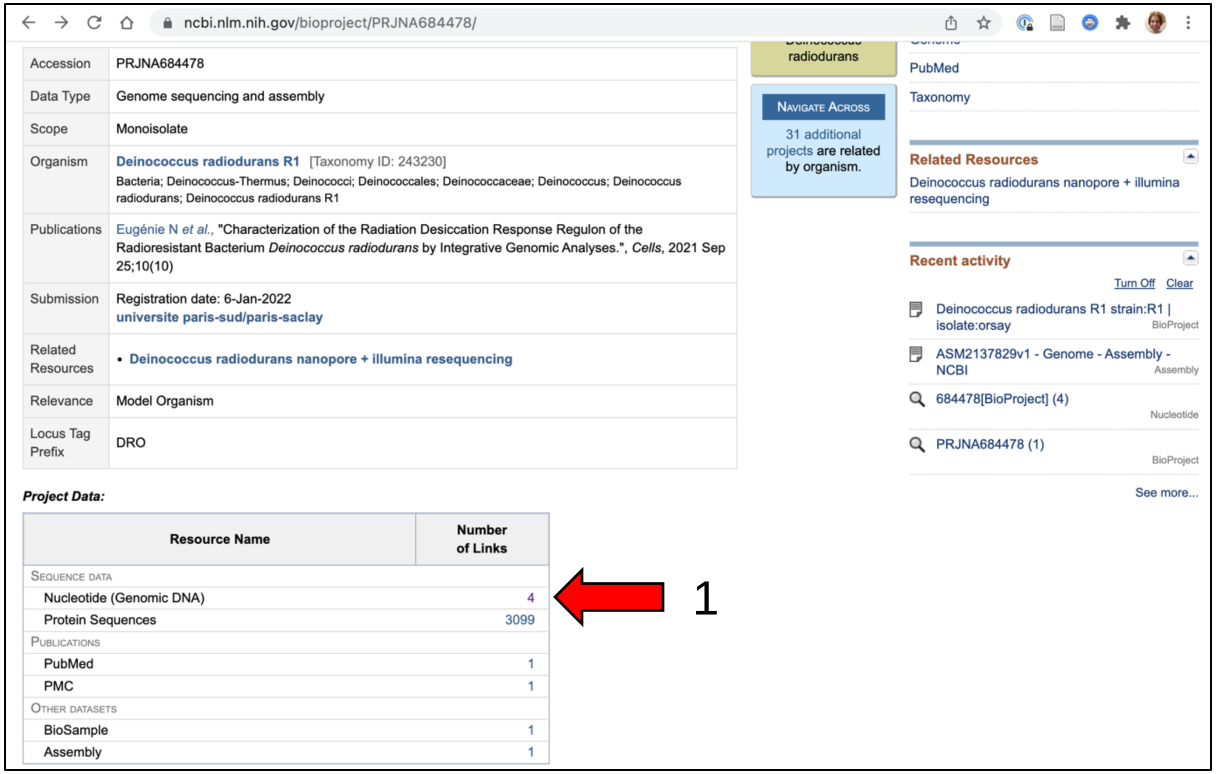
Recent activity

Turn Off Clear

- Deinococcus radiodurans R1 strain:R1 | isolate:orsay — BioProject
- ASM2137829v1 - Genome - Assembly - NCBI — Assembly
- 684478[BioProject] (4) — Nucleotide
- PRJNA684478 (1) — BioProject

Etape 3 : Télécharger le fichier d’annotation du génome (GFF)

- Accéder à la même page que celle consultée pour le fichier génome (FASTA) :



ncbi.nlm.nih.gov/bioproject/PRJNA684478/

Accession	PRJNA684478
Data Type	Genome sequencing and assembly
Scope	Monoisolate
Organism	**Deinococcus radiodurans R1** [Taxonomy ID: 243230] Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales; Deinococcaceae; Deinococcus; Deinococcus radiodurans; Deinococcus radiodurans R1
Publications	Eugénie N *et al.*, "Characterization of the Radiation Desiccation Response Regulon of the Radioresistant Bacterium *Deinococcus radiodurans* by Integrative Genomic Analyses.", *Cells*, 2021 Sep 25;10(10)
Submission	Registration date: 6-Jan-2022 **universite paris-sud/paris-saclay**
Related Resources	• **Deinococcus radiodurans nanopore + illumina resequencing**
Relevance	Model Organism
Locus Tag Prefix	DRO

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic DNA)	4
Protein Sequences	3099
PUBLICATIONS	
PubMed	1
PMC	1
OTHER DATASETS	
BioSample	1
Assembly	1

radiodurans

NAVIGATE ACROSS

31 additional projects are related by organism.

