

Galaxy objectives are :

First, making bioinfo Linux tools accessible to biogists.

Then, it is possible to add Linux tools by developpers into Galaxy workbench.

Then, Galaxy is used to hide the complexity of the infrastructure and to allow creation, execution and sharing of workflows.

To access to Galaxy, you need to have an LDAP Genotoul login and password.

L'objectif de la plateforme Galaxy est de rendre les logiciels de bioinformatique sous Linux accessibles aux biologistes. Pour cela, Galaxy permet de masquer la complexité de l'infrastructure et permet la création, le partage et l'exécution de chaînes de traitement.

L'interface web de Galaxy est conviviale et permet la création et la supervision de workflows.

Cet outil est aussi bien destiné aux biologistes et qu'aux développeurs.

Pour vous connecter à la plateforme Galaxy, il vous sera nécessaire de vous authentifier avec votre login et votre mot passe LDAP Genotoul.

GALAXY pour vos traitements bioinformatiques
<http://sigenae-workbench.toulouse.inra.fr>

« sig-learning » pour votre auto-formation continue en ligne
<http://sig-learning.toulouse.inra.fr>



29/01/2013



Vos traitements
bioinformatiques avec
GALAXY





Vidéo disponible
sur « sig-learning »

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



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.

Galaxy Project

Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences
Jeremy Goedge, Anton Nekrutenko, James Taylor, 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)
- URGI

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.**

Navigation:

- Plateforme
- Vos données
- Historique
- Workflow
- Bioinfo
- Vous



Des outils accessibles à tous.. 

Inutile de savoir :

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

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Inutile de s'inquiéter pour son disque dur:

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

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.



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

Vos données

Historique

Workflow

Bioinfo

Vous



Les biologistes peuvent :

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



Et ceci de manière très intuitive !

Les bioinformaticiens peuvent :

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



Contexte d'utilisation dans un laboratoire

- ✓ Complémentaire au « **cahier de laboratoire** », avec **archivage** des données de séquençage
→ 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)

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



Navigation menu:

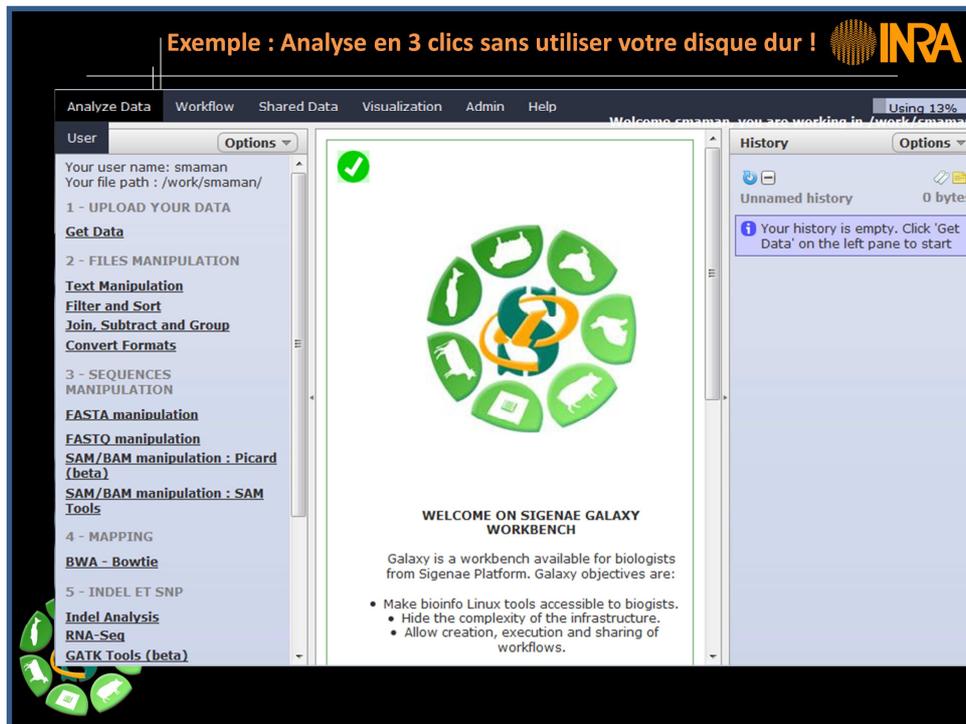
- Plateforme
- Vos données
- Historique
- Workflow
- Bioinfo
- Vous

The image shows a screenshot of the Sigena de Galaxy web interface. At the top, an orange callout box contains the text: "Accès à l'instance Sigena de Galaxy : <http://sigena-workbench.toulouse.inra.fr> Puis renseigner vos login et mot de passe LDAP Genotoul." The interface itself has a navigation bar with "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", and "Help". The main content area is divided into three panes: "User" on the left, a central workspace with a green circular logo, and "History" on the right. The "User" pane shows the user name "smaman" and file path "/work/smaman/". The "History" pane shows "Unnamed history" with "0 bytes" and a message: "Your history is empty. Click 'Get Data' on the left pane to start." At the bottom, another orange callout box contains the text: "Vos données sont protégées. Vos jobs sont envoyés sur le cluster. Inutile de savoir programmer De nombreux outils bioinformatiques sont intégrés dans Galaxy."

