Travaux pratiques – Galaxy* (partie 1)

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

check for updates

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

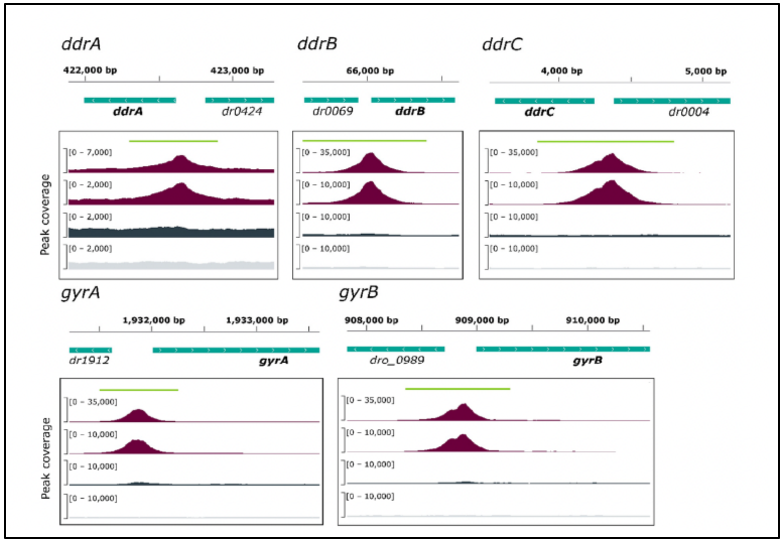
Received: 19 July 2021
Accepted: 10 September 2021
Published: 25 September 2021

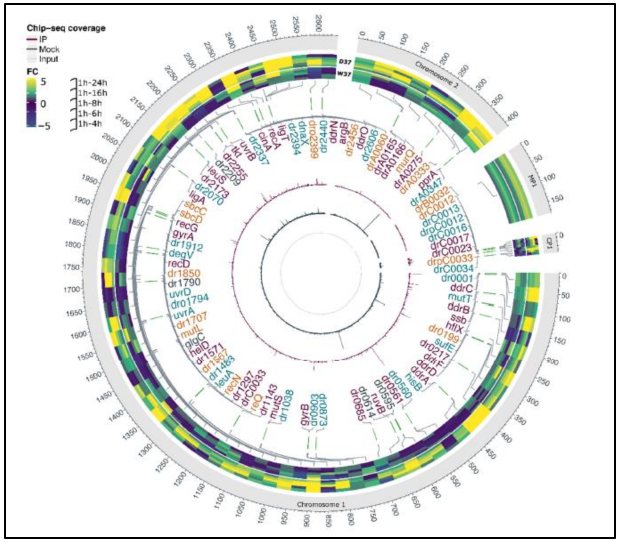
1. Introduction

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

Cells **2021**, *10*, 2536. https://doi.org/10.3390/cells10102536 https://www.mdpi.com/journal/cells

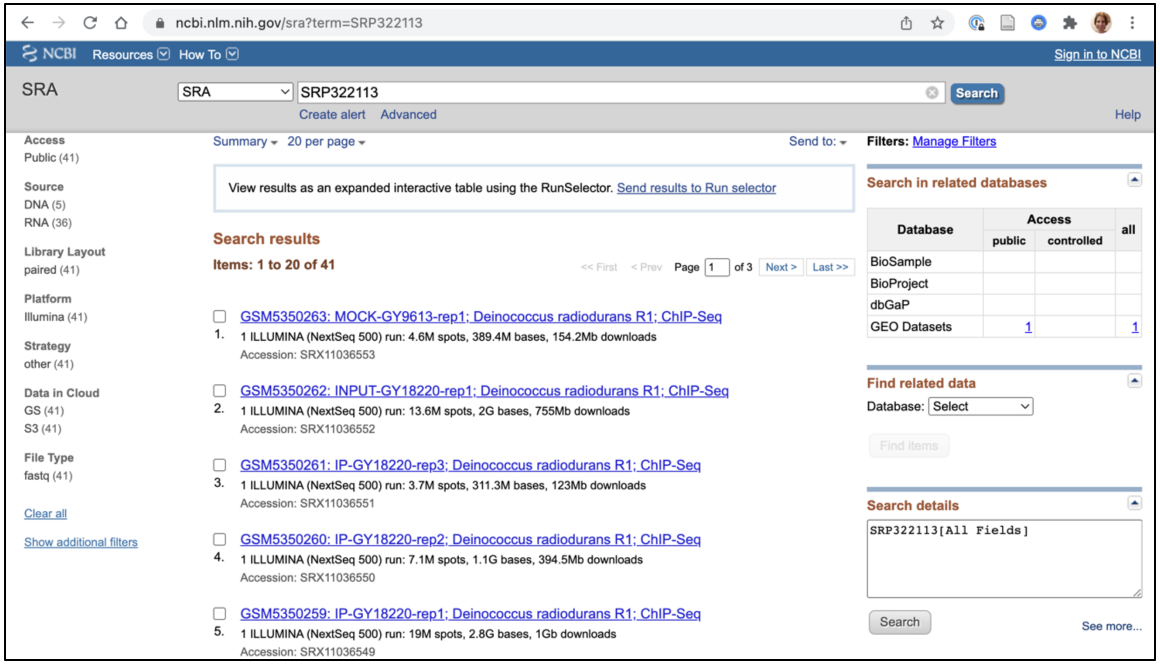




Etape 2 : Importer des fichiers FASTQ (RNAseq)

- Disponibilité des données sur SRA :

  https://www.ncbi.nlm.nih.gov/sra?term=SRP322113



SRA | SRP322113

Access
Public (41)

Source
DNA (5)
RNA (36)

Library Layout
paired (41)

Platform
Illumina (41)

Strategy
other (41)

Data in Cloud
GS (41)
S3 (41)

File Type
fastq (41)

Clear all

Show additional filters

View results as an expanded interactive table using the RunSelector. Send results to Run selector

Search results

Items: 1 to 20 of 41

<< First < Prev Page 1 of 3 Next > Last >>

☐ 1. GSM5350263: MOCK-GY9613-rep1; Deinococcus radiodurans R1; ChIP-Seq
1 ILLUMINA (NextSeq 500) run: 4.6M spots, 389.4M bases, 154.2Mb downloads
Accession: SRX11036553