PubMed

Taxonomy

Related Resources

Deinococcus radiodurans nanopore + illumina resequencing

Recent activity

Turn Off Clear

Deinococcus radiodurans R1 strain:R1 | isolate:orsay — BioProject

ASM2137829v1 - Genome - Assembly - NCBI — Assembly

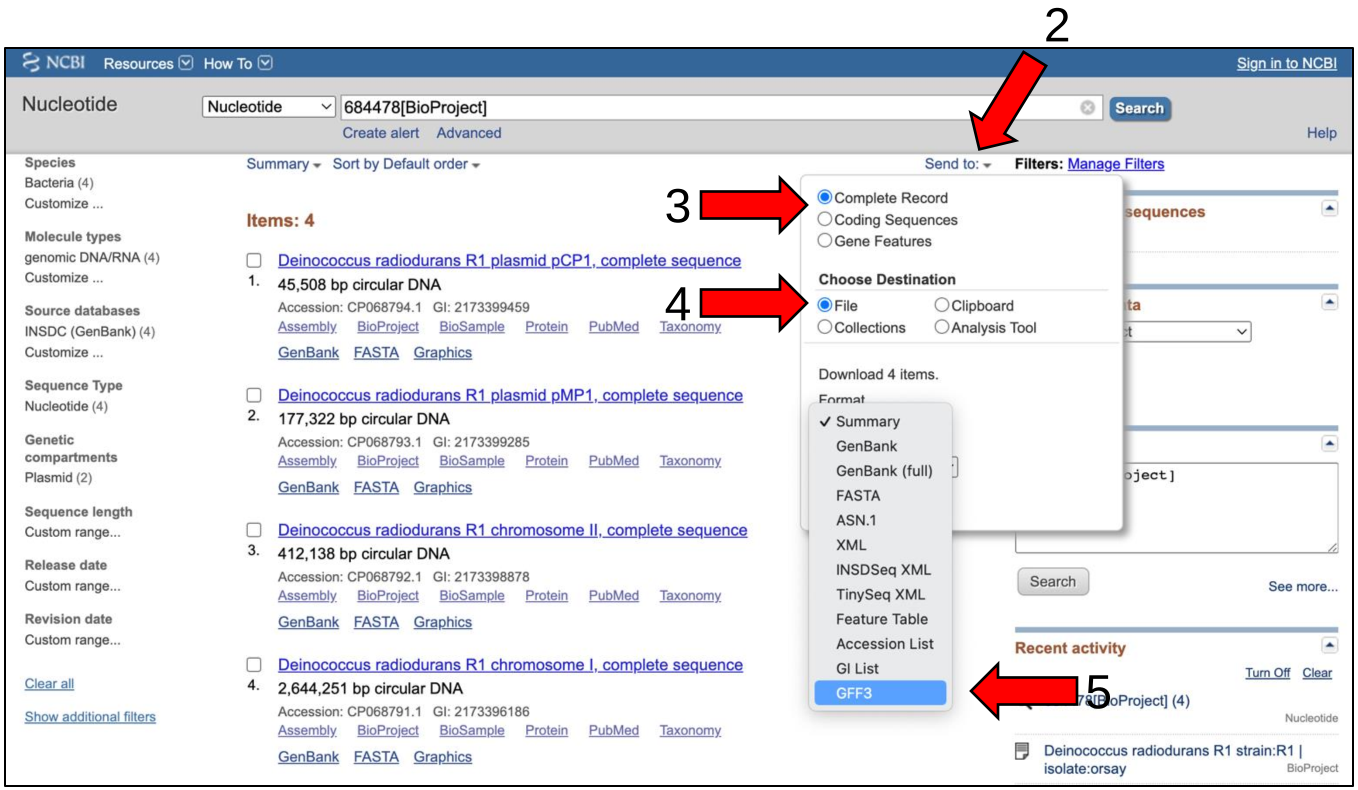
684478[BioProject] (4) — Nucleotide

PRJNA684478 (1) — BioProject

See more...