Accès à l'instance Sigena de Galaxy :
<http://sigena-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 ! 



Using 13%

Analyze Data Workflow Shared Data Visualization Admin Help

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

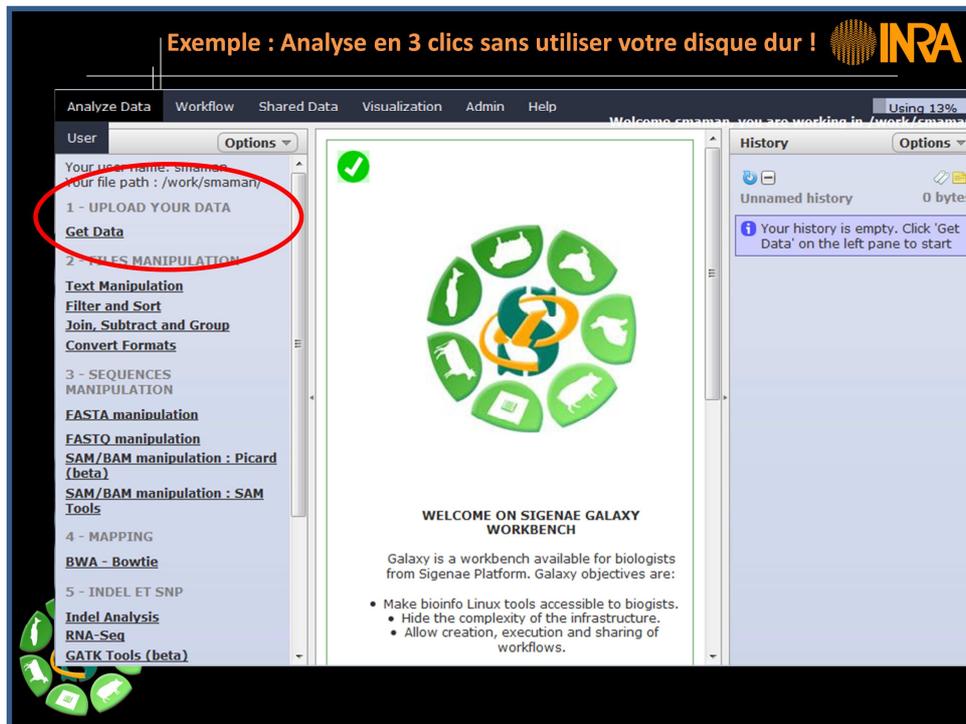
Your history is empty. Click 'Get Data' on the left pane to start

WELCOME ON SIGENAE GALAXY WORKBENCH

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

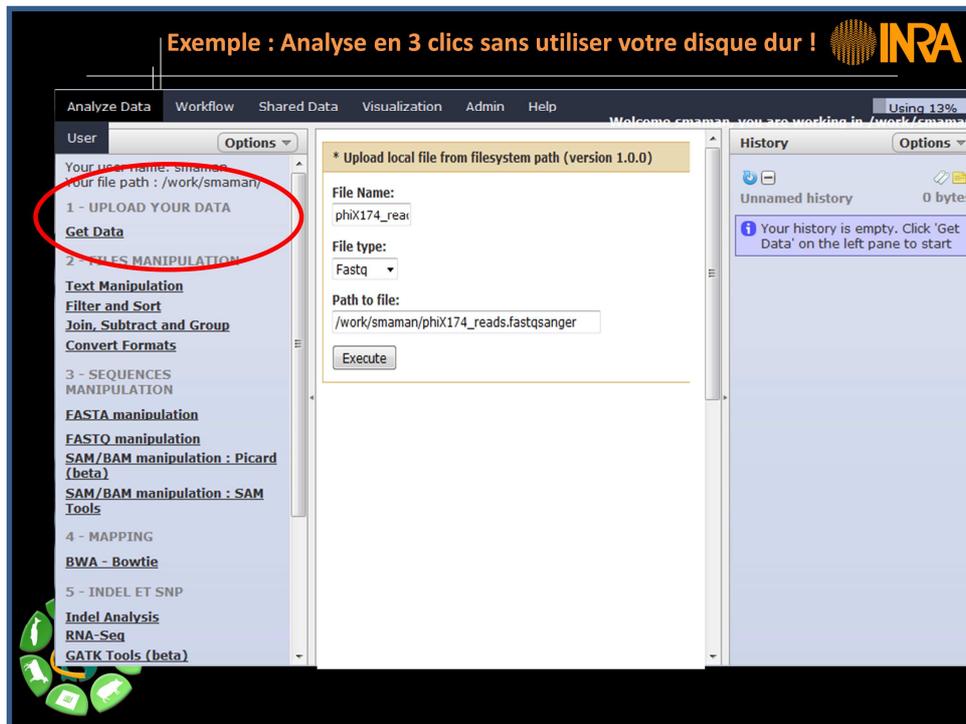
- Make bioinfo Linux tools accessible to biologists.
- Hide the complexity of the infrastructure.
- Allow creation, execution and sharing of workflows.