☐ 2. GSM5350262: INPUT-GY18220-rep1; Deinococcus radiodurans R1; ChIP-Seq
1 ILLUMINA (NextSeq 500) run: 13.6M spots, 2G bases, 755Mb downloads
Accession: SRX11036552

☐ 3. GSM5350261: IP-GY18220-rep3; Deinococcus radiodurans R1; ChIP-Seq
1 ILLUMINA (NextSeq 500) run: 3.7M spots, 311.3M bases, 123Mb downloads
Accession: SRX11036551

☐ 4. GSM5350260: IP-GY18220-rep2; Deinococcus radiodurans R1; ChIP-Seq
1 ILLUMINA (NextSeq 500) run: 7.1M spots, 1.1G bases, 394.5Mb downloads
Accession: SRX11036550

☐ 5. GSM5350259: IP-GY18220-rep1; Deinococcus radiodurans R1; ChIP-Seq
1 ILLUMINA (NextSeq 500) run: 19M spots, 2.8G bases, 1Gb downloads
Accession: SRX11036549

Search in related databases

Database	Access		all
	public	controlled	
BioSample			
BioProject			
dbGaP			
GEO Datasets	1		1

Find related data

Database: Select

Find items

Search details

SRP322113[All Fields]

Search

See more...

Etape 2 : Importer des fichiers FASTQ (RNAseq)

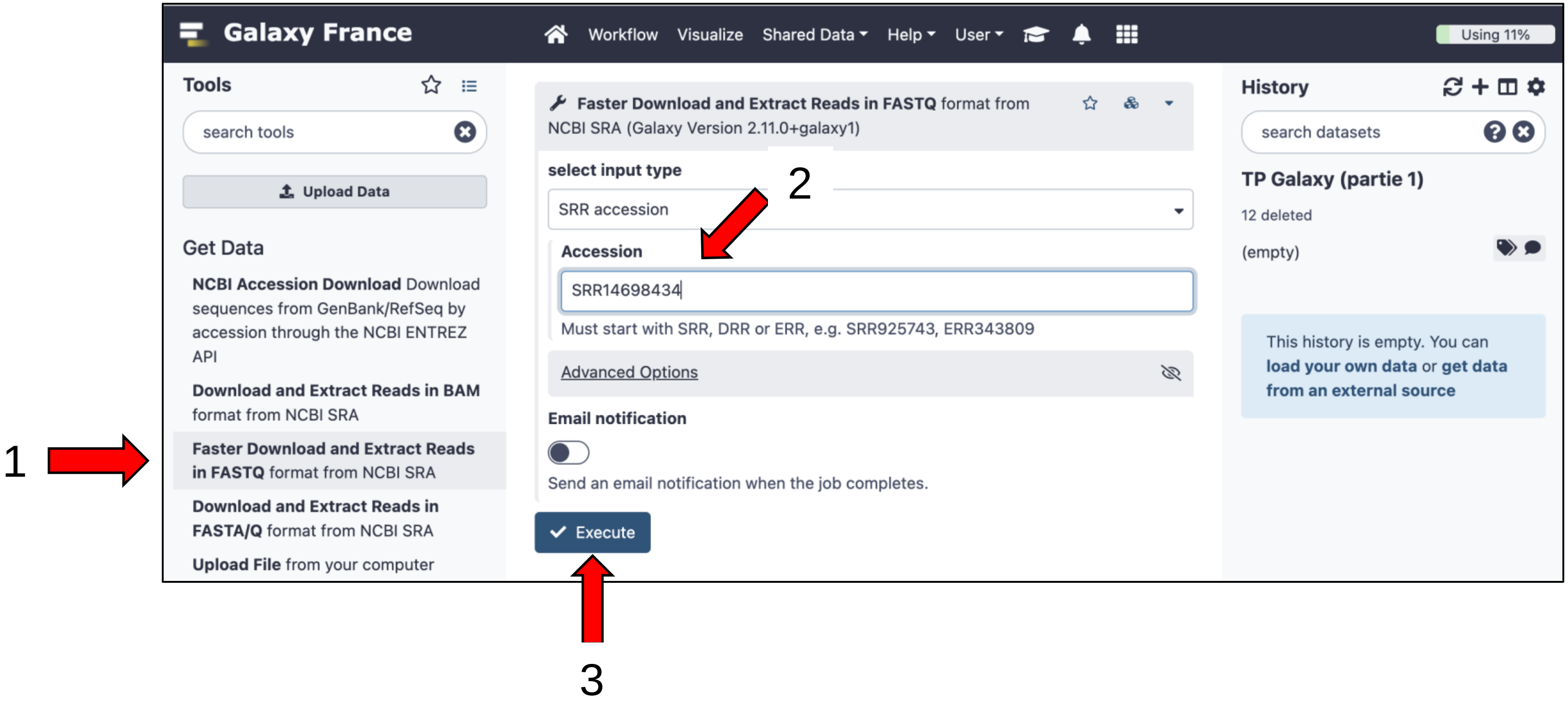
- Disponibilité des données sur SRA :

  https://www.ncbi.nlm.nih.gov/sra?term=SRP322113

- Première partie du TP réalisée sur un unique échantillon RNAseq :
  - SRR14698434

Etape 2 : Importer des fichiers FASTQ (RNAseq)

- Importer le fichier FASTQ dans l'historique de travail
  - Outil : Get Data / Faster Download and Extract Reads in FASTQ



Faster Download and Extract Reads in FASTQ format from NCBI SRA (Galaxy Version 2.11.0+galaxy1)

select input type
SRR accession

Accession
SRR14698434
Must start with SRR, DRR or ERR, e.g. SRR925743, ERR343809

Advanced Options

Email notification
Send an email notification when the job completes.

