



Travaux pratiques – Galaxy* (partie 2)

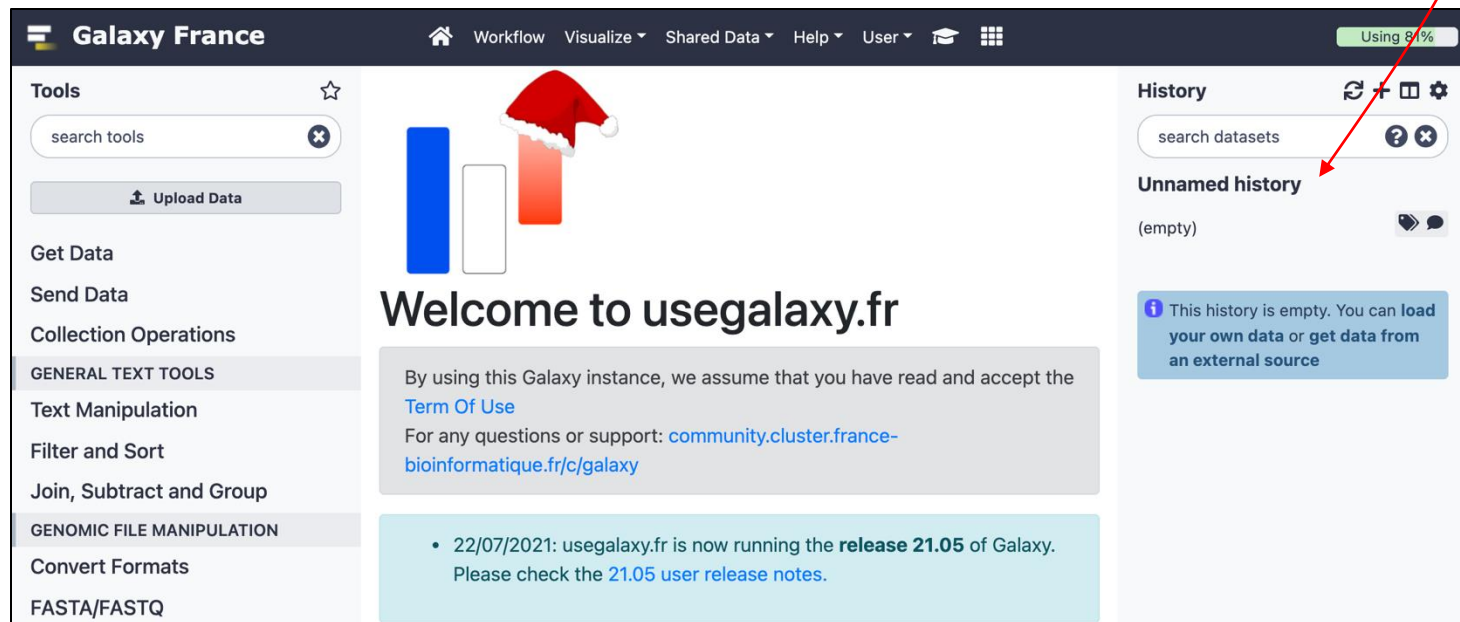
Gaëlle Lelandais et Fabrice Confalonieri

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



Données utilisées pour le TP



Article

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

Nicolas Eugénie , Yvan Zivanovic , Gaelle Lelandaïs, 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.lelandaïs@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 (F.C.)
[†] 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

1. Introduction

Deinococcus radiodurans is one of the most resistant bacteria to genotoxic agent exposure and desiccation isolated to date [1–4]. Unlike radioresistant 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

check for updates

Citation: Eugénie, N.; Zivanovic, Y.; Lelandaïs, 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

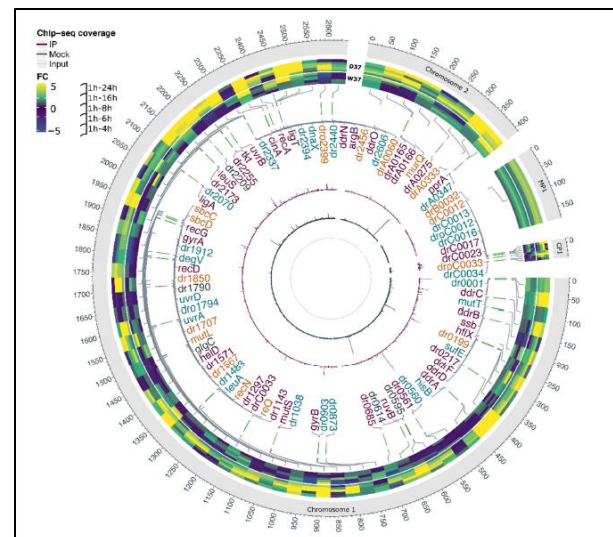
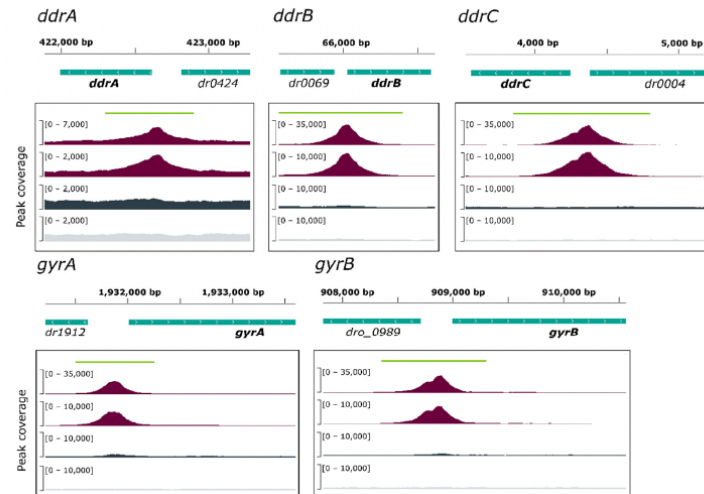
Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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>