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



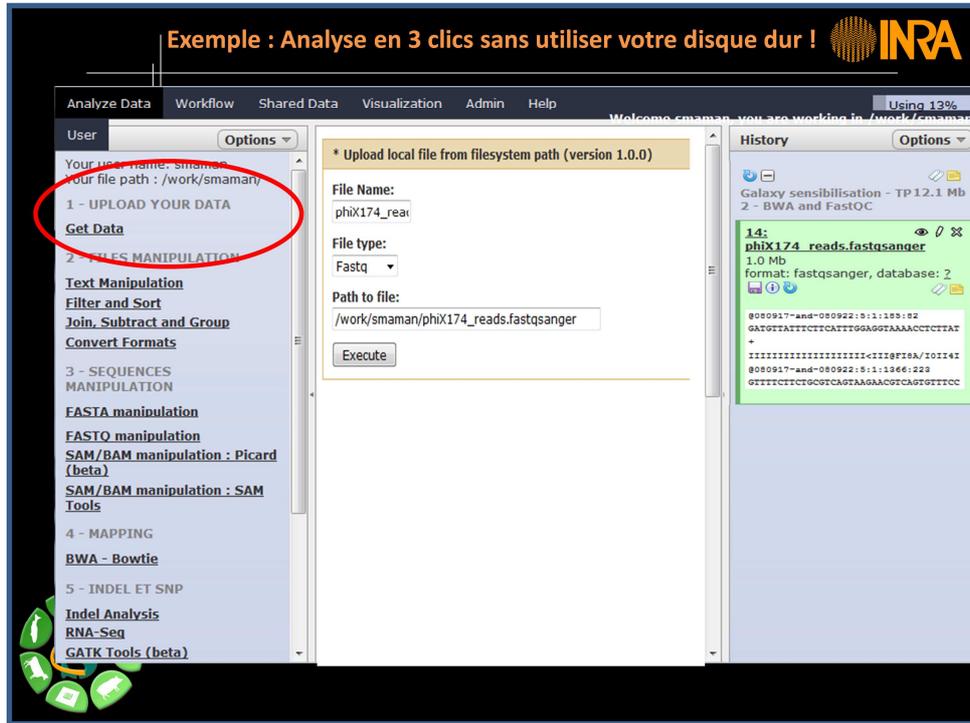
The screenshot shows the Sigena Galaxy Workbench interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', and 'Help'. The left sidebar contains a menu with the following items: 'User', 'Options', 'Your user name: sigena', 'Your file path: /work/smamary', '1 - UPLOAD YOUR DATA', 'Get Data' (circled in red), '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', and 'GATK Tools (beta)'. The main panel displays a green checkmark, a circular logo with icons, and the text 'WELCOME ON SIGENAE GALAXY WORKBENCH'. Below this, it states 'Galaxy is a workbench available for biologists from Sigena Platform. Galaxy objectives are:' followed by a list: '• Make bioinfo Linux tools accessible to biogists.', '• Hide the complexity of the infrastructure.', and '• Allow creation, execution and sharing of workflows.'. The right sidebar shows a 'History' panel with 'Options', 'Unnamed history', '0 bytes', and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.

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



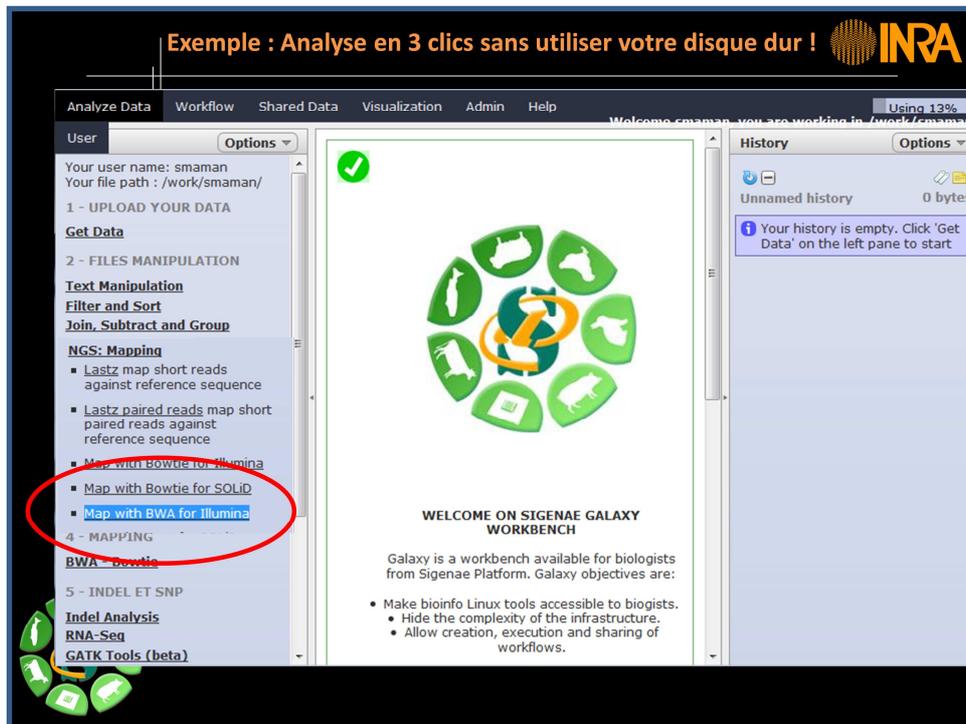
The screenshot displays a software interface with a dark blue header and a light blue sidebar. The sidebar contains a menu with categories: '1 - UPLOAD YOUR DATA', '2 - FILES MANIPULATION', '3 - SEQUENCES MANIPULATION', '4 - MAPPING', and '5 - INDEL ET SNP'. Under '1 - UPLOAD YOUR DATA', the 'Get Data' option is circled in red. The main panel shows a workflow step titled '* Upload local file from filesystem path (version 1.0.0)'. It includes input fields for 'File Name' (phiX174_read), 'File type' (Fastq), and 'Path to file' (/work/smaman/phiX174_reads.fastqsanger), followed by an 'Execute' button. A 'History' panel on the right shows 'Unnamed history' with '0 bytes' and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', and 'Help'. The top right corner shows 'Using 13%' and a 'Welcome smaman, you are working in /work/smaman' message.