Execute

Get Data

NCBI Accession Download Download sequences from GenBank/RefSeq by accession through the NCBI ENTREZ API

Download and Extract Reads in BAM format from NCBI SRA

Faster Download and Extract Reads in FASTQ format from NCBI SRA

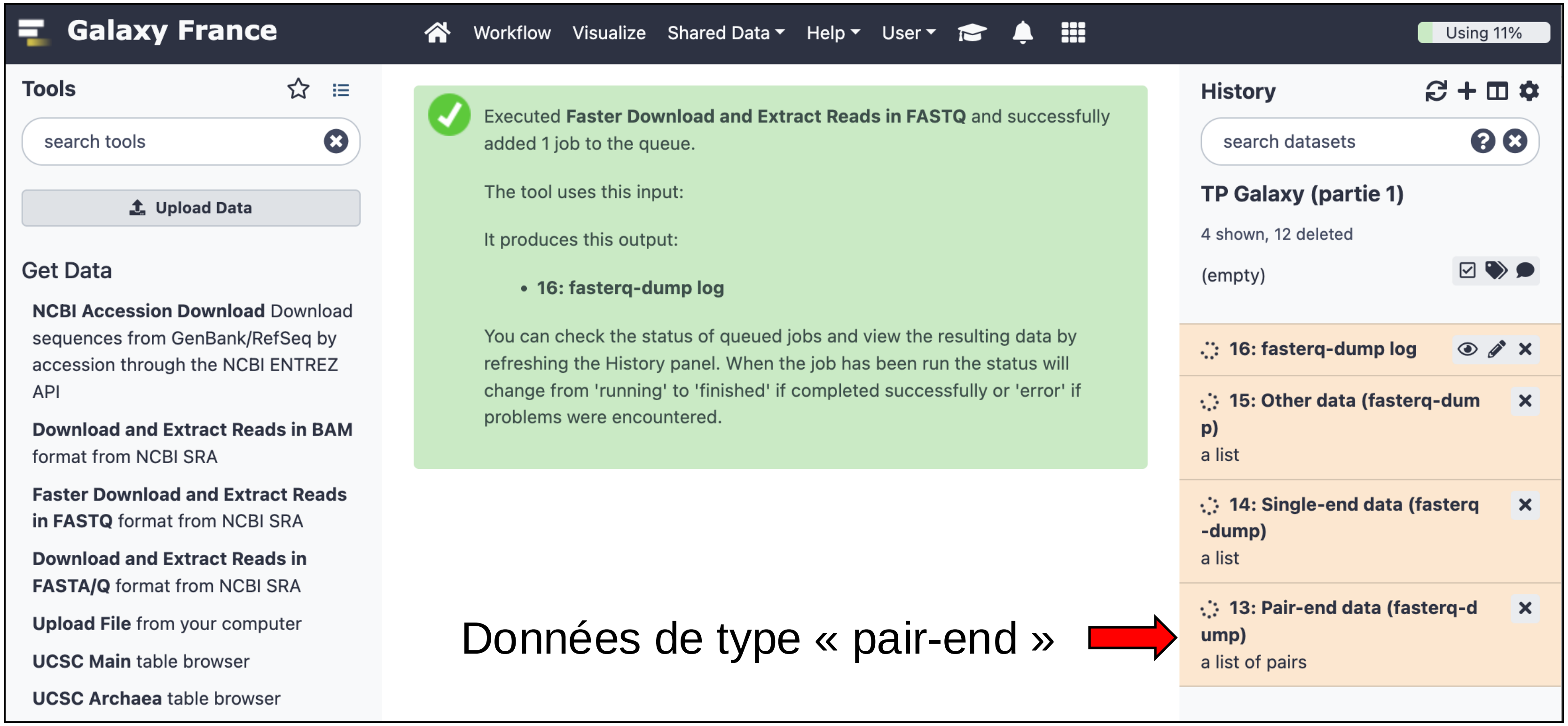
Download and Extract Reads in FASTA/Q format from NCBI SRA

Upload File from your computer

TP Galaxy (partie 1)
12 deleted
(empty)

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

Temps attente (un peu long…)



Get Data

NCBI Accession Download Download sequences from GenBank/RefSeq by accession through the NCBI ENTREZ API

Download and Extract Reads in BAM format from NCBI SRA

Faster Download and Extract Reads in FASTQ format from NCBI SRA

Download and Extract Reads in FASTA/Q format from NCBI SRA

Upload File from your computer

UCSC Main table browser

UCSC Archaea table browser

Executed **Faster Download and Extract Reads in FASTQ** and successfully added 1 job to the queue.

The tool uses this input:

It produces this output:

- **16: fasterq-dump log**

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 1)

4 shown, 12 deleted

(empty)

16: fasterq-dump log

15: Other data (fasterq-dump)
a list

14: Single-end data (fasterq-dump)
a list

13: Pair-end data (fasterq-dump)
a list of pairs

Données de type « pair-end »

Historique de secours

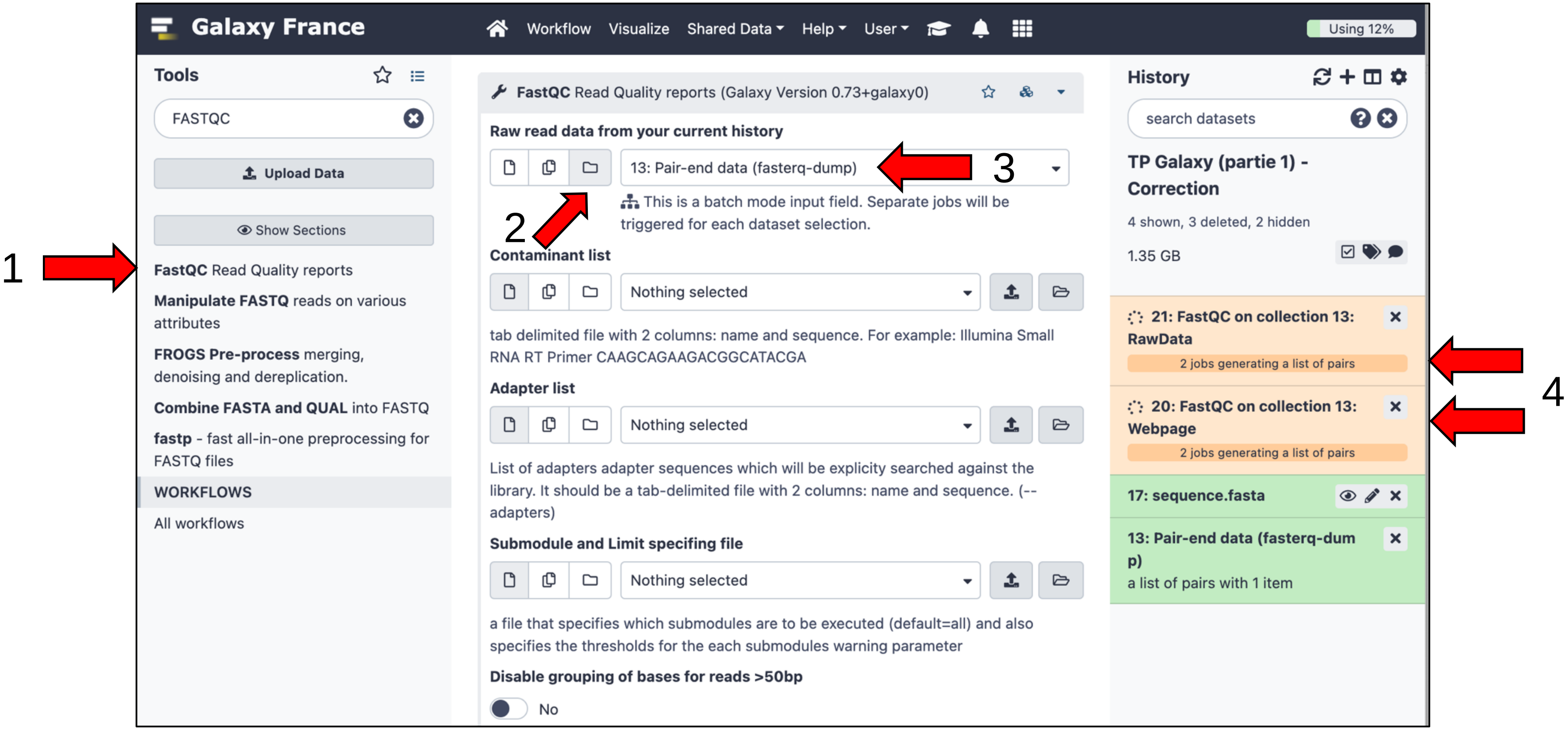
- Si l’importation des fichiers FASTQ depuis la banque de données SRA est trop longue, vous pouvez importer les données de mon historique de secours :