Etape 3 : Télécharger le fichier d’annotation du génome (GFF)

- Sauvegarder le fichier « GFF3 » sur votre ordinateur de travail :



Nucleotide: 684478[BioProject]

Species: Bacteria (4), Customize ...
Molecule types: genomic DNA/RNA (4), Customize ...
Source databases: INSDC (GenBank) (4), Customize ...
Sequence Type: Nucleotide (4)
Genetic compartments: Plasmid (2)
Sequence length: Custom range...
Release date: Custom range...
Revision date: Custom range...
Clear all
Show additional filters

Items: 4

1. ☐ Deinococcus radiodurans R1 plasmid pCP1, complete sequence
45,508 bp circular DNA
Accession: CP068794.1 GI: 2173399459
Assembly BioProject BioSample Protein PubMed Taxonomy
GenBank FASTA Graphics

2. ☐ Deinococcus radiodurans R1 plasmid pMP1, complete sequence
177,322 bp circular DNA
Accession: CP068793.1 GI: 2173399285
Assembly BioProject BioSample Protein PubMed Taxonomy
GenBank FASTA Graphics

3. ☐ Deinococcus radiodurans R1 chromosome II, complete sequence
412,138 bp circular DNA
Accession: CP068792.1 GI: 2173398878
Assembly BioProject BioSample Protein PubMed Taxonomy
GenBank FASTA Graphics

4. ☐ Deinococcus radiodurans R1 chromosome I, complete sequence
2,644,251 bp circular DNA
Accession: CP068791.1 GI: 2173396186
Assembly BioProject BioSample Protein PubMed Taxonomy
GenBank FASTA Graphics

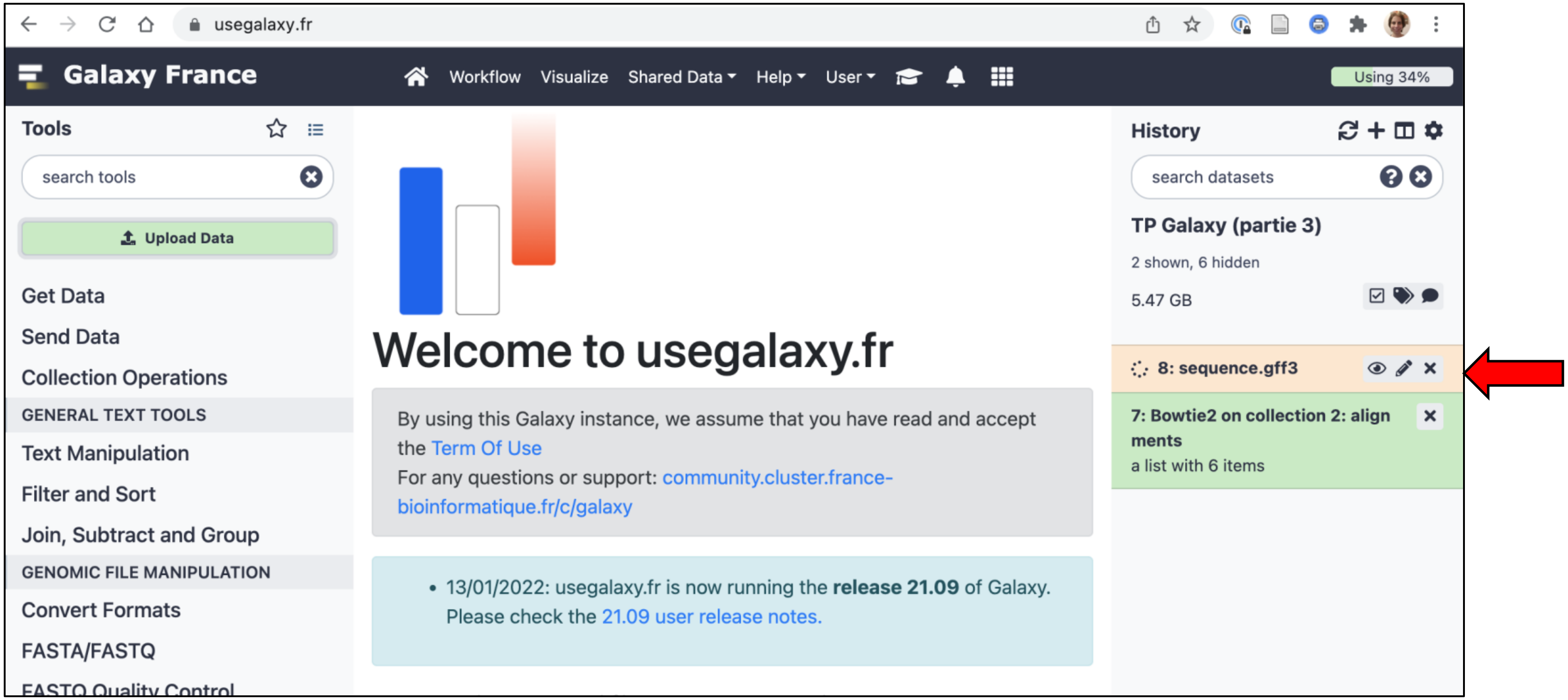
● Complete Record
○ Coding Sequences
○ Gene Features