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



The screenshot displays the Galaxy web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', and 'Help'. The main content area is titled '* Upload local file from filesystem path (version 1.0.0)'. It features a 'File Name' field with the value 'phiX174_read', a 'File type' dropdown set to 'Fastq', and a 'Path to file' field with the value '/work/smaman/phiX174_reads.fastqsanger'. An 'Execute' button is located below these fields. On the left, a sidebar menu lists various tools, with '1 - UPLOAD YOUR DATA' and 'Get Data' circled in red. The right sidebar shows a 'History' panel with a job entry for '14: phiX174_reads.fastqsanger' (1.0 Mb) in 'fastqsanger, database: 2' format, displaying a preview of sequencing reads.

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



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

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)

WELCOME ON SIGENAE GALAXY WORKBENCH

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

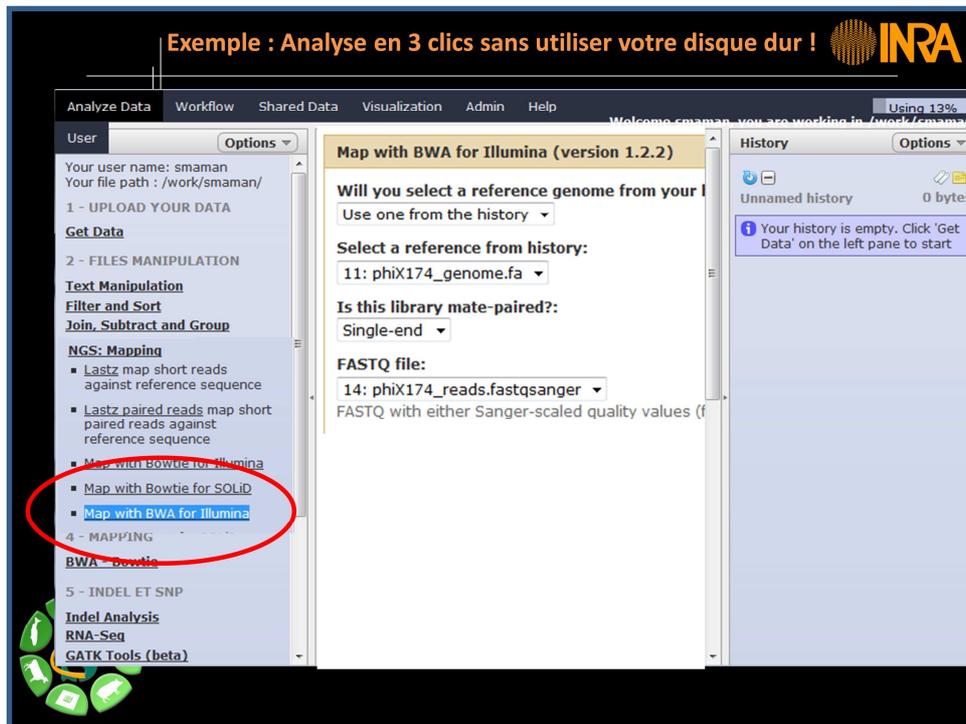
- Make bioinfo Linux tools accessible to biogists.
- Hide the complexity of the infrastructure.
- Allow creation, execution and sharing of workflows.

History Options

Unnamed history 0 bytes

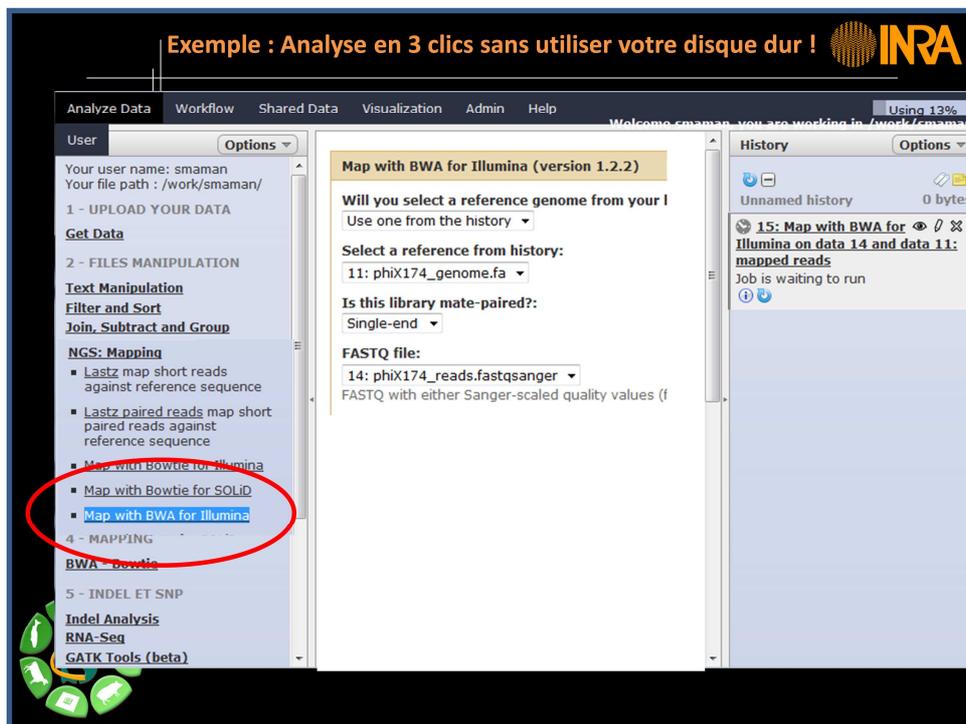
Your history is empty. Click 'Get Data' on the left pane to start