https://usegalaxy.fr/u/gaellelelandais/h/backup---tp-galaxy-partie-1

- Passez ensuite à l’étape suivante !

Etape 3 : Contrôle de la qualité des séquences (1/2)

- Appliquer l’outil FASTQC sur le fichier FASTQ



FASTQC

FastQC Read Quality reports

Manipulate FASTQ reads on various attributes

FROGS Pre-process merging, denoising and dereplication.

Combine FASTA and QUAL into FASTQ

fastp - fast all-in-one preprocessing for FASTQ files

FastQC Read Quality reports (Galaxy Version 0.73+galaxy0)

Raw read data from your current history

13: Pair-end data (fasterq-dump)

This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Contaminant list

Nothing selected

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA

Adapter list

Nothing selected

List of adapters adapter sequences which will be explicity searched against the library. It should be a tab-delimited file with 2 columns: name and sequence. (--adapters)

Submodule and Limit specifing file

Nothing selected

a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for the each submodules warning parameter

Disable grouping of bases for reads >50bp

No

search datasets

TP Galaxy (partie 1) - Correction

4 shown, 3 deleted, 2 hidden

1.35 GB

21: FastQC on collection 13: RawData

2 jobs generating a list of pairs

20: FastQC on collection 13: Webpage

2 jobs generating a list of pairs

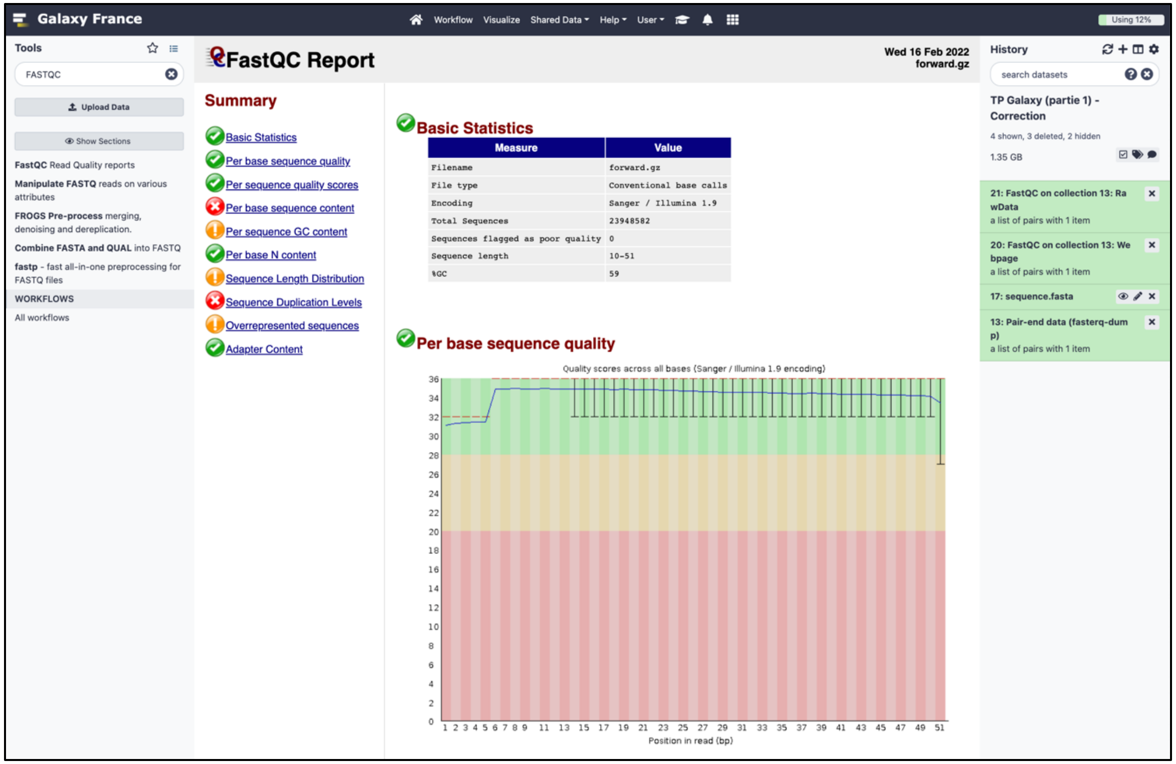
17: sequence.fasta

13: Pair-end data (fasterq-dump)

a list of pairs with 1 item

Etape 3 : Contrôle de la qualité des séquences (2/2)

- Accéder aux graphiques de résultats



FastQC Read Quality reports

Manipulate FASTQ reads on various attributes

FROGS Pre-process merging, denoising and dereplication.

