

Tools

Options ▾

Your user name: smaman

Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text ManipulationFilter and SortJoin, Subtract and GroupConvert Formats3 - SEQUENCES
MANIPULATIONFASTA manipulationFASTQ manipulationSAM/BAM manipulation : Picard
(beta)SAM/BAM manipulation : SAM
Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel AnalysisRNA-SeqGATK Tools (beta)

6 - SRNASEQ

Analyse des miRNAAnnotationsAlignement sur reference**WELCOME ON SIGENAE GALAXY WORKBENCH**

Galaxy is a workbench available for biologists from Sigena Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biologists.
 - Hide the complexity of the infrastructure.
- Allow creation, execution and sharing of workflows.

History

Options ▾



TP FastQC

54.0 Mb

[8: FastQC_data 5.html](#) [6: GM.fastqsanger](#) [5: h1.fastqsanger](#) [4: FastQC_data
18.html](#) [3: FASTQ Summary
Statistics on data 18](#) [2: FASTQ Summary
Statistics on data 18](#)

76 lines, 1 comments
format: tabular, database: ?
Info: 99115 fastq reads were processed.
Based upon quality values and sequence characters, the input data is valid for: sanger
Input ASCII range: '#'(35) - 'C'(67)
Input decimal range: 2 - 34
Epilog : job finished at ven mai 11 10:36:43 CEST 2012



1	2	3	4	5	6
#column	count	min	max	sum	mean
1	99115	2	33	3194703	32.2
2	99115	2	34	3156652	31.8
3	99115	2	34	3145060	31.7
4	99115	2	34	3120431	31.4
5	99115	2	34	3096075	31.2

Vos traitements bioinformatiques avec GALAXY



<http://sigenae-workbench.toulouse.inra.fr>



Vidéo disponible
sur « sig-learning »

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.



Galaxy Project

Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences

Jeremy Goecks¹, Anton Nekrutenko^{2,3}, James Taylor^{1,3} and The Galaxy Team



Equipe "Galaxy project" :

- Le Center for Comparative Genomics and Bioinformatics - Penn State,
- Des départements "Biology" et "Mathematics and Computer Science" de l'Université d'Emory.



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Anton Nekrutenko
Penn State



Nate Coraor
Penn State



James Taylor
Emory



Une « Galaxy » parmi tant d'autres



Serveur public (<https://main.g2.bx.psu.edu/>):

- Gratuit
- Quota limité : pour se familiariser à l'outil sur des petits jeux de données.
- Données non protégées



Nombreuses autres instances :

- Curie (Nebula)
 - URG
- groupe de travail IFB



Une communauté nationale et internationale très active :

- Listes de diffusion (US, FR)
- Wiki
- Twitter
- "Galaxy tour de France"

L'instance locale Siganae de Galaxy :

- Maintenu par Siganae.
 - Intégration des outils et scripts "locaux".
- **Présentation des particularités de l'instance Siganae.**

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Inutile de savoir :

- ✓ Lancer une ligne de commande
- ✓ Programmer en perl, python, shell ...
- ✓ Lancer un script

Plateforme

Inutile de s'inquiéter pour son disque dur:

- ✓ Jobs lancés sur Genotoul.
- ✓ Pas d'archivage de fichiers sur votre PC.

Vos données

Historique

Workflow

Inutile d'attendre la fin d'un traitement:

- ✓ Possible de lancer plusieurs jobs en parallèle
- ✓ Partir prendre un café, consulter ses mails, ..fermer Internet !
- ✓ Puis voir les résultats le lendemain matin.

Bioinfo

Vous



Galaxy « la bioinformatique pour tous »



Galaxy est :

- « Open source ».
- Développé et maintenu par une communauté active.
- Une plateforme proposant un ensemble d'outils bioinformatiques.
- Accessible : <http://sigenae-workbench.toulouse.inra.fr/>
- Une "constellation" d'outils (analyser, manipuler, visualiser)

Plateforme

Les biologistes peuvent :

Vos données

- Lancer des traitements sans Linux.
- Dupliquer des traitements.
- Partager des analyses complètes.

Et ceci de manière très intuitive !

Historique

Workflow

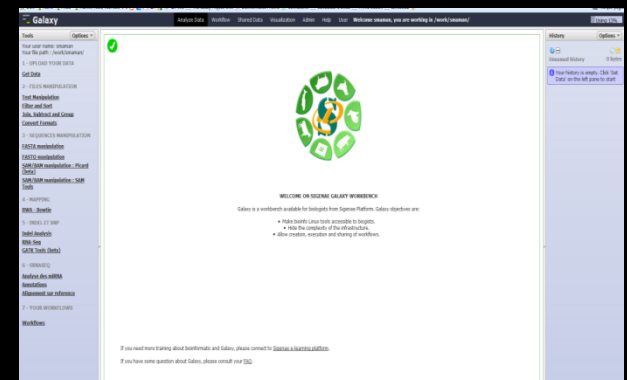
Bioinfo

Vous



Les bioinformaticiens peuvent :

- Faire ajouter des outils.
- Partager des outils (Tool Shed).
- Partager des traitements.



- ✓ Complémentaire au « **cahier de laboratoire** »
→ Retrouver les données, les outils, les références pour la **publication**

- ✓ Manipuler les informations contenues dans un fichier, de façon **simple** et **rapide**.

- ✓ Autres fonctionnalités intéressantes :
 - ✓ "**mapping**" des séquences,
 - ✓ analyse des régions de **variation** ("indel", substitution) ...

- ✓ Construction de **workflow** résumant l'ensemble des fonctionnalités utilisées.

- ✓ Intégration de **nos propres outils** (outils très utiles et fréquemment utilisés)