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

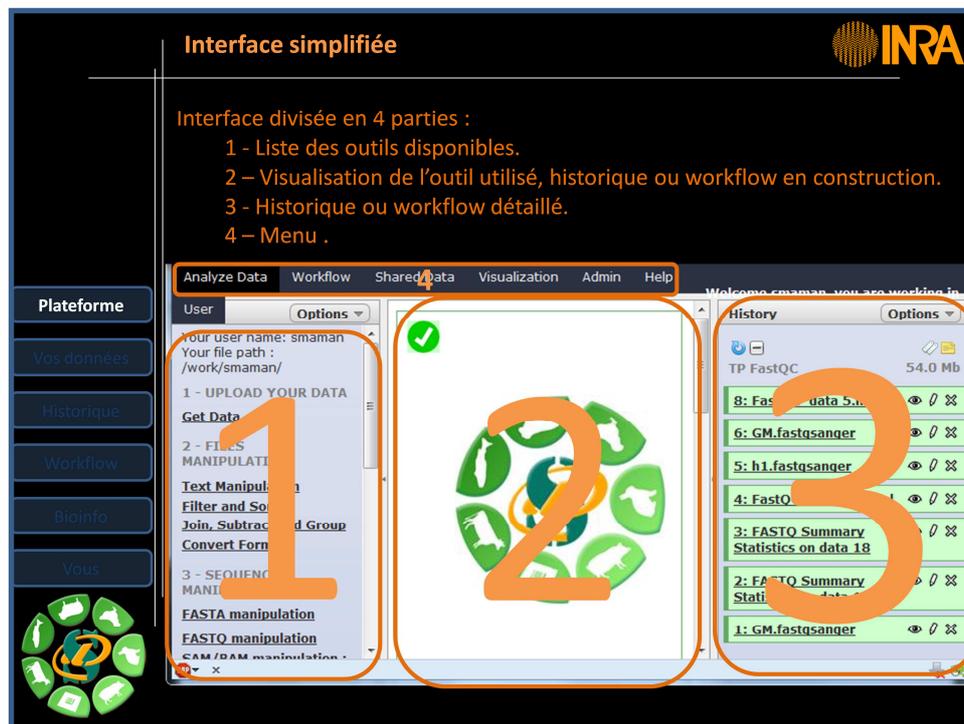


The screenshot displays a software interface with a dark theme. At the top, there is a navigation bar with tabs: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', and 'Help'. A status bar on the right indicates 'Using 13%'. Below the navigation bar, the main workspace is divided into three panes. The left pane, titled 'User', shows the user's name 'smaman' and file path '/work/smaman/'. It contains a menu with categories: '1 - UPLOAD YOUR DATA', '2 - FILES MANIPULATION', '3 - NGS: Mapping', '4 - MAPPING', and '5 - INDEL ET SNP'. Under '3 - NGS: Mapping', there are sub-items: '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', and 'Map with BWA for Illumina'. The 'Map with BWA for Illumina' item is highlighted with a red circle. The middle pane, titled 'Map with BWA for Illumina (version 1.2.2)', contains configuration options: 'Will you select a reference genome from your history?' (set to 'Use one from the history'), 'Select a reference from history:' (set to '11: phiX174_genome.fa'), 'Is this library mate-paired?:' (set to 'Single-end'), and 'FASTQ file:' (set to '14: phiX174_reads.fastqsanger'). The right pane, titled 'History', shows 'Unnamed history' with '0 bytes' and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'. At the bottom left of the interface, there are several green circular icons representing different actions like upload, download, and refresh.

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



The screenshot displays a web-based bioinformatics workflow interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', and 'Help'. A user profile section on the left shows the user 'smaman' with a file path of '/work/smaman/'. The main workflow area is divided into several steps: '1 - UPLOAD YOUR DATA', '2 - FILES MANIPULATION', '4 - MAPPING', and '5 - INDEL ET SNP'. Under step 4, the 'BWA - Bowtie' section is expanded, and the option 'Map with BWA for Illumina' is highlighted with a red circle. The right-hand panel shows the configuration for this step, titled 'Map with BWA for Illumina (version 1.2.2)'. It prompts the user to select a reference genome from a history, with '11: phiX174_genome.fa' selected. The library type is set to 'Single-end', and the FASTQ file is '14: phiX174_reads.fastqsanger'. A history panel on the far right shows a job titled '15: Map with BWA for Illumina on data 14 and data 11: mapped reads' that is currently waiting to run.