The screenshot shows the NCBI SRA search results for the term SRP322113. The page includes a search bar, a left sidebar with filters, and a main results area. The results are displayed as a list of 5 items, each with a checkbox, a link to the run, and details about the run (e.g., 1 ILLUMINA (NextSeq 500) run: 4.6M spots, 389.4M bases, 154.2Mb downloads). The results are paginated, showing items 1 to 20 of 41. On the right, there is a section for 'Search in related databases' with a table showing the number of results in various databases (BioSample, BioProject, dbGaP, GEO Datasets) and a 'Find related data' section with a dropdown menu and a 'Find items' button. At the bottom right, there is a 'Search details' section with a search bar and a 'Search' button.

NCBI Resources How To Sign in to NCBI

SRA SRA SRP322113 Search

Create alert Advanced Help

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](#)

Summary 20 per page Send to Filters: [Manage 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 >>

☐ [GSM5350263: MOCK-GY9613-rep1: Deinococcus radiodurans R1: ChIP-Seq](#)

1. 1 ILLUMINA (NextSeq 500) run: 4.6M spots, 389.4M bases, 154.2Mb downloads
Accession: SRX11036553

☐ [GSM5350262: INPUT-GY18220-rep1: Deinococcus radiodurans R1: ChIP-Seq](#)

2. 1 ILLUMINA (NextSeq 500) run: 13.6M spots, 2G bases, 755Mb downloads
Accession: SRX11036552

☐ [GSM5350261: IP-GY18220-rep3: Deinococcus radiodurans R1: ChIP-Seq](#)

3. 1 ILLUMINA (NextSeq 500) run: 3.7M spots, 311.3M bases, 123Mb downloads
Accession: SRX11036551

☐ [GSM5350260: IP-GY18220-rep2: Deinococcus radiodurans R1: ChIP-Seq](#)

4. 1 ILLUMINA (NextSeq 500) run: 7.1M spots, 1.1G bases, 394.5Mb downloads
Accession: SRX11036550

☐ [GSM5350259: IP-GY18220-rep1: Deinococcus radiodurans R1: ChIP-Seq](#)

5. 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](#)

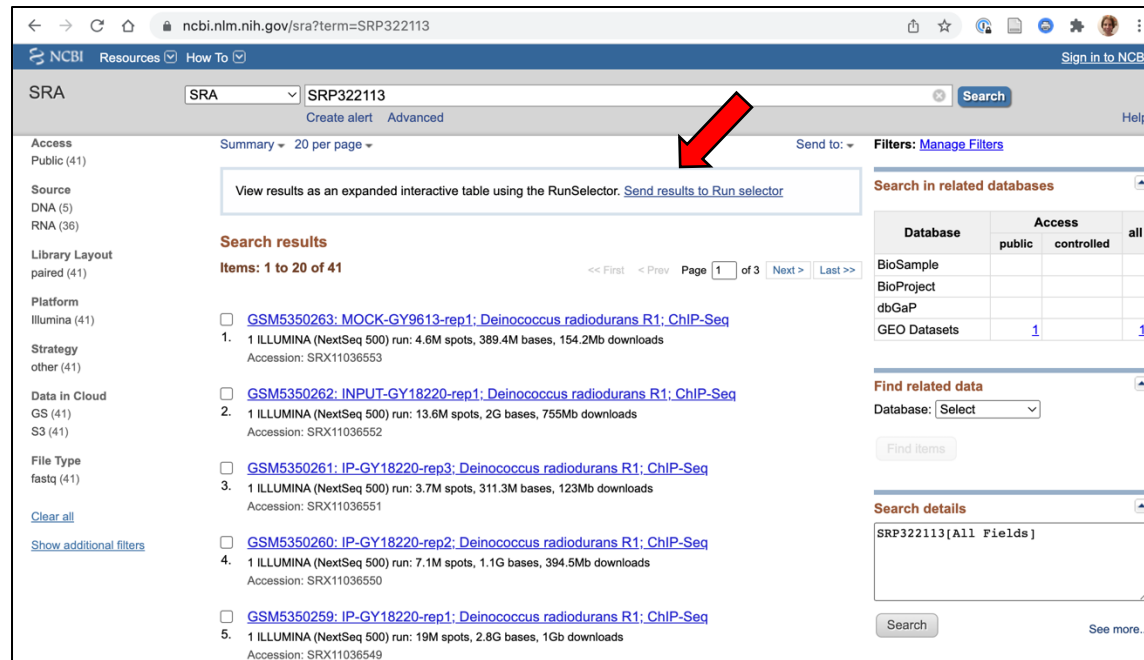
Search details

SRP322113[All Fields]