Combine FASTA and QUAL into FASTQ

fastp - fast all-in-one preprocessing for FASTQ files

WORKFLOWS

All workflows

FastQC Report

Wed 16 Feb 2022
forward.gz

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ⚠ Per sequence GC content
- ✓ Per base N content
- ⚠ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ⚠ Overrepresented sequences
- ✓ Adapter Content

✓ Basic Statistics

Measure	Value
Filename	forward.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	23948582
Sequences flagged as poor quality	0
Sequence length	10-51
%GC	59

✓ Per base sequence quality

History

search datasets

TP Galaxy (partie 1) - Correction

4 shown, 3 deleted, 2 hidden

1.35 GB

21: FastQC on collection 13: RawData
a list of pairs with 1 item

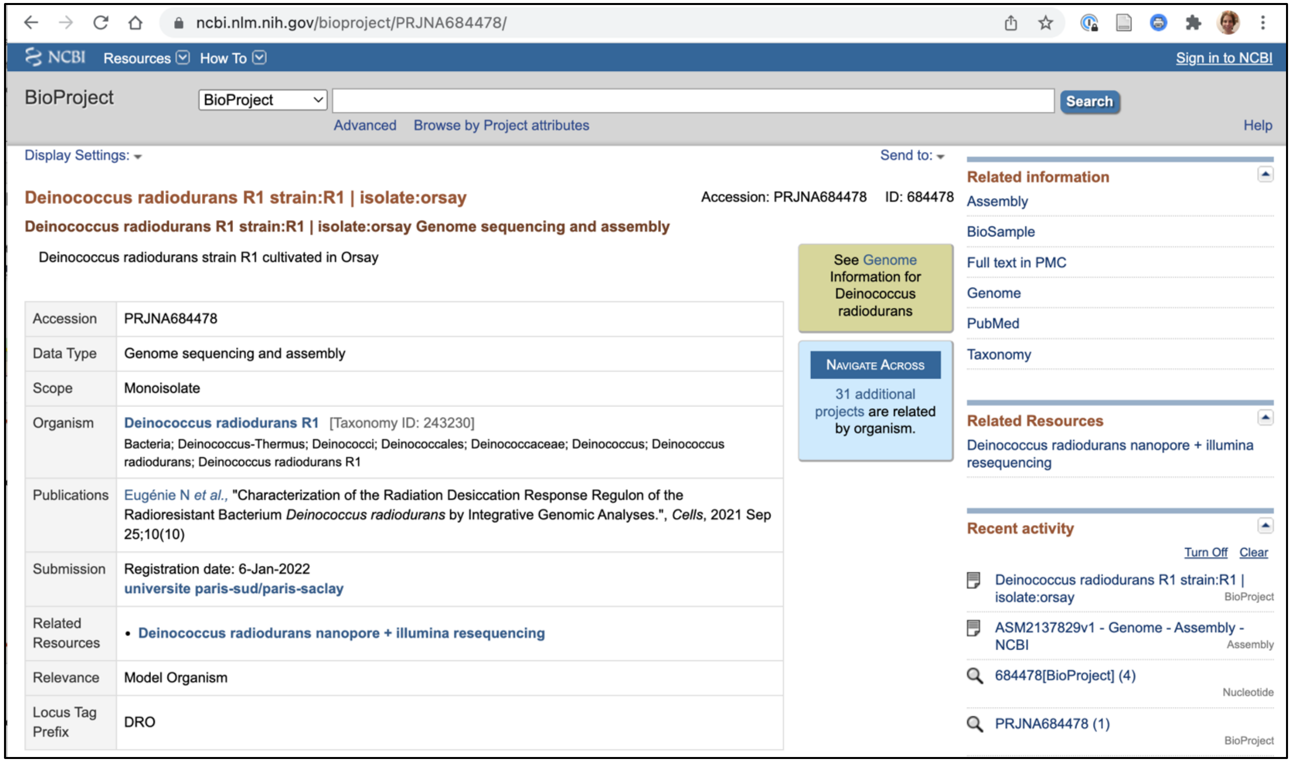
20: FastQC on collection 13: Webpage
a list of pairs with 1 item

17: sequence.fasta

13: Pair-end data (fasterq-dump)
a list of pairs with 1 item

Etape 4 : Télécharger le génome de référence

- 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

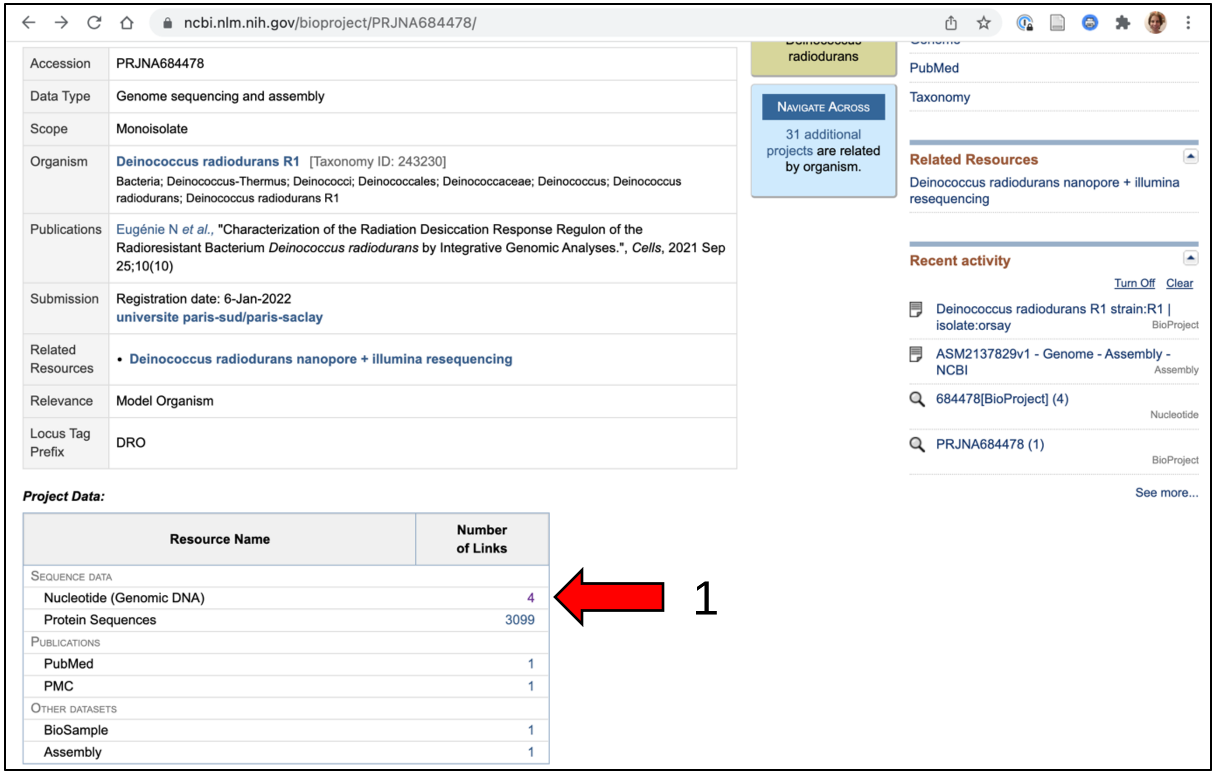
ASM2137829v1 - Genome - Assembly - NCBI Assembly