Votre écran est divisé en trois parties :

- 1- A gauche : vous trouverez la liste des outils linux et bioinformatiques disponibles pour le traitement de vos fichiers et de vos données.
- 2 - A droite, l'ensemble des fichiers de données que vous utilisez ainsi que les outils que vous avez sélectionné. Cette colonne se nomme "historique" car elle liste les fichiers et les outils utilisés. Chaque historique peut être nommé par vos soins et est, automatiquement, archivé.
- 3 - Au centre, selon le menu sélectionné en haut de l'écran, s'affichent les datasets, les workflow, les historiques ...

Your screen is divided into three parts:

- 1 - Left: There is a list of Linux and bioinformatics tools available to process your files and data.
- 2 - On the right, all your datasets and tools used to process these datasets are listed here. This column is called "historic". Each history can be named and is automatically archived.
- 3 - In the center, above you will find Galaxy menu and below, Galaxy interface displays datasets, workflows, histories ...

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.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

TOOL

Upload File from your computer

1: GM.fastqsanger 17.4 Mb
format: fastqsanger, database: 2

```

@HWI-ST-EAS627_1:8:1:17:202
TGGTTGGAGACCCAGATGGTCTTCAGCTCC
+
BB@=A-C987AB??-?B7B7B33>7A8?><1:9=A
@HWI-ST-EAS627_1:8:1:66:1050
CAGAAGTAGAGCAGAAAGAAAGCGAACCTTCCG
                    
```

DATASET (S)

WORKFLOW

HISTORY

TP FastQC 54.0 Mb

8: FastQC_data 5.html

6: GM.fastqsanger

5: h1.fastqsanger

4: FastQC_data 18.html

Generate pileup
Select the BAM file to generate pileup file for
Select a reference genome
output1 (tabular)

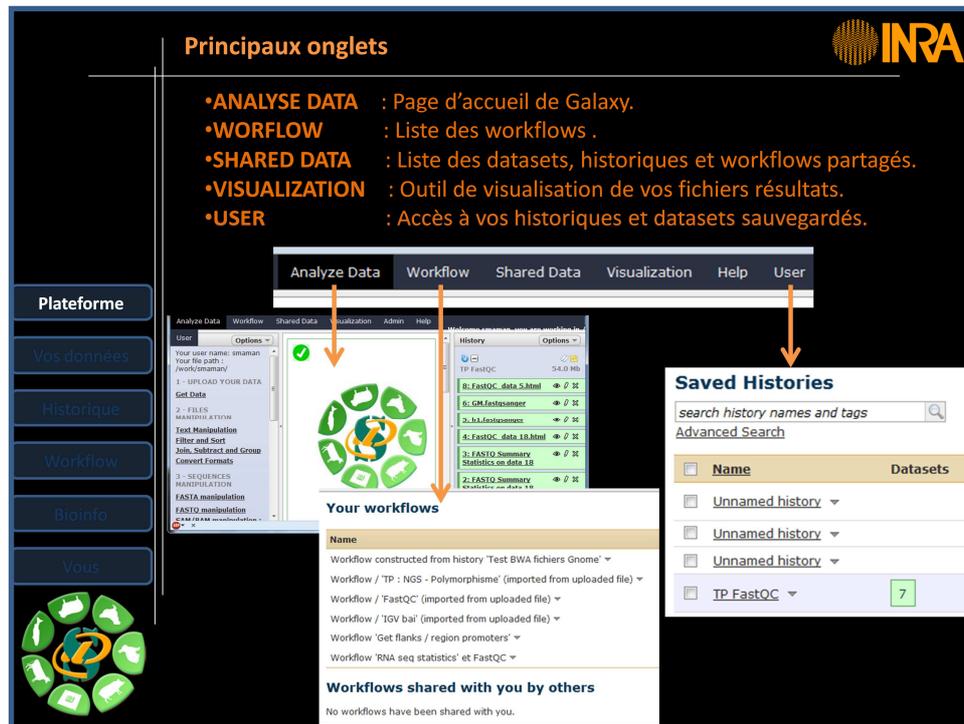
FASTA-to-Tabular
Convert these sequences
output (tabular)

SAM-to-BAM
Convert SAM file
Using reference file
output1 (bam)

Upload local file from filesystem path
out1 (bam, txt, fastqsanger, csfasta, qual, bed, gff, gtf, vcf, sam, fasta, pdf, xsq)

Generate (arrow from tool to dataset)

Dont la liste (arrow from history to tool)



Depuis la barre du menu principal en haut de votre écran, vous avez accès aux onglets suivants :