Choose Destination
● File ○ Clipboard
○ Collections ○ Analysis Tool

Download 4 items.
Format: GFF3

Deinococcus radiodurans R1 strain:R1 | isolate:orsay

Etape 4 : Importer le fichier d'annotation dans l'historique



Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and accept the Term Of Use

For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy

- 13/01/2022: usegalaxy.fr is now running the **release 21.09** of Galaxy. Please check the 21.09 user release notes.

TP Galaxy (partie 3)

2 shown, 6 hidden

5.47 GB

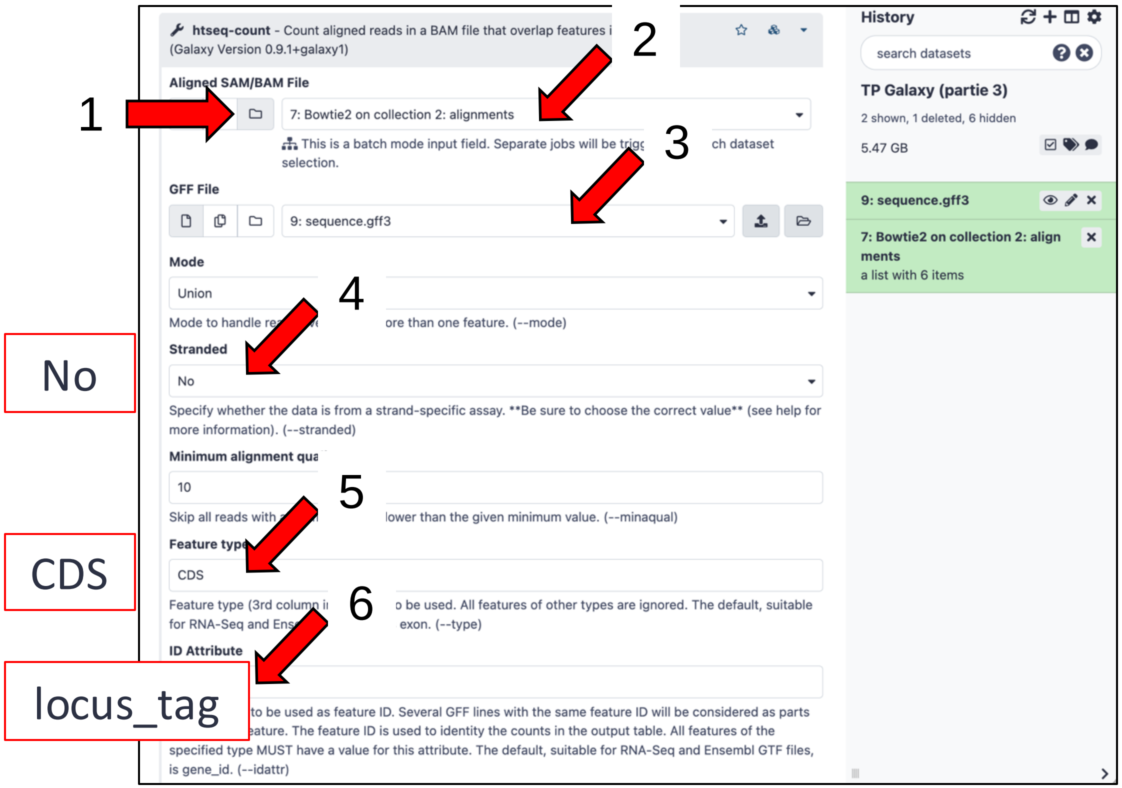
8: sequence.gff3

7: Bowtie2 on collection 2: alignments

a list with 6 items

Etape 5 : Quantifier les niveaux d’expression des gènes

- Utiliser l’outil HTSeq-Count. Utiliser les paramètres présentés ci-dessous :