[Search](#) [See more...](#)

Etape 2 : Importer des fichiers FASTQ (RNAseq)

- Utilisation de l'outil « Run selector » pour choisir les échantillons :



The screenshot shows the NCBI SRA Run Selector interface. The search term 'SRP322113' is entered in the search bar. The results are displayed as a list of 5 items, each with a checkbox and a link to the Run Selector. A red arrow points to the link 'View results as an expanded interactive table using the RunSelector. Send results to Run selector'.

Search results
Items: 1 to 20 of 41

Accession	Run	Platform	Strategy	File Type
GSM5350263	MOCK-GY9613-rep1: Deinococcus radiodurans R1: ChIP-Seq	ILLUMINA (NextSeq 500)	run: 4.6M spots, 389.4M bases, 154.2Mb downloads	Accession: SRX11036553
GSM5350262	INPUT-GY18220-rep1: Deinococcus radiodurans R1: ChIP-Seq	ILLUMINA (NextSeq 500)	run: 13.6M spots, 2G bases, 755Mb downloads	Accession: SRX11036552
GSM5350261	IP-GY18220-rep3: Deinococcus radiodurans R1: ChIP-Seq	ILLUMINA (NextSeq 500)	run: 3.7M spots, 311.3M bases, 123Mb downloads	Accession: SRX11036551
GSM5350260	IP-GY18220-rep2: Deinococcus radiodurans R1: ChIP-Seq	ILLUMINA (NextSeq 500)	run: 7.1M spots, 1.1G bases, 394.5Mb downloads	Accession: SRX11036550
GSM5350259	IP-GY18220-rep1: Deinococcus radiodurans R1: ChIP-Seq	ILLUMINA (NextSeq 500)	run: 19M spots, 2.8G bases, 1Gb downloads	Accession: SRX11036549

Search in related databases

Database	Access	public	controlled	all
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)

- Sélectionner les 3 réplicats 1H, ainsi que les 3 réplicats 4H, exporter un fichier texte avec les identifiants d'accèsion des échantillons :

Common Fields

BioProject	PRJNA734175
Consent	PUBLIC
Center Name	GEO
DATASTORE filetype	FASTQ, SRA
DATASTORE provider	GS, NCBI, S3
DATASTORE region	gs.US, ncbi.public, s3.us-east-1
Instrument	NextSeq 500
LibraryLayout	PAIRED
Organism	Deinococcus radiodurans R1

Select

	Runs	Bytes	Bases	Download	Cloud Data Delivery	Computing
Total	41	28.72 Gb	75.37 G	Metadata or Accession List or JWT Cart		
Selected	6	4.07 Gb	10.65 G	Metadata or Accession List or JWT Cart	Deliver Data	Galaxy