- Analyse Data : pour accéder à la première page de Galaxy. Cette page est divisée en trois zones (à gauche, les outils, au centre les paramètres, et à droite l'historique des fichiers de données et des traitements).
- Workflow : Liste de l'ensemble de vos workflows et, éventuellement, de ceux que vous partagez avec d'autres utilisateurs de Galaxy.
- Shared data : Liste des fichiers de données que vous avez téléchargés dans votre interface Galaxy ainsi que les historiques, les workflows et les graphiques de visualisation de vos résultats que vous avez éventuellement partagés avec d'autres utilisateurs de Galaxy.
- Visualisation : Il s'agit d'un outil de visualisation, comme IGV, mais spécifique à Galaxy. Cette outil vous permet de visualiser vos fichiers résultats.
- Help : Il s'agit d'un ensemble de liens vers les pages Internet du projet Galaxy. A ne pas confondre avec l'administration de l'outil Galaxy au sein de l'INRA. Ces liens donnent accès à : un support, un wiki, des vidéos tutoriels ainsi que le texte à copier pour citer Galaxy dans vos publications.
- User : cet onglet vous identifie (login et nom public) et vous permet d'accéder à vos historiques et fichiers de données sauvegardés.

Les onglets les plus couramment utilisés, tout au moins lors de vos premiers pas dans l'usage de l'interface Galaxy, sont « Analyse Data » et « User / Saved Histories ».

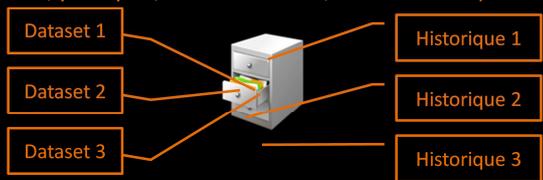
INRA

En résumé ...

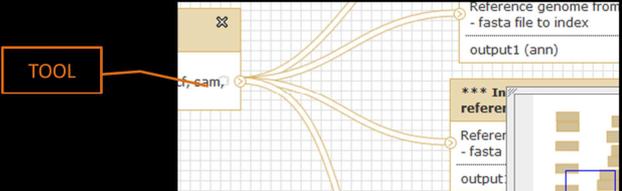
Un **DATASET** est un fichier de données (privées, publiques, fichiers d'entrée, fichier résultat) :



Votre **HISTORIQUE** est un « répertoire » qui « liste » l'ensemble de vos fichiers de données (privées, publiques, 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 (privées et publiques) ?

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

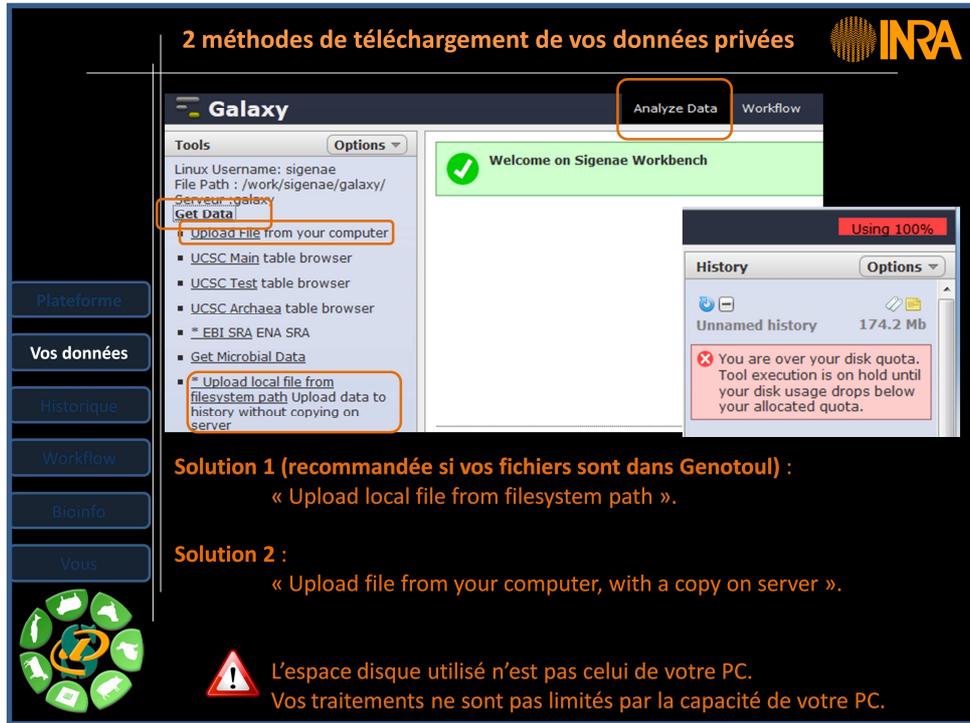
Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.



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.

Dans un premier temps, vous importez vos fichiers de données grâce à l'interface « Analyse Data / Get Data ». Puis vos fichiers de données téléchargés sont automatiquement archivés dans « User / Saved Datasets ».

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

Pour accéder à UCSC : menu « Analyze Data », puis « Get Data » / « UCSC Main table browser ».

Pour télécharger des fichiers de données, sélectionner vos paramètres, puis cliquer sur « Get output », puis sélectionner les données dont vous avez besoin et téléchargez les avec « Send query to Galaxy ».

Le fichier choisi est automatiquement importé dans vos datasets Galaxy.

To access UCSC: menu "Analyze Data", then "Get Data" / "UCSC Main table browser".

To download data files : select your settings, click on "Get output", select the data you need and download them with "Send query to Galaxy."

The selected file is automatically imported into your Galaxy history.

Présentation de la plateforme Galaxy.