Plateforme

Vos données

Historique

Workflow

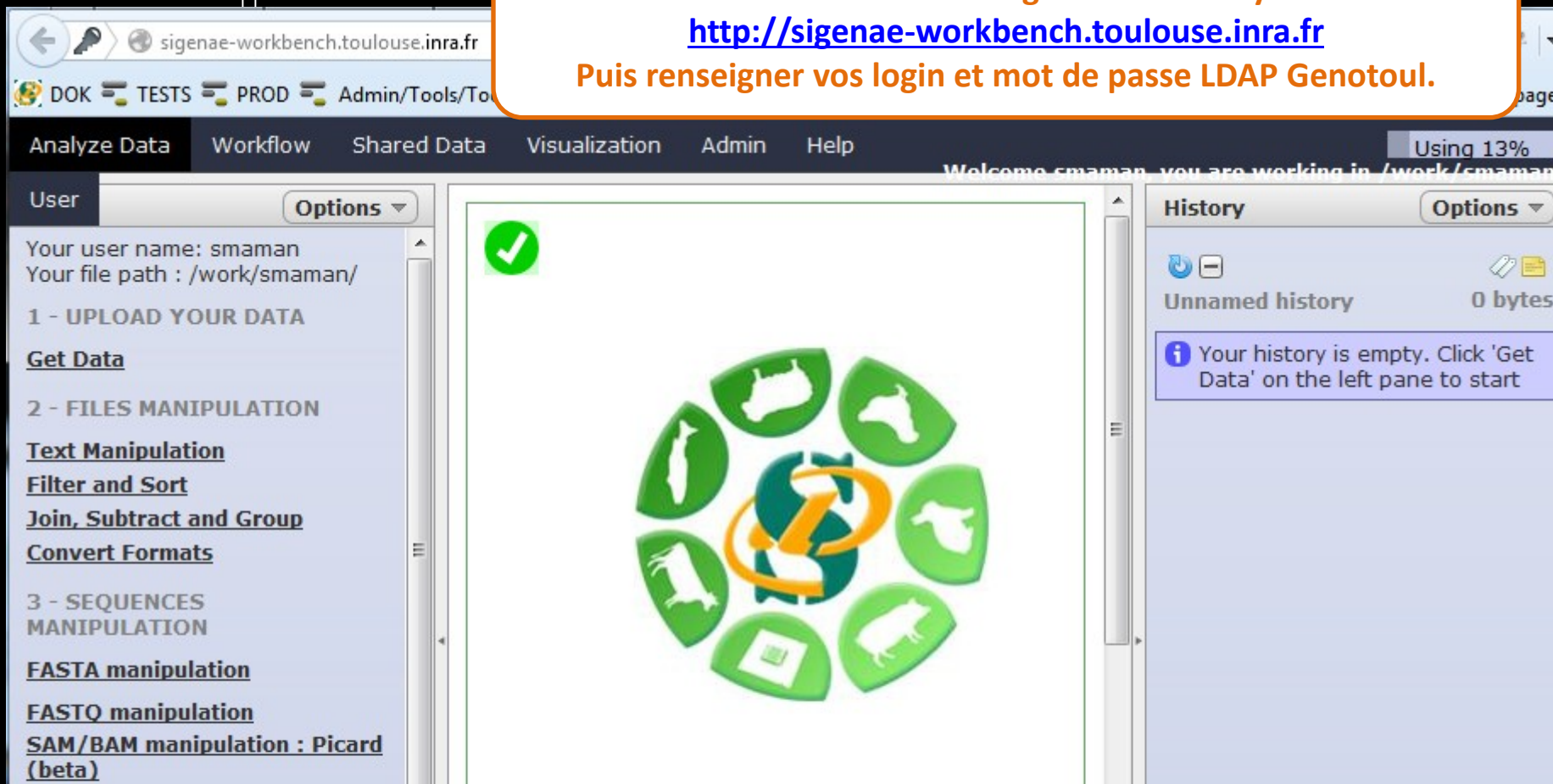
Bioinfo

Vous



→ Galaxy devient **VOTRE BOITE A OUTILS**.

Accès à l'instance Sigenae de Galaxy :
<http://sigenae-workbench.toulouse.inra.fr>
 Puis renseigner vos login et mot de passe LDAP Genotoul.



Vos données sont protégées.
 Vos jobs sont envoyés sur le cluster.
 Inutile de savoir programmer
 De nombreux outils bioinformatiques sont intégrés dans Galaxy.

Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats

3 - SEQUENCES MANIPULATION

FASTA manipulation
FASTQ manipulation
SAM/BAM manipulation : Picard (beta)
SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis
RNA-Seq
GATK Tools (beta)

Options

✓



WELCOME ON SIGENAE GALAXY WORKBENCH

Galaxy is a workbench available for biologists from Sigeneae Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biologists.
 - Hide the complexity of the infrastructure.
 - Allow creation, execution and sharing of workflows.

History Options

Unnamed history 0 bytes

i Your history is empty. Click 'Get Data' on the left pane to start



Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Analyze Data Workflow Shared Data Visualization Admin Help

Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats

3 - SEQUENCES MANIPULATION

FASTA manipulation
FASTQ manipulation
SAM/BAM manipulation : Picard (beta)
SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis
RNA-Seq
GATK Tools (beta)

History Options

Unnamed history 0 bytes

WELCOME ON SIGENAE GALAXY WORKBENCH

Galaxy is a workbench available for biologists from Sigeneae Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biologists.
- Hide the complexity of the infrastructure.
- Allow creation, execution and sharing of workflows.

Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats

3 - SEQUENCES MANIPULATION

FASTA manipulation
FASTQ manipulation
SAM/BAM manipulation : Picard (beta)
SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis
RNA-Seq
GATK Tools (beta)

*** Upload local file from filesystem path (version 1.0.0)**

File Name:
phiX174_read

File type:
Fastq

Path to file:
/work/smaman/phiX174_reads.fastqsanger

Execute

History Options

Unnamed history 0 bytes

Your history is empty. Click 'Get Data' on the left pane to start

Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Analyze Data Workflow Shared Data Visualization Admin Help

Using 13%

Welcome smaman - you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats

3 - SEQUENCES MANIPULATION

FASTA manipulation
FASTQ manipulation
SAM/BAM manipulation : Picard (beta)
SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis
RNA-Seq
GATK Tools (beta)

* Upload local file from filesystem path (version 1.0.0)

File Name:
phiX174_read

File type:
Fastq

Path to file:
/work/smaman/phiX174_reads.fastqsanger

Execute

History Options

Galaxy sensibilisation - TP 12.1 Mb
2 - BWA and FastQC

14:
phiX174_reads.fastqsanger
1.0 Mb
format: fastqsanger, database: ?

```
@080917-and-080922:5:1:185:82
GATGTTATTTCTTCATTTGGAGGTAACCTTTAT
+
IIIIIIIIIIIIIIIIII<III@FI8A/I0II4I
@080917-and-080922:5:1:1366:223
GTTTCTTCTGCGTCAGTAAGAACGTCAGTGTTCC
```



Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation
Filter and Sort
Join, Subtract and Group

NGS: Mapping

- Lastz map short reads against reference sequence
- Lastz paired reads map short paired reads against reference sequence
- Map with Bowtie for Illumina
- Map with Bowtie for SOLiD
- Map with BWA for Illumina

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis
RNA-Seq
GATK Tools (beta)

History Options

Unnamed history 0 bytes

i Your history is empty. Click 'Get Data' on the left pane to start

✓



WELCOME ON SIGENAE GALAXY WORKBENCH

Galaxy is a workbench available for biologists from Sigenae Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biologists.
 - Hide the complexity of the infrastructure.
 - Allow creation, execution and sharing of workflows.

Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Analyze Data Workflow Shared Data Visualization Admin Help

Using 13%

Welcome smaman - you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

NGS: Mapping

- Lastz map short reads against reference sequence
- Lastz paired reads map short paired reads against reference sequence
- Map with Bowtie for Illumina
- Map with Bowtie for SOLiD
- Map with BWA for Illumina

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

Map with BWA for Illumina (version 1.2.2)

Will you select a reference genome from your history?

Use one from the history ▼

Select a reference from history:

11: phiX174_genome.fa ▼

Is this library mate-paired?:

Single-end ▼

FASTQ file:

14: phiX174_reads.fastqsanger ▼

FASTQ with either Sanger-scaled quality values (1-40) or Phred-scaled quality values (0-40)

History Options

Unnamed history 0 bytes

ⓘ Your history is empty. Click 'Get Data' on the left pane to start



Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Analyze Data Workflow Shared Data Visualization Admin Help

Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

NGS: Mapping

- Lastz map short reads against reference sequence
- Lastz paired reads map short paired reads against reference sequence
- Map with Bowtie for Illumina
- Map with Bowtie for SOLiD
- Map with BWA for Illumina

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

Map with BWA for Illumina (version 1.2.2)

Will you select a reference genome from your history?

Use one from the history

Select a reference from history:

11: phiX174_genome.fa

Is this library mate-paired?:

Single-end

FASTQ file:

14: phiX174_reads.fastqsanger

FASTQ with either Sanger-scaled quality values (I

History Options

Unnamed history 0 bytes

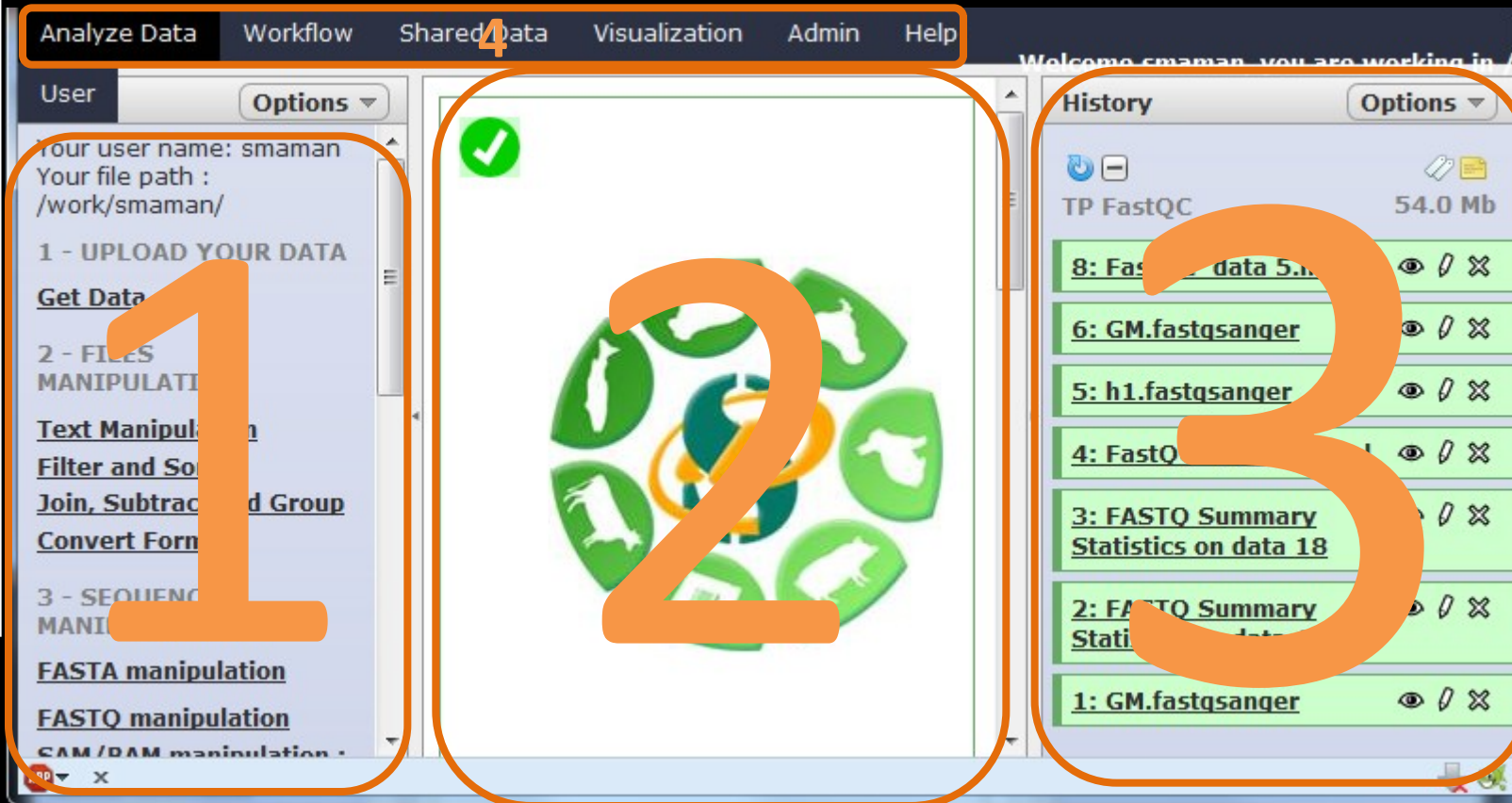
15: Map with BWA for Illumina on data 14 and data 11: mapped reads

Job is waiting to run

Interface simplifiée

Interface divisée en 4 parties :

- 1 - Liste des outils disponibles.
- 2 - Visualisation de l'outil utilisé, historique ou workflow en construction.
- 3 - Historique ou workflow détaillé.
- 4 - Menu .



The screenshot shows the INRA web interface with the following components highlighted by numbered orange callouts:

- 1**: A sidebar menu on the left containing options like 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', and 'Help'. Below this is a 'User' section with fields for 'Your user name: smaman' and 'Your file path: /work/smaman/'.
- 2**: The main content area, which is currently displaying a circular graphic with green and blue segments and a central orange arrow.
- 3**: A 'History' panel on the right side, showing a list of recent actions with details like 'TP FastQC', '8: FastQC data 5.1', '6: GM.fastqsanger', '5: h1.fastqsanger', '4: FastQC', '3: FASTQ Summary Statistics on data 18', '2: FASTQ Summary Statistics on data 18', and '1: GM.fastqsanger'.
- 4**: The top navigation bar, which includes the same set of menu items as the sidebar: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', and 'Help'.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Le vocabulaire spécifique à Galaxy

- TOOL** : Outil bioinformatique ou de traitement de fichiers.
- DATASET** : Fichier de données téléchargé dans Galaxy (fichier entrant) ou fichier généré par Galaxy (fichier résultat).
- HISTORY** : Liste des datasets (entrants et résultants) générés par les tools.
- WORKFLOW** : Schématisation des traitements.

■ [Upload File from your computer](#)

TOOL

génère

1: GM.fastqsanger




17.4 Mb
format: fastqsanger, database: ?





```

@HWUSI-EAS627_1:8:1:17:202
TGGTTGGAGACCCCGAGATGGTTCCTTCAGCTCC
+
BB@=A<9@?AB??=?B?B?B33>?A@?><1:9=A
@HWUSI-EAS627_1:8:1:66:1050
CAGAAGTAGAGCAGAAGAAGACGACCTCCG
    
```

forme

History Options ▾



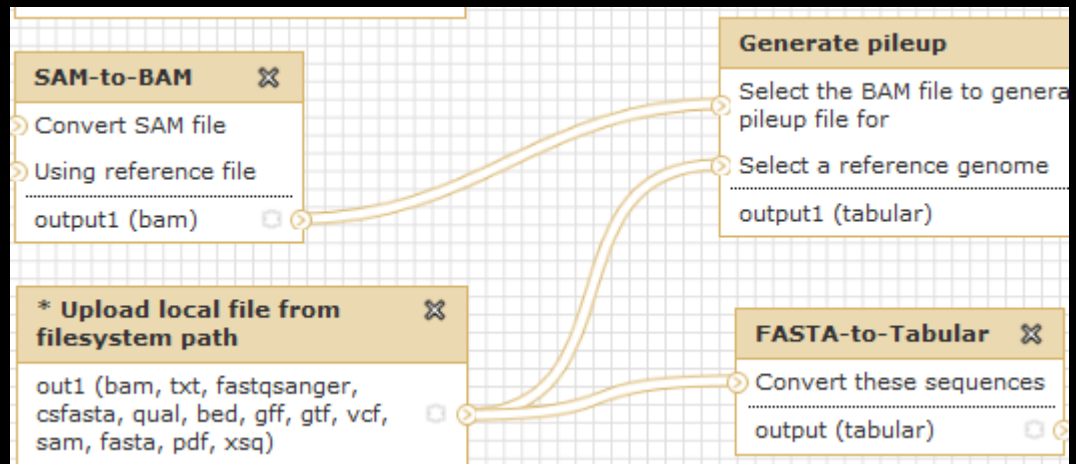

TP FastQC 54.0 Mb

- 8: FastQC data 5.html   
- 6: GM.fastqsanger   
- 5: h1.fastqsanger   
- 4: FastQC data 18.html   

DATASET (S)

HISTORY

WORKFLOW



Plateforme

Vos données

Historique

Workflow

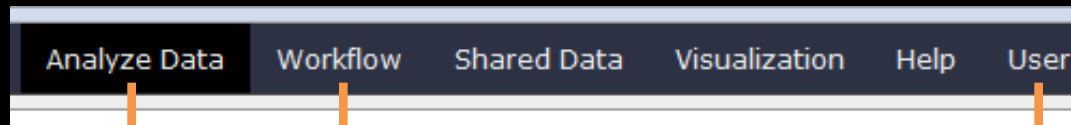
Bioinfo

Vous



Principaux onglets

- **ANALYSE DATA** : Page d'accueil de Galaxy.
- **WORKFLOW** : Liste des workflows .
- **SHARED DATA** : Liste des datasets, historiques et workflows partagés.
- **VISUALIZATION** : Outil de visualisation de vos fichiers résultats.
- **USER** : Accès à vos historiques et datasets sauvegardés.