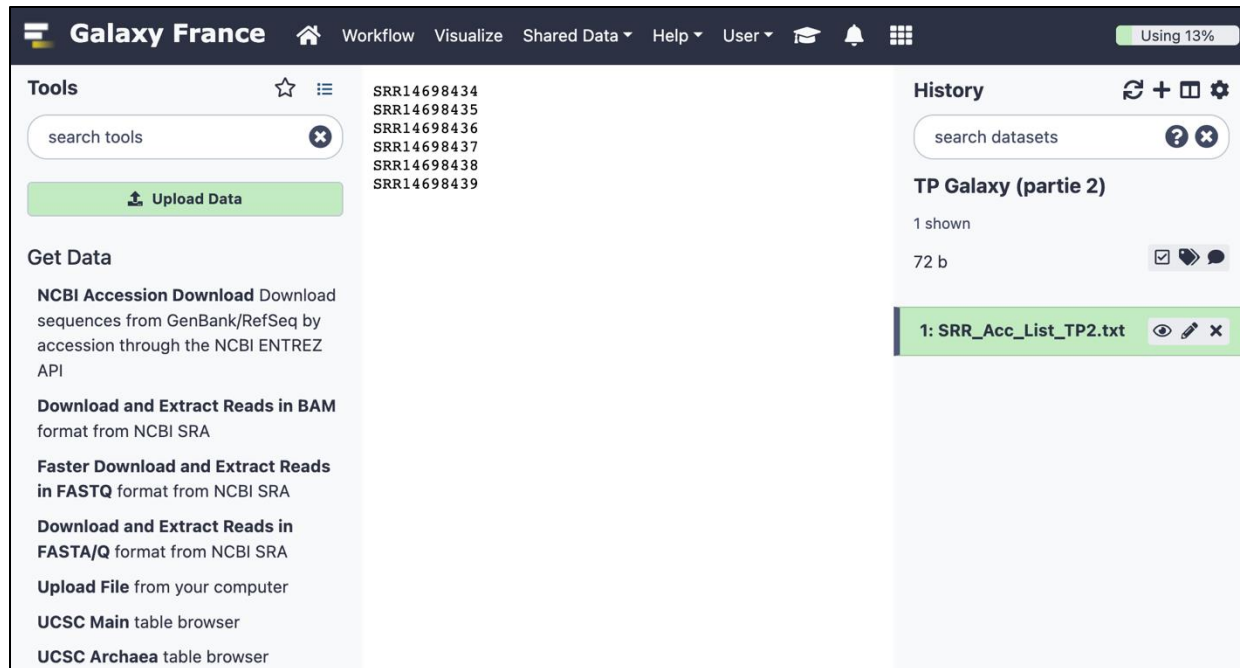
You selected 6 items

	Run	BioSample	Assay Type	AvgSpotLen	Bases	Bytes	Experiment	Genotype	GEO_Accession	LibrarySelection	LibrarySource	Sample Name	source_name	treatment	time_point
☒	SRR14698434	SAMN19486405	RNA-Seq	84	2.03 G	793.10 Mb	SRX11036513	deltaddr/O/ prepUts:ddrO+	GSM5350223	cDNA	TRANSCRIPTOMIC	GSM5350223	17370_1H-rep1	37degreeC	1H
☒	SRR14698435	SAMN19486404	RNA-Seq	84	1.63 G	636.52 Mb	SRX11036514	deltaddr/O/ prepUts:ddrO+	GSM5350224	cDNA	TRANSCRIPTOMIC	GSM5350224	17370_1H-rep2	37degreeC	1H
☒	SRR14698436	SAMN19486403	RNA-Seq	84	1.78 G	695.57 Mb	SRX11036515	deltaddr/O/ prepUts:ddrO+	GSM5350225	cDNA	TRANSCRIPTOMIC	GSM5350225	17370_1H-rep3	37degreeC	1H
☒	SRR14698437	SAMN19486402	RNA-Seq	84	1.49 G	596.25 Mb	SRX11036516	deltaddr/O/ prepUts:ddrO+	GSM5350226	cDNA	TRANSCRIPTOMIC	GSM5350226	17370_4H-rep1	37degreeC	4H
☒	SRR14698438	SAMN19486401	RNA-Seq	84	1.85 G	722.35 Mb	SRX11036517	deltaddr/O/ prepUts:ddrO+	GSM5350227	cDNA	TRANSCRIPTOMIC	GSM5350227	17370_4H-rep2	37degreeC	4H
☒	SRR14698439	SAMN19486419	RNA-Seq	84	1.86 G	722.39 Mb	SRX11036518	deltaddr/O/ prepUts:ddrO+	GSM5350228	cDNA	TRANSCRIPTOMIC	GSM5350228	17370_4H-rep3	37degreeC	4H

(*) Le génotype de ces échantillons est D37

Etape 2 : Importer des fichiers FASTQ (RNAseq)

- Importer le fichier des identifiants SRA dans l'historique Galaxy :



The screenshot shows the Galaxy France web interface. The top navigation bar includes 'Galaxy France', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'User', and a 'Using 13%' indicator. The left sidebar contains 'Tools' (with a search bar and 'Upload Data' button) and 'Get Data' (with links to NCBI Accession Download, Download and Extract Reads in BAM, Faster Download and Extract Reads in FASTQ, Download and Extract Reads in FASTA/Q, Upload File from your computer, UCSC Main table browser, and UCSC Archaea table browser). The main area displays a list of SRR accession numbers: SRR14698434, SRR14698435, SRR14698436, SRR14698437, SRR14698438, and SRR14698439. The right sidebar shows the 'History' panel with a search bar and a list of datasets. The first dataset, 'TP Galaxy (partie 2)', is expanded, showing '1 shown' and '72 b'. Below this, a file named '1: SRR_Acc_List_TP2.txt' is highlighted in green, and a red arrow points to it.

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

The screenshot displays the Galaxy France web interface. On the left, the 'Tools' panel shows the 'Faster Download and Extract Reads in FASTQ' tool selected under the 'Get Data' category (indicated by red arrow 1). The main panel shows the tool's configuration: 'select input type' is set to 'List of SRA accession, one per line' (indicated by red arrow 2); the 'sra accession list' is set to '1: SRR_Acc_List_TP2.txt' (indicated by red arrow 3); and the 'Execute' button is highlighted (indicated by red arrow 4). The 'History' panel on the right shows the workflow steps: '1: SRR_Acc_List_TP2.txt', '2: Pair-end data (fasterq-dump)', '3: Single-end data (fasterq-dump)', '4: Other data (fasterq-dump)', and '5: fasterq-dump log'.

Temps attente (un peu long...)

The screenshot shows the Galaxy France web interface. The top navigation bar includes links for Workflow, Visualize, Shared Data, Help, User, and a notification bell. A status bar on the right indicates 'Using 11%'. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by a 'Get Data' section with links to NCBI Accession Download, Download and Extract Reads in BAM, Faster Download and Extract Reads in FASTQ, Download and Extract Reads in FASTA/Q, Upload File, UCSC Main table browser, and UCSC Archaea table browser.