684478[BioProject] (4) Nucleotide

PRJNA684478 (1) BioProject

Etape 4 : Télécharger le génome de référence

- Télécharger les 4 séquences sous la forme d’un fichier 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

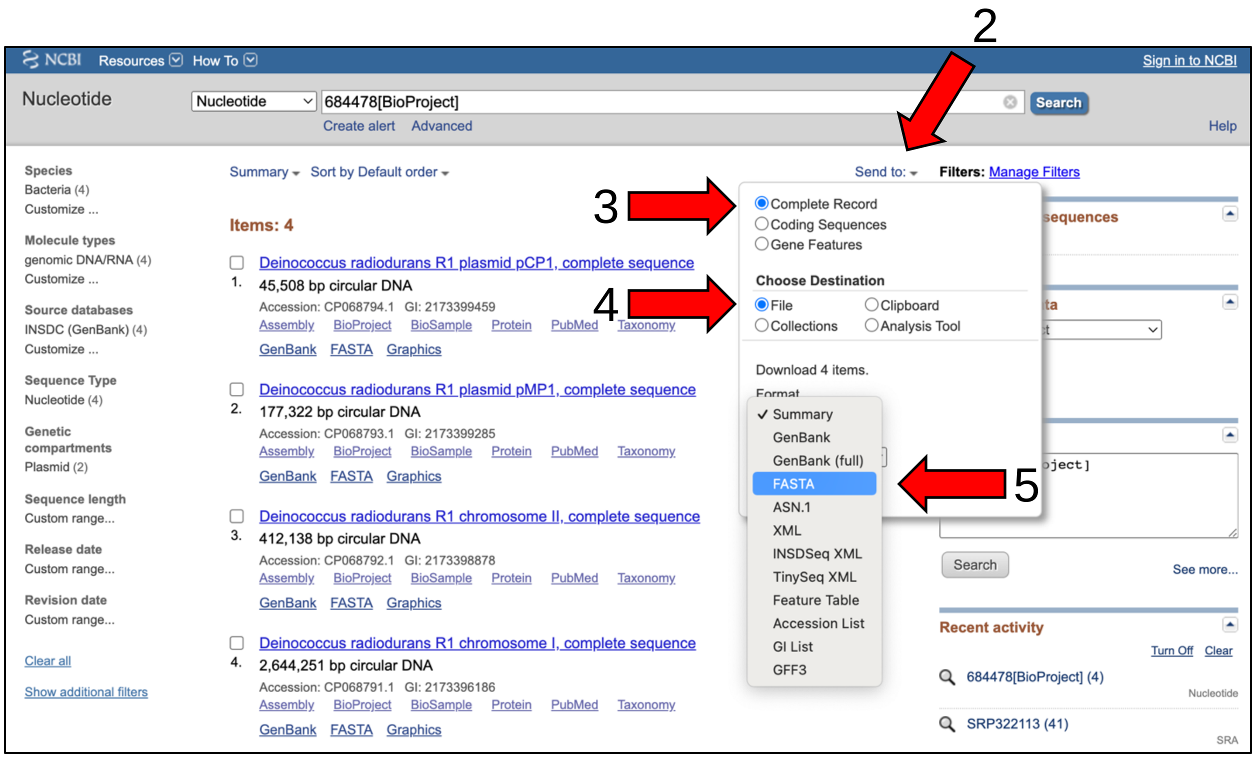
PRJNA684478 (1) — BioProject

See more...

1

Etape 4 : Télécharger le génome de référence

- Télécharger les 4 séquences sous la forme d'un fichier FASTA :



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

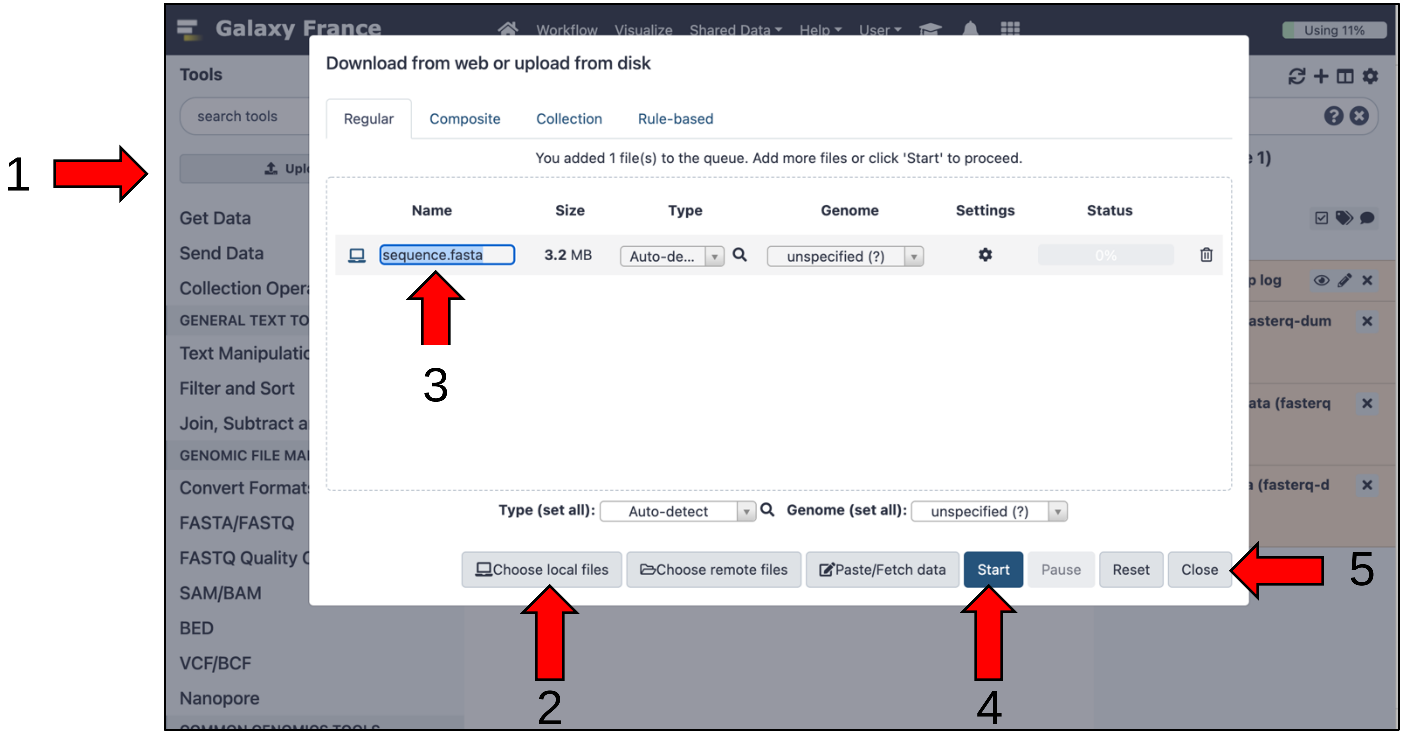
● Complete Record
○ Coding Sequences
○ Gene Features