Plateforme

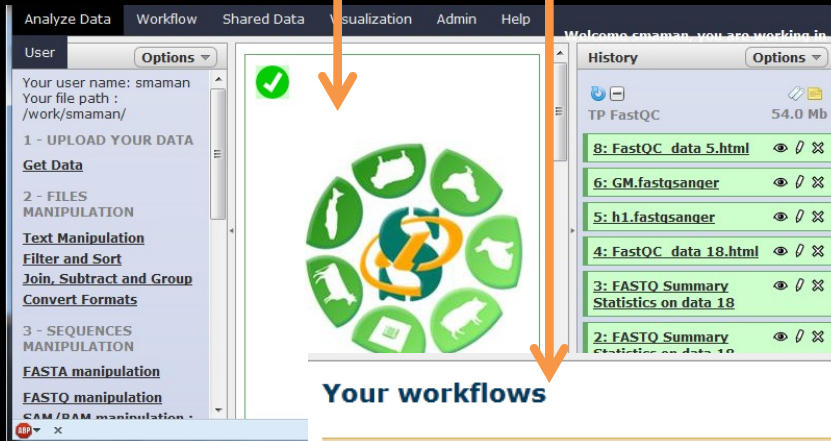
Vos données

Historique

Workflow

Bioinfo

Vous



Your workflows

Name
Workflow constructed from history 'Test BWA fichiers Gnome' ▼
Workflow / 'TP : NGS - Polymorphisme' (imported from uploaded file) ▼
Workflow / 'FastQC' (imported from uploaded file) ▼
Workflow / 'IGV bai' (imported from uploaded file) ▼
Workflow 'Get flanks / region promoters' ▼
Workflow 'RNA seq statistics' et FastQC ▼

Saved Histories

search history names and tags

Advanced Search

☐ Name	Datasets
☐ <u>Unnamed history</u> ▼	
☐ <u>Unnamed history</u> ▼	
☐ <u>Unnamed history</u> ▼	
☐ <u>TP FastQC</u> ▼	7

Workflows shared with you by others

No workflows have been shared with you.

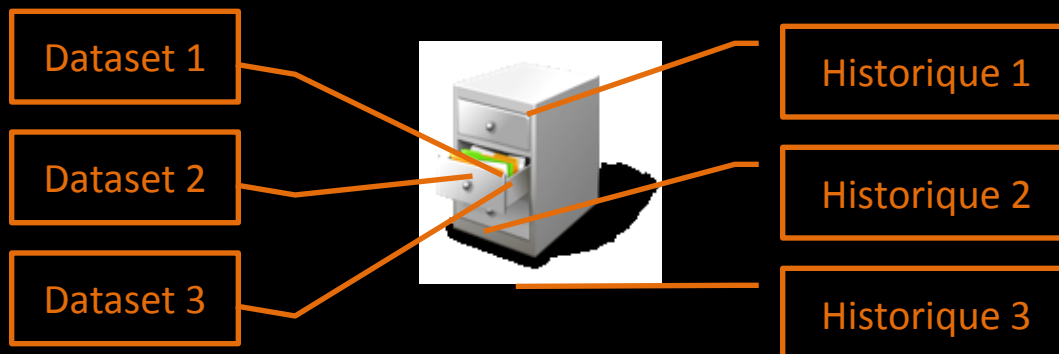


En résumé ...

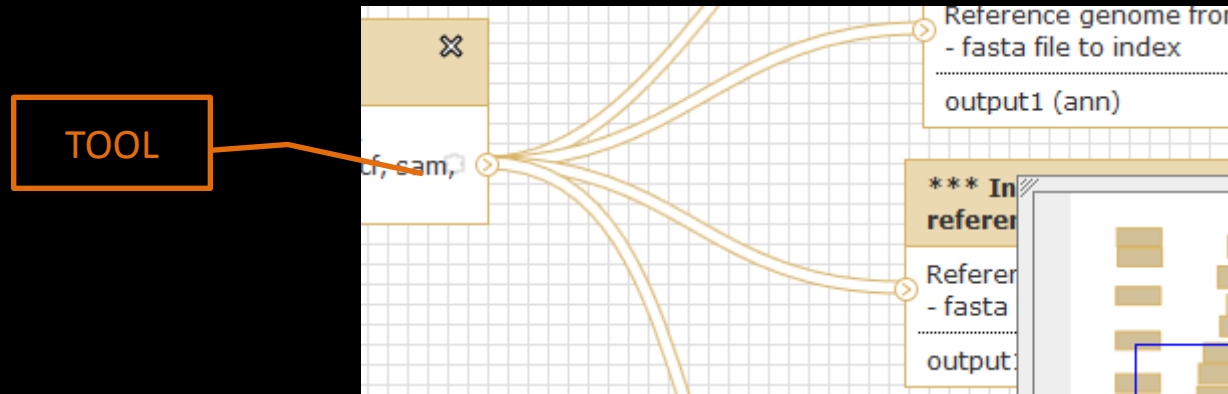
Un **DATASET** est un fichier de données (fichiers d'entrée, fichier résultat) :



Votre **HISTORIQUE** est un « répertoire » qui « liste » l'ensemble de vos fichiers de données (fichiers d'entrée, fichier résultat) utilisés ou générés :



Votre **WORKFLOW** est une représentation de vos traitements : outils utilisés, fichiers



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



2 méthodes de téléchargement de vos données privées



Solution 1 (recommandée si vos fichiers sont dans Genotoul) :

« Upload local file from filesystem path ».

Solution 2 :

« Upload file from your computer, with a copy on server ».



L'espace disque utilisé n'est pas celui de votre PC.
Vos traitements ne sont pas limités par la capacité de votre PC.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Téléchargement de données publiques



Données UCSC, Ensembl, NG6, BIOMART :



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Table Browser

clade: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Prediction Tracks track: UCSC Genes add custom tracks

track hubs

table: knownGene describe table schema

region: ☒ genome ☐ ENCODE Pilot regions ☐ position chr21:33031597-33041570 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

output format: BED - browser extensible data Send output to ☒ Galaxy ☐ GREAT

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

Galaxy

Analyze Data Workflow Shared Data Admin Help User ?

Welcome smaman, you are w

Tools Options

Linux Username: smaman
File Path: /work/smaman/galaxy/
Serveur: galaxy

Get Data

- Upload local file from filesystem path Upload data to history without copying on server
- Upload File from your computer
- EBI SRA ENA SRA
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- BX main browser
- Get Microbial Data
- BioMart Central server

EMBL-EBI

Databases Tools Research Training Industry About Us Help

ENA

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation ... [more](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

NEWS AND ANNOUNCEMENTS

CRAM toolkit 0.7 released
7 Mar 2012

CRAM toolkit 0.7 has been released. More information with download, installation and usage instructions are available at: [http://www.ebi.ac.uk/ena/](#)

Text search

Enter search query, for example: BN000065



Dataset

Présentation de la plateforme Galaxy.

Comment récupérer vos données (privées et publiques) ?

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

History

Options

TP FastQC 54.0 Mb

8: FastQC data 5.html

6: GM.fastqsanger

5: h1.fastqsanger

4: FastQC data 18.html

3: FASTQ Summary Statistics on data 18

2: FASTQ Summary Statistics on data 18

76 lines, 1 comments
format: tabular, database: ?
Info: 99115 fastq reads were processed.
Based upon quality values and sequence characters, the input data is valid for: sanger
Input ASCII range: '#'(35) - 'C'(67)
Input decimal range: 2 - 34
Epilog : job finished at ven mai 11 10:36:43 CEST 2012

1	2	3	4	5	6
#column	count	min	max	sum	mean
1	99115	2	33	3194703	32.2
2	99115	2	34	3156652	31.8
3	99115	2	34	3145060	31.7

• Conserver toutes les étapes de vos analyses .

• Partager vos analyses.

• A chaque run d'un outil, une nouvelle dataset est créée. Les données ne sont pas écrasées.

• Répéter, autant de fois que nécessaire, une analyse.