The main content area displays a green success message: '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.

The right sidebar shows the 'History' panel with a search bar and a list of datasets. The first dataset is '16: fasterq-dump log'. Below it are '15: Other data (fasterq-dump)' (a list), '14: Single-end data (fasterq-dump)' (a list), and '13: Pair-end data (fasterq-dump)' (a list of pairs). A red arrow points from the text 'Données de type « pair-end »' to the '13: Pair-end data' entry.

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-2>

- Passez ensuite à l'étape suivante !

Etape 3 : Reproduire le TP Galaxy (partie 1)

- Contrôle de la qualité des séquences avec le logiciel FASTQC
- Alignement des séquences sur le génome de référence avec le logiciel BOWTIE2

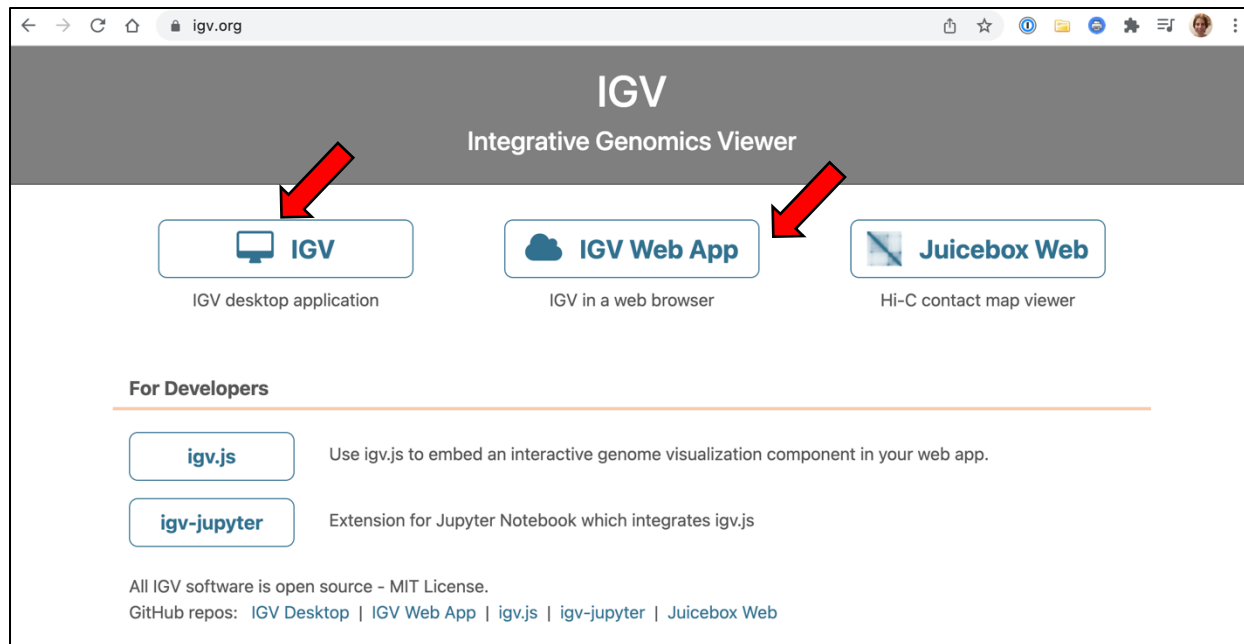
Travaux pratiques – Galaxy (partie 1)

Gaëlle Lelandais et Fabrice Confalonieri

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

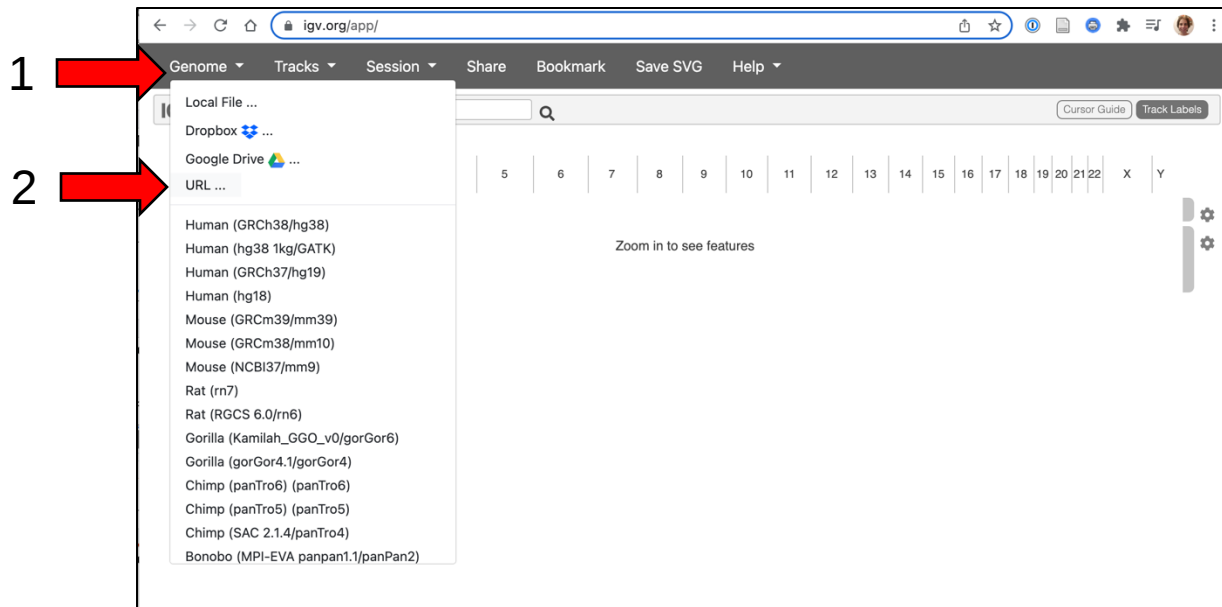
Etape 4 : Visualisation des résultats avec IGV