htseq-count - Count aligned reads in a BAM file that overlap features i
(Galaxy Version 0.9.1+galaxy1)

Aligned SAM/BAM File

7: Bowtie2 on collection 2: alignments

This is a batch mode input field. Separate jobs will be trigg ch dataset selection.

GFF File

9: sequence.gff3

Mode

Union

Mode to handle re ore than one feature. (--mode)

Stranded

No

Specify whether the data is from a strand-specific assay. **Be sure to choose the correct value** (see help for more information). (--stranded)

Minimum alignment qua

10

Skip all reads with lower than the given minimum value. (--minaqual)

Feature type

CDS

Feature type (3rd column i o be used. All features of other types are ignored. The default, suitable for RNA-Seq and Ens exon. (--type)

ID Attribute

to be used as feature ID. Several GFF lines with the same feature ID will be considered as parts eature. The feature ID is used to identity the counts in the output table. All features of the specified type MUST have a value for this attribute. The default, suitable for RNA-Seq and Ensembl GTF files, is gene_id. (--idattr)

History

search datasets

TP Galaxy (partie 3)

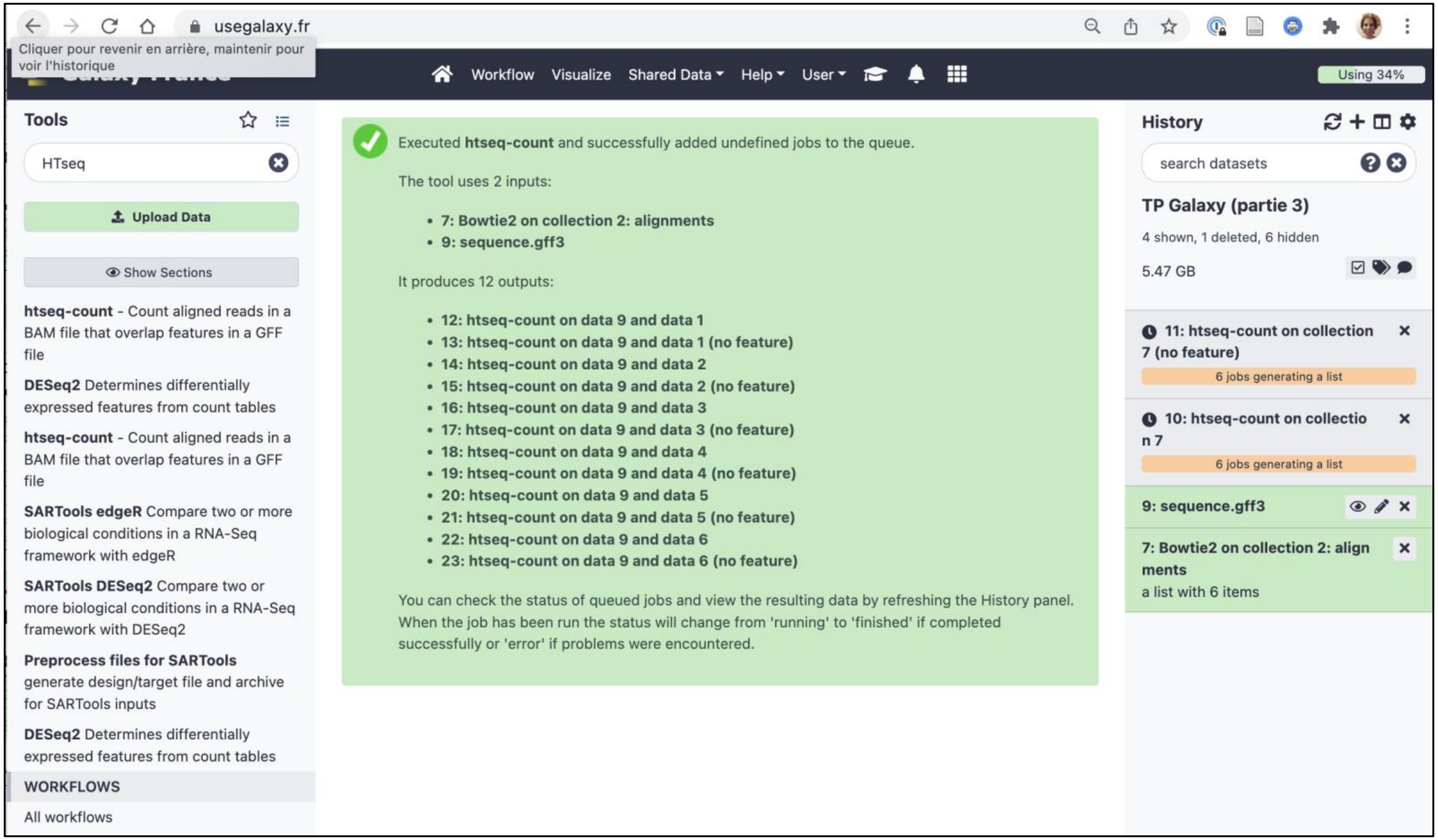
2 shown, 1 deleted, 6 hidden

5.47 GB

9: sequence.gff3

7: Bowtie2 on collection 2: alignments
a list with 6 items

Temps attente (un peu long…)



htseq-count - Count aligned reads in a BAM file that overlap features in a GFF file