☐	SwanPorc ▾	18		0 Tags	Shared	0 bytes	Jul 20
☐	FastQC ▾	6		0 Tags	Shared	17.4 Mb	Apr 20
☐	TP : NGS - Polymorphisme ▾	8	2	0 Tags	Shared	6.6 Gb	Apr 20
☐	TP FastQC ▾	12	16	0 Tags		54.0 Mb	Mar 20
☐	indexation genome ▾	1		0 Tags		46 bytes	Jul 20

For 0 selected histories: Rename Delete Delete Permanently

Historiques sauvegardés et historiques partagés



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Browser address: sigenae-workbench.toulouse.inra.fr/galaxy/history/list_published

Galaxy Sig | Analyze Data | Workflow | **Shared Data** | Visualization | Admin | Help | User | Welcome smaman

Published Histories

search name, annotation, owner, and tags

[Advanced Search](#)

Name	Annotation	Community Rating	Community Tags	Last Updated
RNAseq		★★★★★		1 day ago
TP 1 suite : region promotrices	smaman	★★★★★		2 days ago
TP2 SNP - Mai 2013 - Complet	orue	★★★★★		2 days ago
TP2 SNP - 28 mai 2013 - correction OK	bardou	★★★★★		3 days ago
TP 2 sRNAseq 29 mai 2013	smaman	★★★★★		4 days ago
TP2 SNP - 28 mai 2013	bardou	★★★★★		4 days ago

- Data Libraries
- Published Histories**
- Published Workflows
- Published Visualizations

Workflow | Shared Data | Visualization | Admin | Help | **User** | Welcome smaman | Using 30%

Logged in as smaman@toulouse.inra.fr

Logout

[Saved Histories](#)

[Saved Datasets](#)

[Public Name](#)

Saved Histories

search history names and tags

[Advanced Search](#)

☐ Name	Datasets	Tags
☐ TP Galaxy project	2	1
☐ miRNA tests	59	21

2: UCSC Main on Human: snp137Common (chr22:1-51304566)

Comment lancer un job sans ligne de commande ?



1 – Télécharger vos données :

* Upload local file from filesystem path (version 1.0.0)

File Name:

File type:
Fastq

Path to file:

Galaxy sensibilisation - TP 12.1 Mb
2 - BWA and FastQC

14: **phiX174_reads.fastqsanger**
1.0 Mb
format: fastqsanger, database: 2

```
@080917-and-080922:5:1:185:82
GATGTTATTTCTTCATTGGAGGTAACCTCTTAT
+
IIIIIIIIIIIIIIIIII<III@FI8A/I0II4I
@080917-and-080922:5:1:1366:223
GTTTCTTCTGCGTCAGTAAGAAGCTCAGTGTTC
```

11: **phiX174_genome.fa**
1 sequences
format: fasta, database: 2
Info: uploaded fasta file sur :
ftp://ftp.gmod.org/pub/gmod/Courses/2010/SummerSchoolAmericas/Galaxy/phiX174_genome.fa

```
>phix
GAGTTTATCGCTTCATGACGCAGAAAGTTAACTT
AAATTATCTTGATAAGCAGGAATTACTACTGCTTGT
TGCTGGCGGAAATGAGAAAATTCGACCTATCCTTGC
GCGACCTTTCGCCATCACTAACGATTCTGTCAAAAA
TGGCTTAATATGCTTGGCAGCTTCGTCAAGGACTGGT
```

2 – Choisir un outil dans « Tools » :

NGS: Mapping

- [Lastz map short reads](#) against reference sequence
- [Lastz paired reads map](#) short paired reads against reference sequence
- [Map with Bowtie for Illumina](#)
- [Map with Bowtie for SOLiD](#)
- [Map with BWA for Illumina](#)

Map with BWA for Illumina (version 1.2.2)

Will you select a reference genome from your history?
Use one from the history

Select a reference from history:
11: phiX174_genome.fa

Is this library mate-paired?:
Single-end

FASTQ file:
14: phiX174_reads.fastqsanger

FASTQ with either Sanger-scaled quality values (

3 – Lancer le job en cliquant sur « Executer ».

L'Execution du job en cours est visible dans votre historique.
Fini les lignes de commande !

15: **Map with BWA for Illumina on data 14 and data 11: mapped reads**
Job is waiting to run



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Comment suivre l'exécution de mes jobs



Plateforme

Vos données

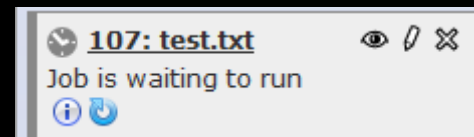
Historique

Workflow

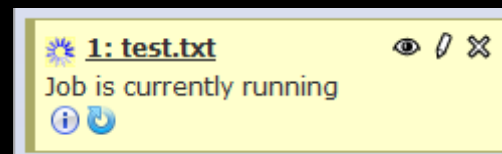
Bioinfo

Vous

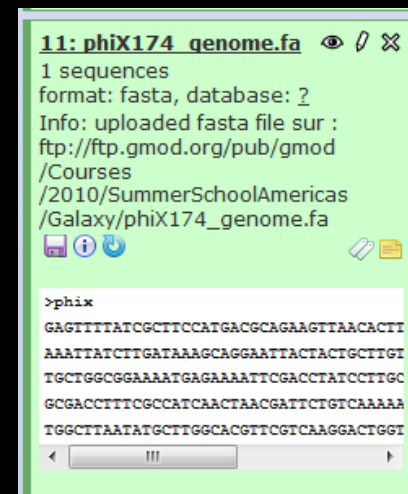
Etat 1 – GRIS : Votre job est en file d'attente.



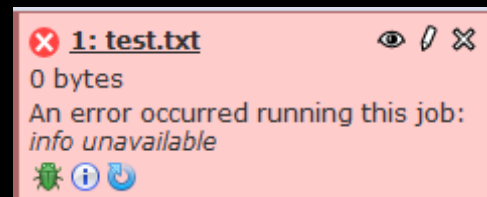
Etat 2 – JAUNE : Votre job est en cours d'exécution



Etat 3 – VERT : Votre job est fini.



Bug - ROUGE : Votre job est planté !



Liste des historiques et exécution de mes jobs



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Galaxy Sig Analyze Data Workflow Shared Data Visualization Admin Help User Welcome smaman Using 30%

Tools Options

1 - UPLOAD YOUR DATA
[Get Data](#)

2 - FILES MANIPULATION
[Text Manipulation](#)
[Filter and Sort](#)
[Join, Subtract and Group](#)
[Convert Formats](#)

3 - SEQUENCES MANIPULATION
[FASTA manipulation](#)
[FASTQ manipulation](#)
[SAM/BAM manipulation : Picard \(beta\)](#)
[SAM/BAM manipulation : SAMtools](#)
[Fetch Sequences](#)

Saved Histories

search history names and tags

Advanced Search

☐ Name	Datasets	Tags	Sharing
☐ TP Galaxy project	2	1	0 Tags
☐ miRNA tests	59	21	0 Tags
☐ TP SNPs calling	84	9	0 Tags
☐ TP RNAseq	88	1	0 Tags
☐ test TP miRNA	36	1	0 Tags
☐ Unnamed history			0 Tags

Analyse OK

Analyse en attente

TP Galaxy project 7.0 Mb

data 2 and data 1
Job is waiting to run

2: UCSC Main on Human: **snp137Common (chr22:1-51304566)**
~180,000 regions
format: bed, database: hg19
view in [GeneTrack](#)
display at Ensembl
Current

Analyse en erreur

Analyse en cours



Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Depuis une page blanche, vous pouvez concevoir un workflow.

Aide : les résultats produits sont typés, il n'est donc pas possible de brancher une dataset sur un mauvais tool !