- Accéder à la page Web de présentation de l'outil. Deux utilisations du logiciel sont possibles, en local ou bien sur un serveur distant.



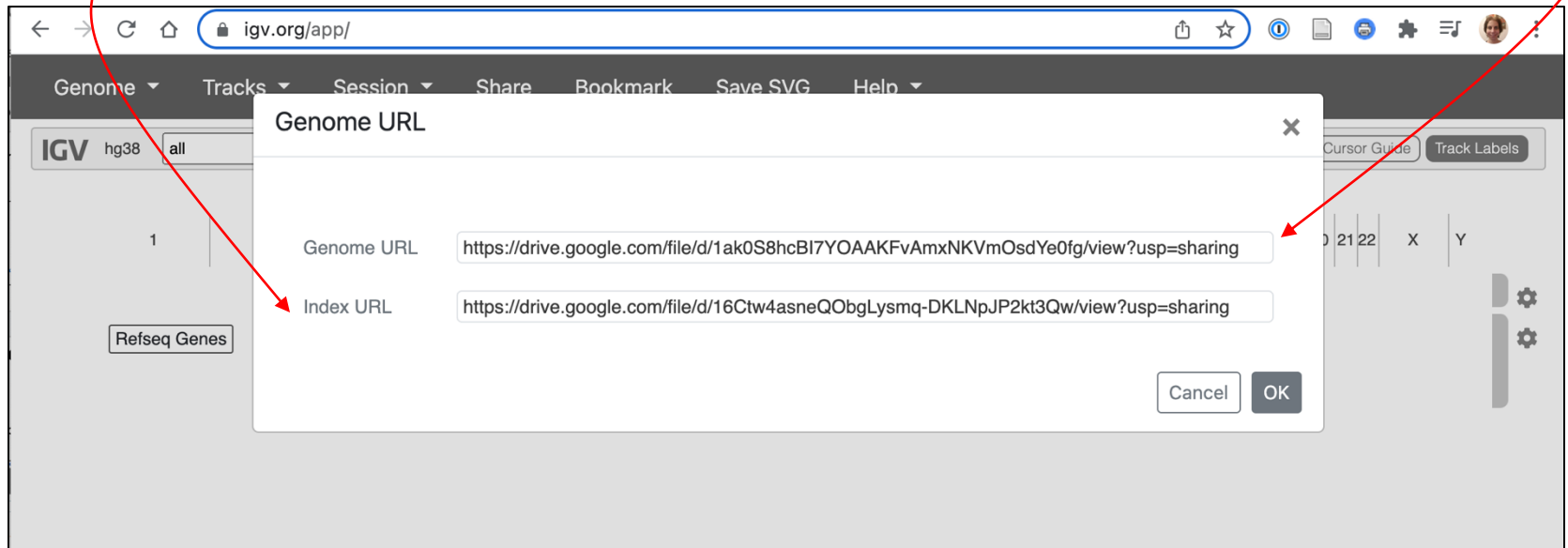
Etape 4 : Visualisation des résultats avec IGV

- Importer le génome de référence (FASTA) dans IGV. Il est conseillé d'utiliser les liens suivants :
 - https://drive.google.com/file/d/1AQ2RI94PoM0J862IOm8IT7pokMe4iku5/view?usp=share_link (Genome link)
 - https://drive.google.com/file/d/174JkfN1Bh2750rPqnhzrtvZUfnmKd8Bb/view?usp=share_link (Index link)



Etape 4 : Visualisation des résultats avec IGV

- Importer le génome de référence (FASTA) dans IGV. Il est conseillé d'utiliser les liens suivants :
 - https://drive.google.com/file/d/1AQ2Rl94PoM0J862lOm8IT7pokMe4iku5/view?usp=share_link (Genome link)
 - https://drive.google.com/file/d/174JkfN1Bh2750rPqnhzrtvZUfnmKd8Bb/view?usp=share_link (Index link)