DESeq2 Determines differentially expressed features from count tables

htseq-count - Count aligned reads in a BAM file that overlap features in a GFF file

SARTools edgeR Compare two or more biological conditions in a RNA-Seq framework with edgeR

SARTools DESeq2 Compare two or more biological conditions in a RNA-Seq framework with DESeq2

Preprocess files for SARTools generate design/target file and archive for SARTools inputs

DESeq2 Determines differentially expressed features from count tables

Executed **htseq-count** and successfully added undefined jobs to the queue.

The tool uses 2 inputs:

- **7: Bowtie2 on collection 2: alignments**
- **9: sequence.gff3**

It produces 12 outputs:

- **12: htseq-count on data 9 and data 1**
- **13: htseq-count on data 9 and data 1 (no feature)**
- **14: htseq-count on data 9 and data 2**
- **15: htseq-count on data 9 and data 2 (no feature)**
- **16: htseq-count on data 9 and data 3**
- **17: htseq-count on data 9 and data 3 (no feature)**
- **18: htseq-count on data 9 and data 4**
- **19: htseq-count on data 9 and data 4 (no feature)**
- **20: htseq-count on data 9 and data 5**
- **21: htseq-count on data 9 and data 5 (no feature)**
- **22: htseq-count on data 9 and data 6**
- **23: htseq-count on data 9 and data 6 (no feature)**

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

TP Galaxy (partie 3)

4 shown, 1 deleted, 6 hidden

5.47 GB

11: htseq-count on collection 7 (no feature)