Plateforme

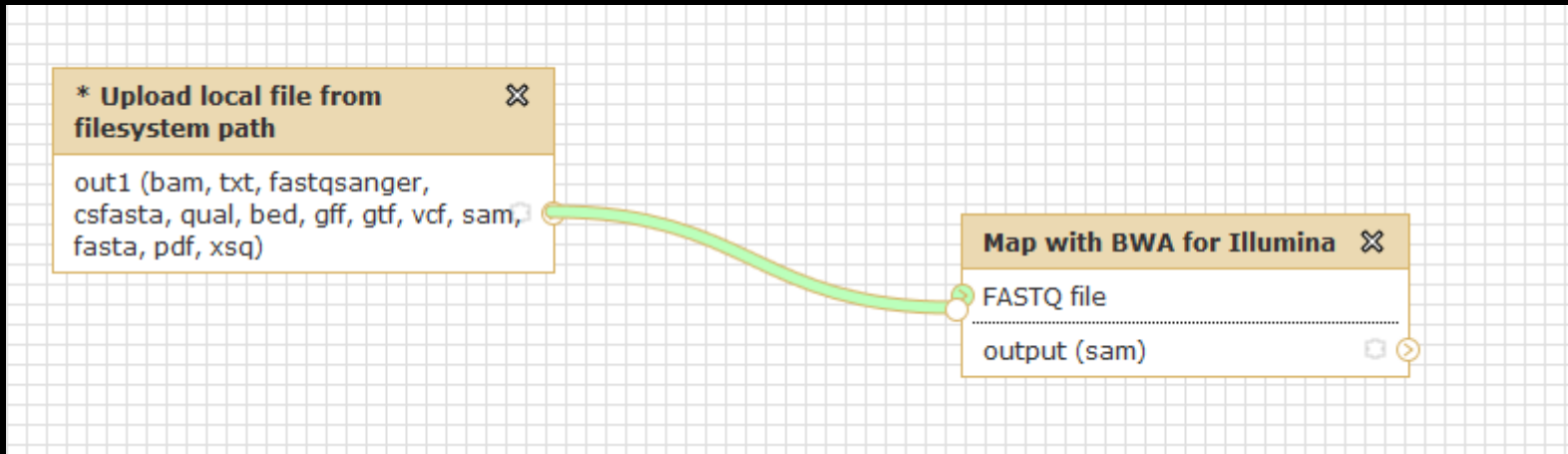
Vos données

Historique

Workflow

Bioinfo

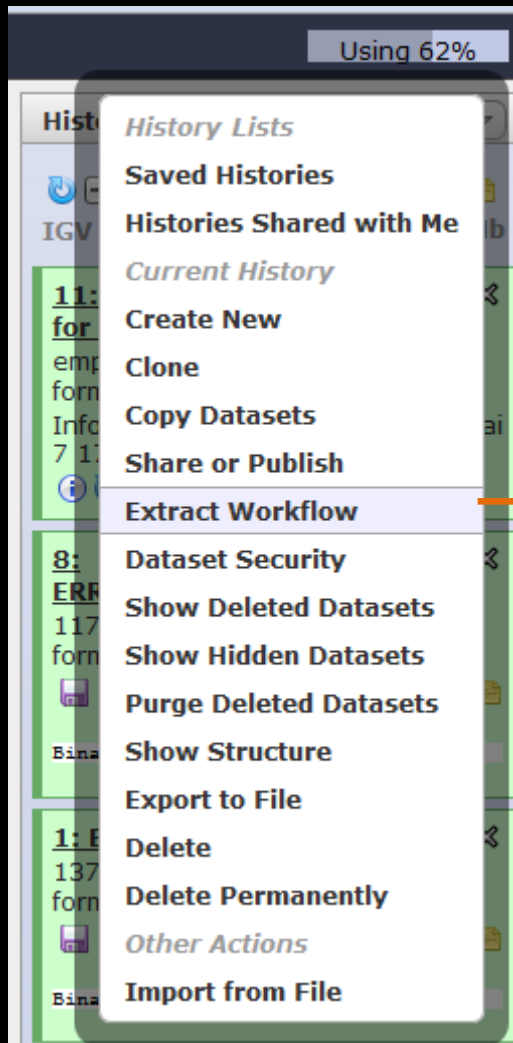
Vous



Exporter votre historique en workflow.



Depuis votre fenêtre « History », vous pouvez extraire un workflow.



Workflow name
Workflow constructed from history 'IGV bai'

Create Workflow Check all Uncheck all

Tool	History items created
* Upload local file from filesystem path ☒ Include "*" Upload local file from filesystem path" in workflow	▶ 1: ERR000017.bam
* Upload local file from filesystem path ☒ Include "*" Upload local file from filesystem path" in workflow	▶ 8: ERR000017.sorte
* BAM sorted to BAI for IGV ☒ Include "*" BAM sorted to BAI for IGV" in workflow	▶ 11: * BAM sorted to

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Lancer plusieurs jobs simultanément

Cliquer sur le menu « Workflow » pour lister vos workflows :



Galaxy Analyze Data **Workflow**

Your workflows

Name
toto ▼
Workflow 'toto' ▼
Workflow 'toto' ▼
Workflow 'toto' ▼

Workflow 'toto' ▼

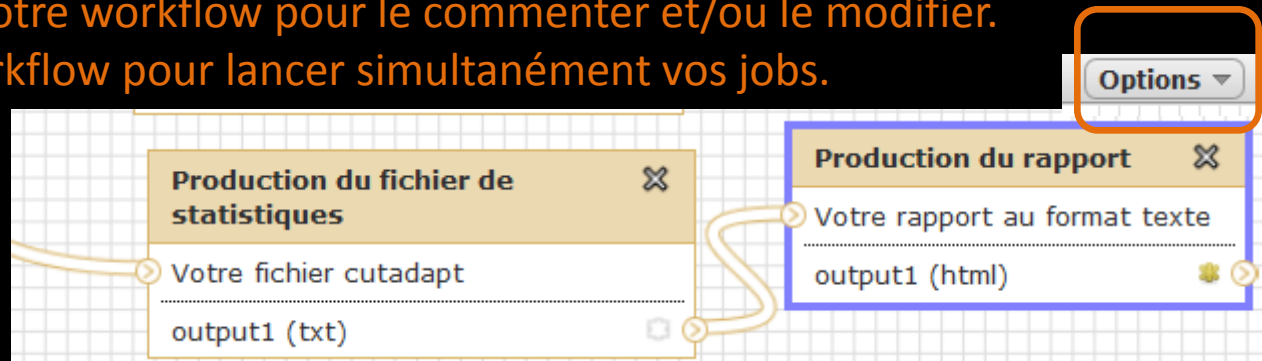
- Edit
- Run
- Share or Publish
- Download or Export
- Clone
- Rename
- View
- Delete

Workflows shared by you by others

Name
Workflow 'Test BWA fichiers Gnome' ▼
Workflow 'Get flanks / region promoters' ▼

Vous pouvez ensuite, depuis le menu « Options », soit :

- Editer votre workflow pour le commenter et/ou le modifier.
- Run workflow pour lancer simultanément vos jobs.



Options ▼

Production du fichier de statistiques ✕

Votre fichier cutadapt
output1 (txt)

Production du rapport ✕

Votre rapport au format texte
output1 (html)

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



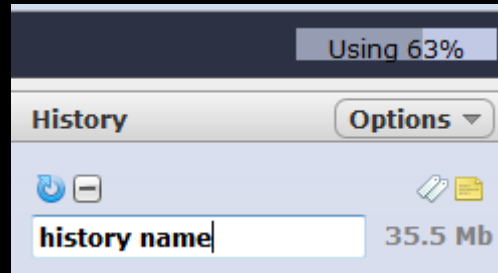
En résumédes datasets aux workflows.



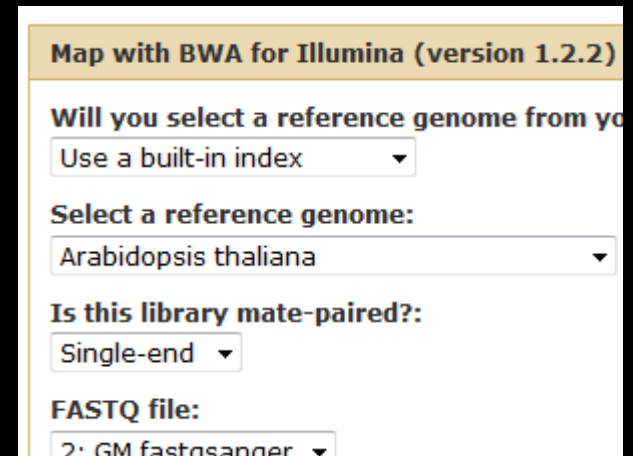
1 – Télécharger vos fichiers de données.

2 – Renommer votre historique.