Etape 4 : Visualisation des résultats avec IGV

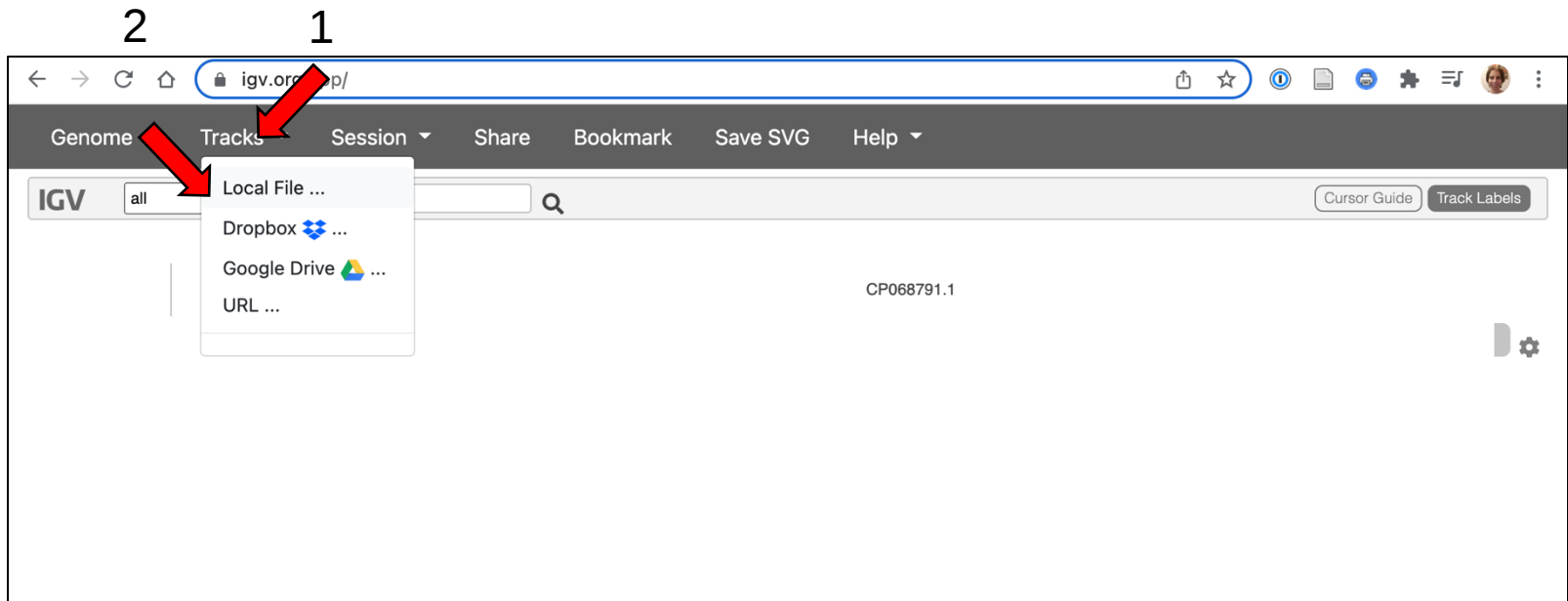
- Télécharger les fichiers BAM et BAM_INDEX, résultats de l'alignement des séquences sur le génome de référence :

The screenshot shows the Galaxy France web interface. On the left is a sidebar with 'Tools' and various categories like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'GENOMIC FILE MANIPULATION', 'Convert Formats', 'FASTA/FASTQ', 'FASTQ Quality Control', 'SAM/BAM', 'BED', 'VCF/BCF', and 'Nanopore'. The main panel displays a table of sequence alignment data with columns: QNAME, FLAG, RNAME, POS, MAPQ, CIGAR, and MRNM. The table lists several alignment records for SRR14698434. On the right is the 'History' panel, which shows a list of datasets. The dataset 'SRR14698434' is selected, and its details are shown, including '1.0 GB' and 'format: bam, database: ?'. Below the details, there are buttons for 'Download', 'load dataset', and 'Download bam_index'. Red arrows and numbers 1, 2, and 3 highlight these buttons: 1 points to the 'Download' button, 2 points to the 'load dataset' button, and 3 points to the 'Download bam_index' button.

QNAME	FLAG	RNAME	POS	MAPQ	CIGAR	MRNM
@HD VN:1.5 SO:coordinate						
@SQ SN:CP068794.1 LN:45508						
@SQ SN:CP068793.1 LN:177322						
@SQ SN:CP068792.1 LN:412138						
@SQ SN:CP068791.1 LN:2644251						
@PG ID:bowtie2 PN:bowtie2 VN:2.4.5 CL:"/shared/ffibstor1/galaxy/mutable-data/dependen						
SRR14698434.751157	99	CP068794.1	1	8	4M2I45M =	
SRR14698434.769218	99	CP068794.1	1	8	4M3I44M =	
SRR14698434.1066036	99	CP068794.1	1	3	4M1I46M =	
SRR14698434.1497896	99	CP068794.1	1	8	4M2I45M =	
SRR14698434.4193949	99	CP068794.1	1	0	4M1I46M =	
SRR14698434.4598291	99	CP068794.1	1	23	4M3I44M =	
SRR14698434.8532678	99	CP068794.1	1	8	4M3I44M =	
SRR14698434.8638104	99	CP068794.1	1	3	4M3I44M =	
SRR14698434.9391477	99	CP068794.1	1	23	4M3I44M =	
SRR14698434.9597271	99	CP068794.1	1	8	4M3I44M =	
SRR14698434.9870353	99	CP068794.1	1	23	4M3I44M =	
SRR14698434.9890462	99	CP068794.1	1	8	4M3I44M =	
SRR14698434.10980062	99	CP068794.1	1	0	4M3I44M =	
SRR14698434.12165596	99	CP068794.1	1	3	4M3I44M =	
SRR14698434.12450692	99	CP068794.1	1	23	4M2I45M =	
SRR14698434.12801713	99	CP068794.1	1	8	4M3I44M =	
SRR14698434.13497802	99	CP068794.1	1	23	4M3I44M =	
SRR14698434.13811826	99	CP068794.1	1	8	4M3I44M =	
SRR14698434.14677147	99	CP068794.1	1	0	4M3I44M =	