Choose Destination
● File ○ Clipboard
○ Collections ○ Analysis Tool

Download 4 items.
Format
✓ Summary

684478[BioProject] (4) Nucleotide
SRP322113 (41) SRA

Etape 5 : Importer le génome de référence dans Galaxy



Download from web or upload from disk

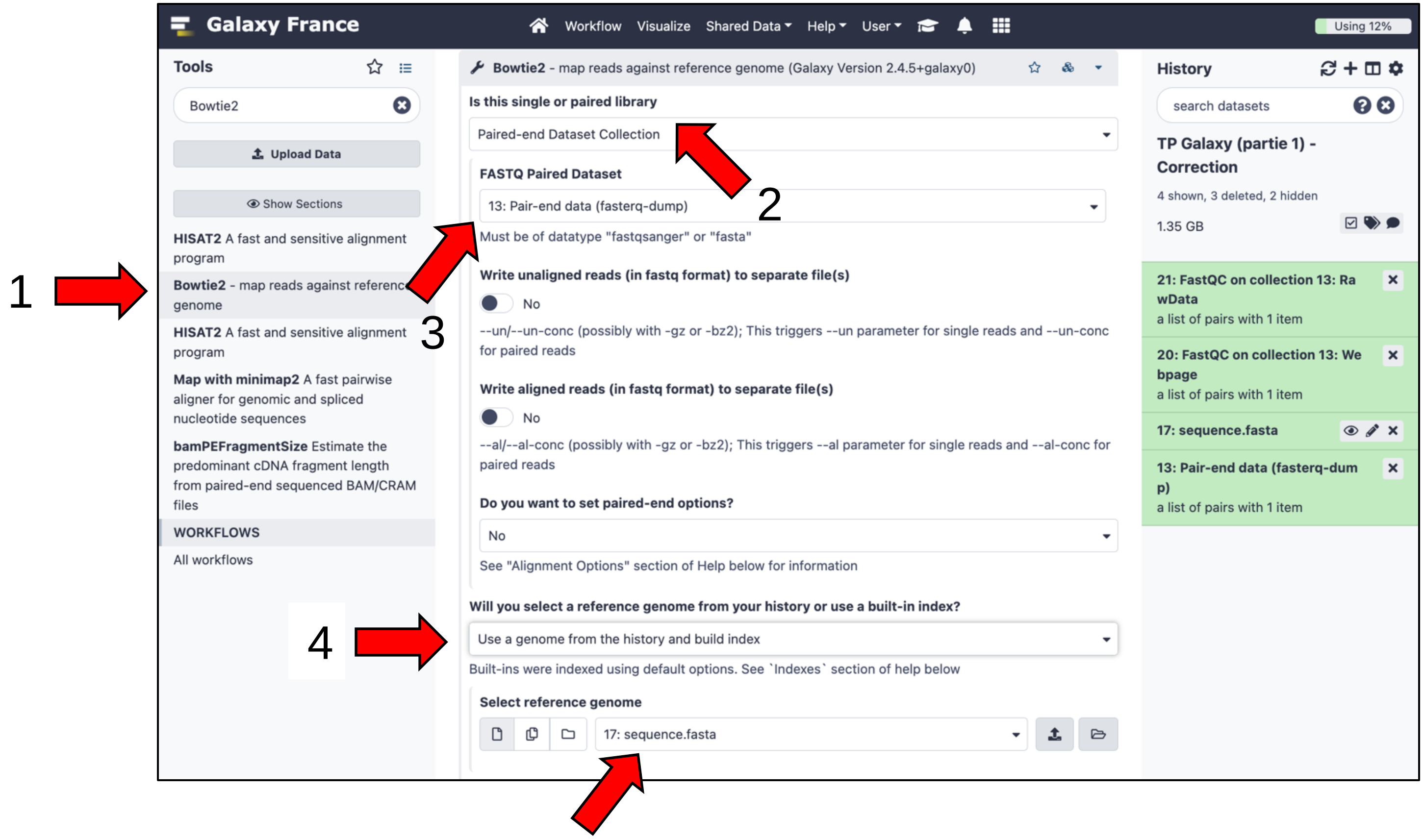
You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
sequence.fasta	3.2 MB	Auto-de...	unspecified (?)		0%

Type (set all): Auto-detect **Genome (set all):** unspecified (?)