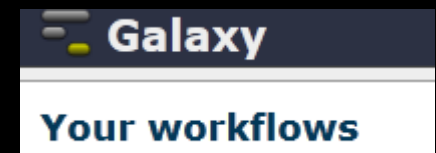
- * Upload local file from filesystem path Upload data to history without copying on server



3 – Utiliser les outils dont vous avez besoin.



4 – Votre historique est sauvegardé automatiquement.
Si nécessaire, exporter votre historique en workflow.



5 – Editer, partager et lancer vos traitements à volonté (run de votre workflow).

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

3 - SEQUENCES MANIPULATION

FASTA manipulation

FASTQ manipulation

SAM/BAM manipulation : Picard (beta)

SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

GET DATA :

Télécharger vos données privées.

Télécharger des données publiques : UCSC, Ensembl, Biomart ...

FILES MANIPULATION :

Manipulation de fichiers texte ou autres : Couper, coller, comparer, soustraire, merger, concatener, sélectionner, filtrer, trier, convertir, grouper ...

SEQUENCE MANIPULATION :

Travailler sur des fichiers FASTA et FastQ, analyse qualité FasQC, Picard tools et samtools.

MAPPING :

BWA, Bowtie, indexation de génome.

Autres :

Recherche d'indel, SNP, RNAseq (TopHat, Cufflinks).

Autres outils en cours de tests et d'ajout .. Selon vos besoins



Ces outils sont nombreux et constituent une bonne alternative à la ligne de commande.

Les traitements sont automatiquement lancés sur GENOTOUL (qsub).

Voici les principaux outils « non bioinfo » proposés :

- Join (des **fichiers lourds**), Substract and Group
- Text Manipulation
- Filter and sort
- Convert Formats



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Select first (version 1.0.0)

Select first:

10

lines

from:

4: UCSC Main on Huma..ne (genome) ▾

Execute

What it does

This tool outputs specified number of lines from the **beginning** of a dataset

Example

Selecting 2 lines from this:

chr7	56632	56652	D17003_CTCF_R6	310	+
chr7	56736	56756	D17003_CTCF_R7	354	+
chr7	56761	56781	D17003_CTCF_R4	220	+
chr7	56772	56792	D17003_CTCF_R7	372	+
chr7	56775	56795	D17003_CTCF_R4	207	+

will produce:

chr7	56632	56652	D17003_CTCF_R6	310	+
chr7	56736	56756	D17003_CTCF_R7	354	+

Outils bioinformatiques

Mapper un FASTQ sur une référence avec BWA.

NGS: Mapping

- Lastz map short reads against reference sequence
- Lastz paired reads map short paired reads against reference sequence
- Map with Bowtie for Illumina
- Map with Bowtie for SOLiD
- Map with BWA for Illumina

Map with BWA for Illumina (version 0.7.17)

Will you select a reference genome?

Use one from the history ▾

Select a reference from history:

29: ERR000017_ref.fasta ▾

Is this library mate-paired?:

Single-end ▾

FASTQ file:

30: ERR003037.fastqsanger ▾

FASTQ with either Sanger-scaled qual

BWA settings to use:

Commonly Used ▾

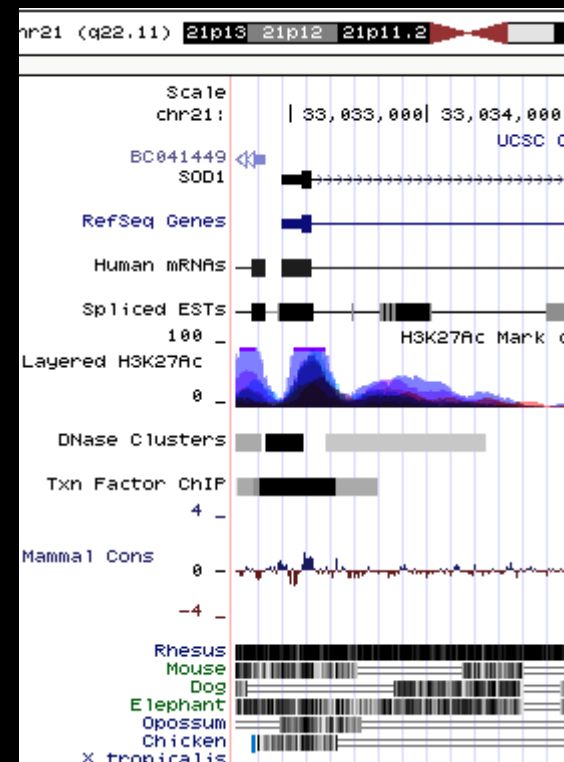
For most mapping needs use Commonly Used

Suppress the header in the output

☐

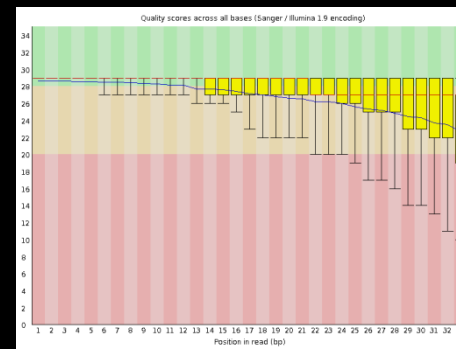
BWA produces SAM with several lines

Execute

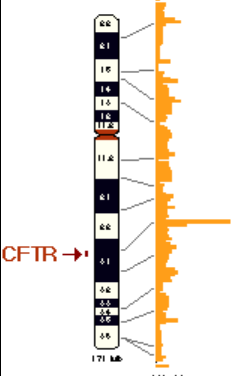


Visualiser la qualité des données avec FASTQC Report.

Visualiser un génome avec UCSC .



The CFTR gene maps to chromosome 7



<http://www.genomeweb.com/2014/04/01/01/>

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



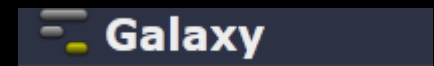
De nombreux outils disponibles :

- Outils de traitement de fichiers
- BWA, FastQC, SAM Tools, Picard Tools ...

Façilité d'ajout de nouveaux scripts / outils selon vos besoins.

Par exemple :

- GATK,
 - Mirdeep2
 - Cutadapt
 - Indexation de génomes
 - Autres ... N'hésitez pas à en faire la demande !
- Mise à jour du menu avec l'ajout d'outils.



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Une FAQ et le lien vers « sig-learning » sont disponibles depuis la page d'accueil.

Shared Data Lab Visualization Admin Help User User **Welcome smaman,**



FAQ on your Galaxy tool

▼ Dataset, history and workflow ?

Step 1 : Import your datasets

First of all, you have to import your data files thanks to "Data Analysis / Get Data" tool. Then your downloaded datasets are automatically archived in "User / Saved Datasets".

Step 2 : Select tools and create your history

Then you select relevant tool in "Data Analysis", on the left side of Galaxy interface.

Vos supports sont disponibles depuis : <http://sig-learning.toulouse.inra.fr>



"If you need more training about bioinformatic and Galaxy, please connect to [Sigenae e-learning platform](#)"

“How to be a good user ?”:

- Maîtrise de son quota
- Organisation de son espace de travail
- Contact: sigenae-support@listes.inra.fr

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Comment gérer son compte ?