6 jobs generating a list

10: htseq-count on collectio n 7

6 jobs generating a list

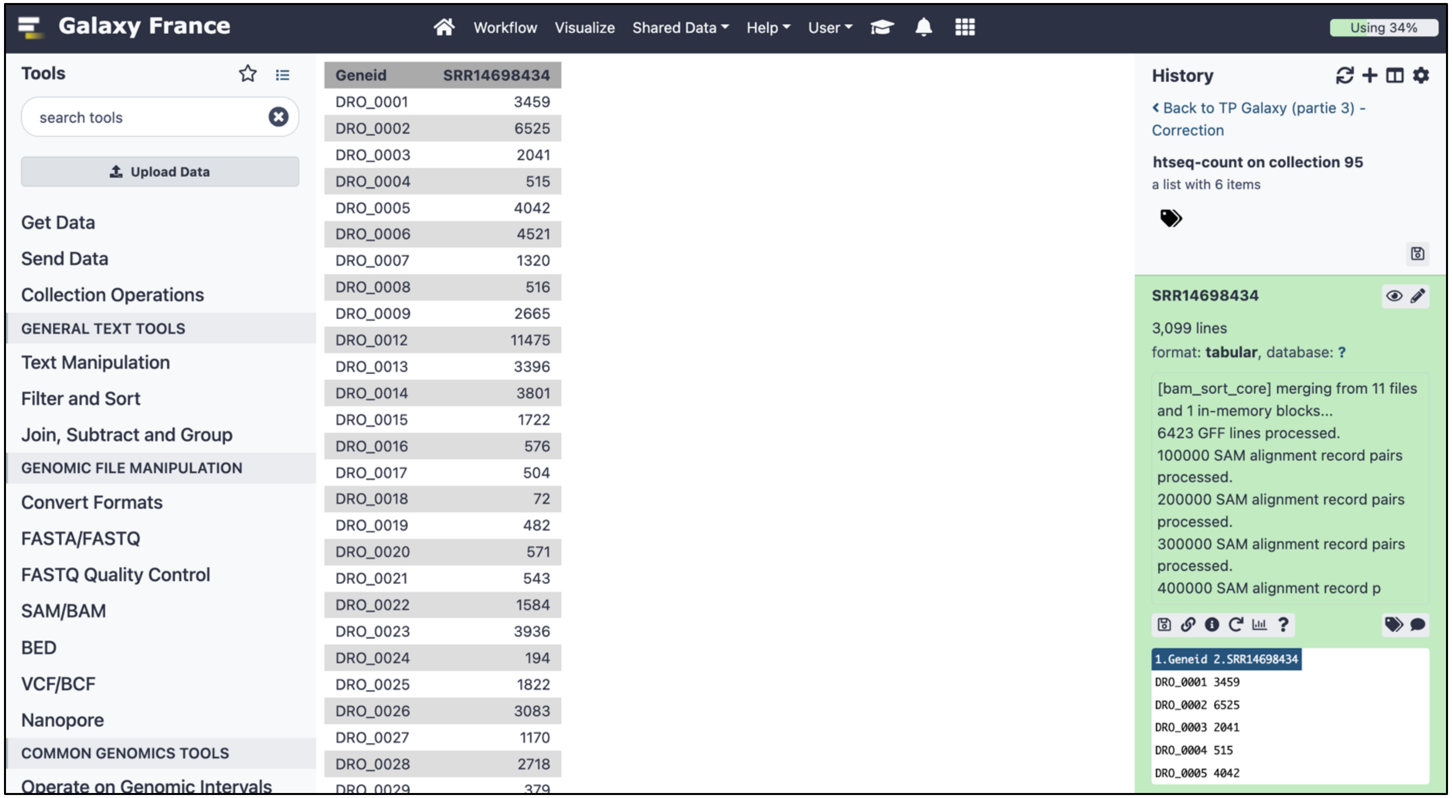
9: sequence.gff3

7: Bowtie2 on collection 2: align ments

a list with 6 items

Etape 5 : Quantifier les niveaux d’expression des gènes

- Vérifier les résultats, des <u>valeurs différentes de 0</u> doivent être notées dans les fichiers.



Geneid	SRR14698434
DRO_0001	3459
DRO_0002	6525
DRO_0003	2041
DRO_0004	515
DRO_0005	4042
DRO_0006	4521
DRO_0007	1320
DRO_0008	516
DRO_0009	2665
DRO_0012	11475
DRO_0013	3396
DRO_0014	3801
DRO_0015	1722
DRO_0016	576
DRO_0017	504
DRO_0018	72
DRO_0019	482
DRO_0020	571
DRO_0021	543
DRO_0022	1584
DRO_0023	3936
DRO_0024	194
DRO_0025	1822
DRO_0026	3083
DRO_0027	1170
DRO_0028	2718
DRO_0029	379

htseq-count on collection 95
a list with 6 items

SRR14698434

3,099 lines
format: **tabular**, database: ?

```
[bam_sort_core] merging from 11 files and 1 in-memory blocks...
6423 GFF lines processed.
100000 SAM alignment record pairs processed.
200000 SAM alignment record pairs processed.
300000 SAM alignment record pairs processed.
400000 SAM alignment record p
```

```
1.Geneid 2.SRR14698434
DRO_0001 3459
DRO_0002 6525
DRO_0003 2041
DRO_0004 515
DRO_0005 4042
```

Fin de la partie 3 ☺

Gaëlle LELANDAIS

Version du document : 11/02/2025