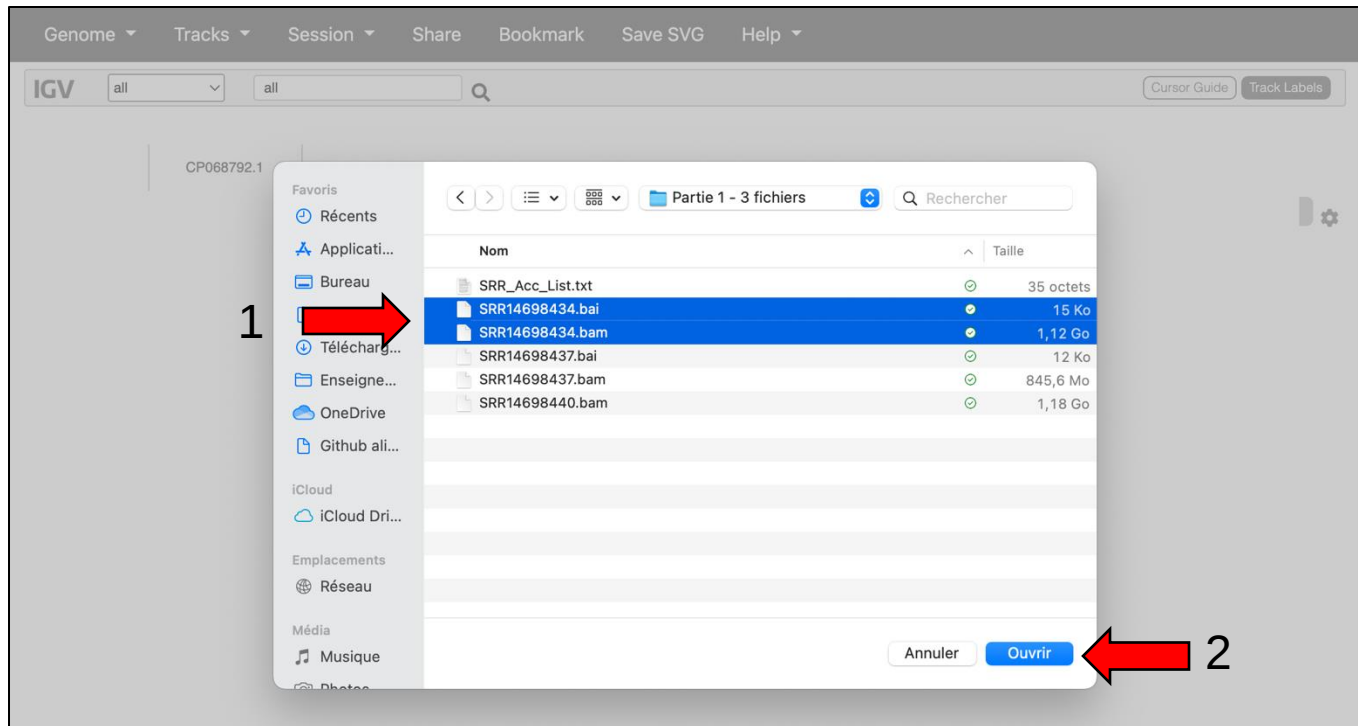
Etape 4 : Visualisation des résultats avec IGV

- Importer les fichiers BAM et BAM_INDEX dans IGV :



Etape 4 : Visualisation des résultats avec IGV

- Importer ensemble les fichiers BAM et BAM_INDEX dans IGV :



Etape 4 : Visualisation des résultats avec IGV

- Et voilà 😊



Conseil : Installation du logiciel IGV en local

- La version Web d'IGV a l'avantage d'exister et d'être gratuite !
- Il lui manque toutefois quelques fonctionnalités utiles pour analyser les résultats.
- Si possible, il est donc recommandé d'installer la version locale d'IGV :
<https://software.broadinstitute.org/software/igv/download>

Fin de la partie 2 😊



Sauf mention contraire, ce contenu est mis à disposition selon les termes de la licence Creative Commons Attribution - Partage dans les mêmes conditions 4.0 International (CC BY-SA 4.0)

Gaëlle LELANDAIS

Version du document : 05/02/2025