Suppression des datasets, historiques et workflows inutiles.
Suppression temporaire / Suppression permanente

Workflow Shared Data Visualization Admin Help User Welcome smaman Using 30%

Workflow	Shared Data	Visualization	Admin	Help	User	Welcome smaman	Using 30%
Chromosome	1	3	0 Tags			328.0	
: fichiers abs du	4		0 Tags			3.6 G	
otation SNP	17		0 Tags			2.4 M	
orted: gnTargetCreator	6	10	0 Tags			3.6 G	
Mirdeep2 sans							
ination de la	51		0 Tags	Accessible		12.8	
ndance intra ni inter							
ory archive	4		0 Tags	Shared		0 byt	
orted: Unnamed	6		0 Tags			742.0	
ry							
iere session	21		0 Tags			1.4 G	
ation Galaxy							
- GALAXY	16		0 Tags			1.6 G	
NGS -	14	6	0 Tags	Shared		0 byt	
morphisme							
NGS RNA Analysis	4	2	0 Tags			41.1	

selected histories: **Rename** **Delete** **Delete Permanently** **Undelet**

History Lists

- Saved Histories
- Histories Shared with Me
- Current History
- Create New
- Clone
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Show Deleted Datasets
- Show Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export to File
- Delete
- Delete Permanently
- Other Actions
- Import from File

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Saved Histories

[Close Advanced Search](#)

name:

tags:

sharing: [private](#) | [shared](#) | [accessible](#) | [published](#) | **all**

status: **active** | [deleted](#) | [all](#)

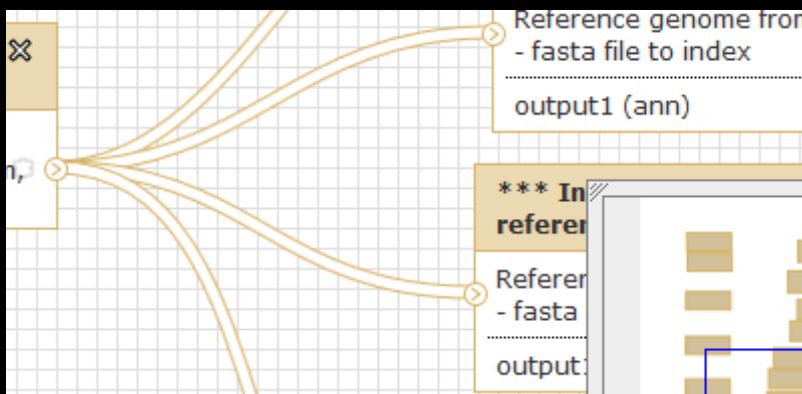
☐ <u>Name</u>	<u>Datasets</u>		<u>Tags</u>
☐ TP Galaxy project ▼	2	1	0 Tags
☐ miRNA tests ▼	59	21	0 Tags
☐ TP SNPs calling ▼	84	9	0 Tags
☐ TP RNAseq ▼	88	1	0 Tags
☐ test TP miRNA ▼	36	1	0 Tags



Comment citer Galaxy dans vos publications ?

Pour vos publications, citer:

- ✓ Les outils utilisés (nom, version).
- ✓ Le workflow généré.
- ✓ Les références « Galaxy project ».



Extract features (version 1.0.0)

Select GFF data:

From:

Column 1 / Sequence name

Extract features:

Multi-select list - hold the appropriate key while clicking to

Execute

What it does

This tool extracts selected features from GFF data.

Primary Publications

If you use or extend Galaxy in your published work, please cite **each** of the following publications:

1. Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 2010 Aug 25;11(8):R86.
2. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. "Galaxy: a web-based genome analysis tool for experimentalists". *Current Protocols in Molecular Biology*. 2010 Jan; Chapter 19:Unit 19.10.1-21.
3. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, "Galaxy: a platform for interactive large-scale genome analysis." *Genome Research*. 2005 Oct; 15(

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



GALAXY

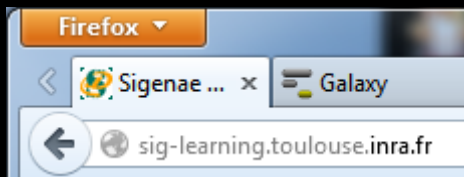
- ✓ Simplicité d'utilisation (sans Linux) pour lancer des calculs sur le cluster.
- ✓ Gérer les résultats et traitements d'une ou plusieurs analyses.
- ✓ Partage de vos datasets, historiques et workflows (intra et inter Galaxy).
- ✓ Présentation schématique de vos traitements grâce aux workflows.
- ✓ Possibilité d'ajout de nouveaux outils selon vos besoins.
(*Outil Siganae avec ligne commande)



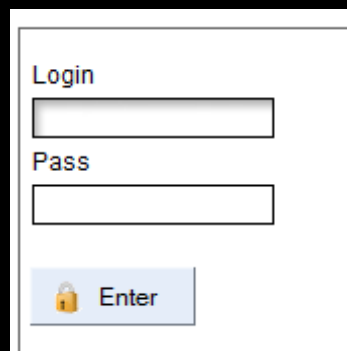
Votre auto-formation
continue en ligne avec
« sig-learning »



- 1 Taper l'adresse de « sig-learning » : <http://sig-learning.toulouse.inra.fr/>



- 2 Authentification



Login

Pass

 Enter

- 3 Accès à la liste des formations auxquelles l'utilisateur s'est inscrit.



Homepage Trainings Profile Agenda Reporting Administration

Trainings

- 1 - Linux & Unix
UNIX1 – SIGENAE Team
- 2 - Cluster (en construction)
CLUSTER – SIGENAE Team
- 3 - Management of large files on Unix and Galaxy
UNIX2 – SIGENAE Team

Plateforme

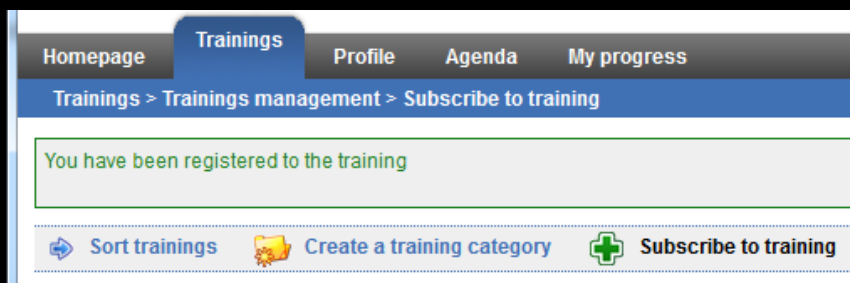
Vos formations



Comment s'inscrire à des formations ?



Il vous est possible de vous inscrire directement en ligne à une formation :
« Trainings » « Trainings management » puis « Subscribe to training » :



Plateforme

Vos
formations

L'inscription s'effectue via une recherche de la formation par mots clés.
Voici donc la liste des formations

(disponibles au 01/2013) :

Trainings	
	1 - Linux & Unix UNIX1 – SIGENAE Team
	2 - Cluster (en construction) CLUSTER – SIGENAE Team
	3 - Management of large files on Unix and Galaxy UNIX2 – SIGENAE Team
	4 - Galaxy GALA01 – SIGENAE Team
	5 - FastQC : quality control tool & reports interpretation FASTQC – SIGENAE Team
	6 - Aligning SGS reads and SNP finding SGS-SNP – SIGENAE Team
	7 - NG6 NG6 – SIGENAE Team
	8 - RNA seq (en construction) RNASEQ – SIGENAE Team
	Demonstration DEMO – SIGENAE Team



Outre une introduction et un carrousel permettant d'accéder aux principaux chapitres de la formation, la page d'accueil de la formation donne accès :



Training plan



Forum



Tests



Links

Plateforme

Vos
formations

TRAINING PLAN : Parcours pédagogique avec les supports en ligne.

FORUM : Support de communication entre stagiaires / formateurs.

TESTS : Tests et exercices.

LINKS : Liens utiles.



Comment ouvrir un compte sur GALAXY Sigenae et sig-learning ?



1 – Demande à compte sur la plateforme BIOINFO GENOTOUL :

<http://bioinfo.genotoul.fr/index.php?id=81>

Ou :

bioinfo.genotoul.fr puis « menu « Help » , puis « Create an account ».

Vous recevrez un login et mot de passe LDAP Genotoul.

2 – Puis utilisez ce login et mot de passe LDAP Genotoul lorsque vous souhaitez accéder à :

.Instance Sigenae de Galaxy : <http://sigenae-workbench.toulouse.inra.fr/>

.« Sig-learning » : <http://sig-learning.toulouse.inra.fr/>

3 - Pour demander une augmentation de votre quota utilisateur sur Galaxy, veuillez vous adresser à :

sigenae-support@listes.inra.fr

Plateforme

Vos
formations



Merci pour votre écoute

Questionnaire

<http://bioinfo.genotoul.fr/index.php?id=130>