Etape 6 : Alignement des « reads » sur le génome

- Appliquer l'outil Bowtie2 sur le fichier FASTQ



Tools search: Bowtie2

HISAT2 A fast and sensitive alignment program

Bowtie2 - map reads against reference genome

HISAT2 A fast and sensitive alignment program

Map with minimap2 A fast pairwise aligner for genomic and spliced nucleotide sequences

bamPEFragmentSize Estimate the predominant cDNA fragment length from paired-end sequenced BAM/CRAM files

WORKFLOWS

All workflows

Bowtie2 - map reads against reference genome (Galaxy Version 2.4.5+galaxy0)

Is this single or paired library
Paired-end Dataset Collection

FASTQ Paired Dataset
13: Pair-end data (fasterq-dump)
Must be of datatype "fastqsanger" or "fasta"

Write unaligned reads (in fastq format) to separate file(s)
No
--un/--un-conc (possibly with -gz or -bz2); This triggers --un parameter for single reads and --un-conc for paired reads

Write aligned reads (in fastq format) to separate file(s)
No
--al/--al-conc (possibly with -gz or -bz2); This triggers --al parameter for single reads and --al-conc for paired reads

Do you want to set paired-end options?
No
See "Alignment Options" section of Help below for information

Will you select a reference genome from your history or use a built-in index?
Use a genome from the history and build index
Built-ins were indexed using default options. See `Indexes` section of help below

Select reference genome
17: sequence.fasta

TP Galaxy (partie 1) - Correction
4 shown, 3 deleted, 2 hidden
1.35 GB

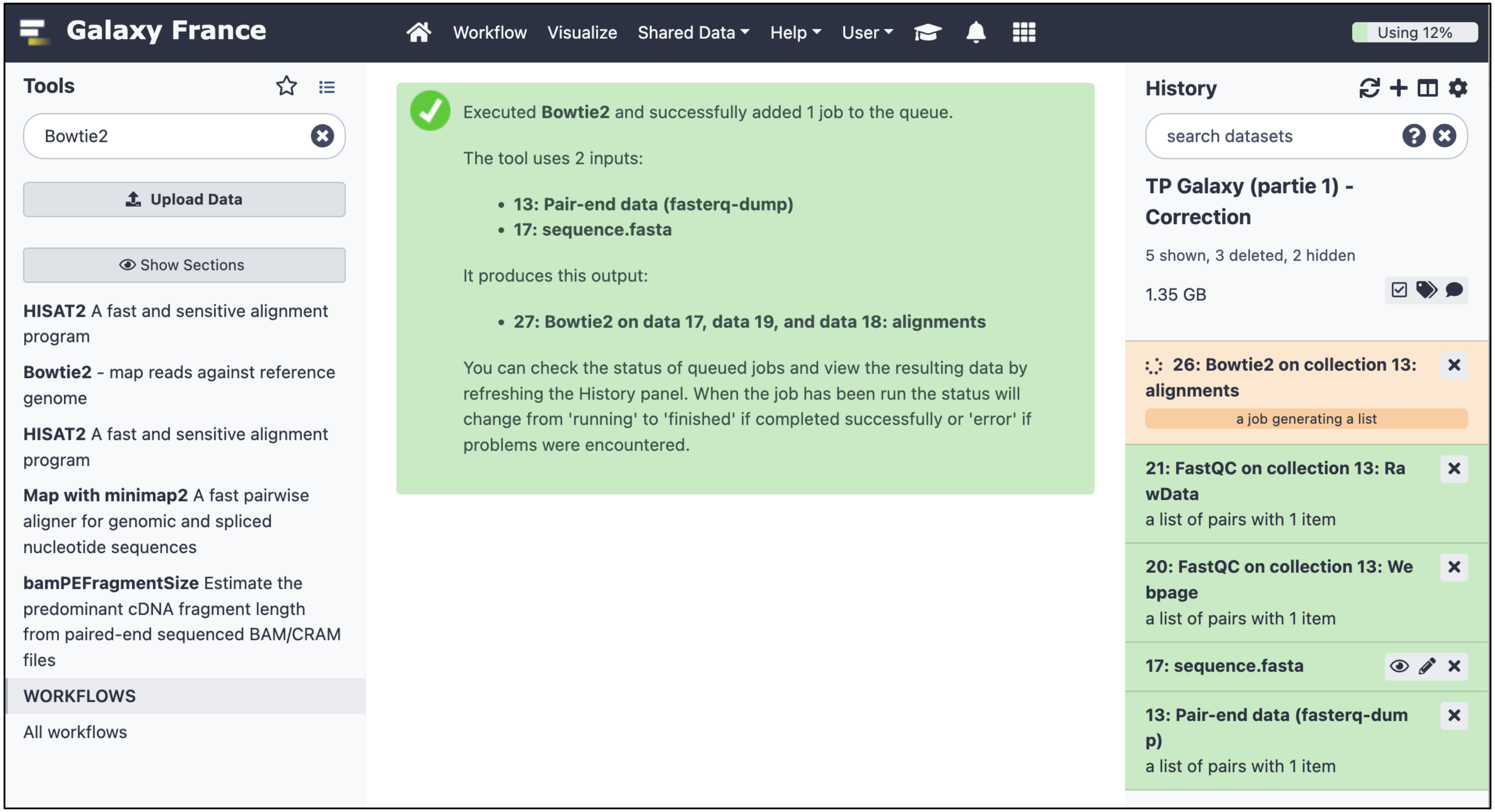
21: FastQC on collection 13: RawData
a list of pairs with 1 item

20: FastQC on collection 13: Webpage
a list of pairs with 1 item

17: sequence.fasta

13: Pair-end data (fasterq-dump)
a list of pairs with 1 item

Temps attente (un peu long…)



Bowtie2

HISAT2 A fast and sensitive alignment program

Bowtie2 - map reads against reference genome

HISAT2 A fast and sensitive alignment program

Map with minimap2 A fast pairwise aligner for genomic and spliced nucleotide sequences

bamPEFragmentSize Estimate the predominant cDNA fragment length from paired-end sequenced BAM/CRAM files

WORKFLOWS

All workflows

Executed **Bowtie2** and successfully added 1 job to the queue.

The tool uses 2 inputs:

- **13: Pair-end data (fasterq-dump)**
- **17: sequence.fasta**

It produces this output:

- **27: Bowtie2 on data 17, data 19, and data 18: alignments**

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 1) - Correction

5 shown, 3 deleted, 2 hidden

1.35 GB

26: Bowtie2 on collection 13: alignments

a job generating a list

21: FastQC on collection 13: RawData

a list of pairs with 1 item

20: FastQC on collection 13: Webpage

a list of pairs with 1 item

17: sequence.fasta

13: Pair-end data (fasterq-dump)

a list of pairs with 1 item

Fin de la partie 1 ☺

Gaëlle LELANDAIS

Version du document : 05/02/2025