Comment récupérer vos données (privées et publiques) ?

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

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

Lancement de traitements bioinformatiques.



Guide pour les utilisateurs Galaxy.

Gestion de vos historiques INRA

Historique

- 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: 2
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 |
|---------|-------|-----|-----|---------|------|
| fcolumn | count | min | max | sum | mean |
| 1 | 99115 | 2 | 33 | 3194703 | 32.2 |
| 2 | 99115 | 2 | 34 | 3156652 | 31.8 |
| 3 | 99115 | 2 | 34 | 3143060 | 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 | |
| FastQC | 6 | 0 Tags | Shared | 17.4 Mb | |
| TP : NGS - Polymorphisme | 8 | 2 | 0 Tags | Shared | 6.6 Gb |
| TP FastQC | 12 | 16 | 0 Tags | 54.0 Mb | |
| indexation genome | 1 | 0 Tags | | 46 bytes | |

For 0 selected histories: Rename Delete Delete Permanently

Comment suivre l'exécution de mes jobs



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



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

107: test.txt 👁️ 🗑️

Job is waiting to run

[i](#) [🔄](#)

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

1: test.txt 👁️ 🗑️

Job is currently running

[i](#) [🔄](#)

Etat 3 – VERT : Votre job est fini.

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

[i](#) [🔄](#)

```

>phiX
GAGTTTATCGCTTCCATGACCGAGAAGTTAACAATT
AAATTACTCTGATAAAGCGAATTAATCTACTGCTGTG
TGGTGGGAGAAATGAGAAATTCAGATCACTCTG
GCGACCTTGGCCATCACTAGATTCCTGCAAAA
TGGCTTAATATGCTGGCACTTCTCAAGAGACTGGT

```

Bug - ROUGE : Votre job est planté !

1: test.txt 👁️ 🗑️

0 bytes
An error occurred running this job:
info unavailable

[i](#) [🔄](#)

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

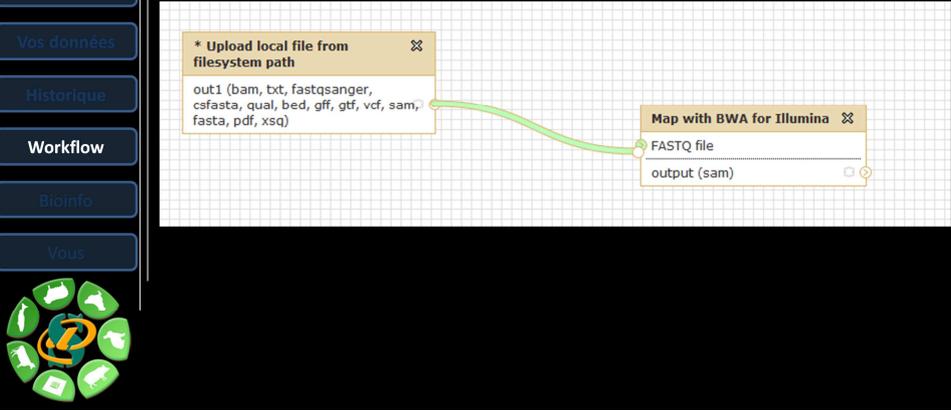
Vous



Créer un workflow 

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 !



Pour lancer plusieurs jobs en même temps, vous pouvez enregistrer un workflow à partir d'un historique donné (**History panel / click Options → Extract Workflow**).

To run multiple jobs simultaneously, you can save a workflow from a given historical (**History panel / click Options → Extract Workflow**).

Exporter votre historique en workflow.



The screenshot shows the IGV interface with the 'History' menu open. The 'Options' button in the 'History' header is circled in orange, with an arrow pointing to the 'Extract Workflow' option in the dropdown menu. The 'Extract Workflow' dialog box is open, showing a workflow name 'Workflow constructed from history: IGV bai', a 'Create Workflow' button, and a list of tools to be included in the workflow. The tools are:

| Tool | History items created |
|---|-----------------------|
| * Upload local file from filesystem path <input checked="" type="checkbox"/> Include "*" Upload local file from filesystem path" in workflow | 1: ERR000017.bam |
| * Upload local file from filesystem path <input checked="" type="checkbox"/> Include "*" Upload local file from filesystem path" in workflow | 8: ERR000017.sorte |
| * BAM sorted to BAI for IGV <input checked="" type="checkbox"/> Include "*" BAM sorted to BAI for IGV" in workflow | 11: * BAM sorted to |

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

INRA

Lancer plusieurs jobs simultanément

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

The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with 'Analyze Data' and 'Workflow' buttons. Below this is a section titled 'Your workflows' with a table listing workflow names. A right-click context menu is open over one of the workflow names, showing options like 'Edit', 'Run', 'Share or Publish', 'Download or Export', 'Clone', 'Rename', 'View', and 'Delete'. To the left of the main content is a sidebar with buttons for 'Plateforme', 'Vos données', 'Historique', 'Workflow', 'Bioinfo', and 'Vous'. Below the sidebar is a circular icon with various workflow symbols. Below the workflow list, there is a section titled 'Workflows shared by others' with a table listing shared workflows. Below this, there is a workflow diagram with two main steps: 'Production du fichier de statistiques' and 'Production du rapport'. The 'Production du rapport' step has an 'Options' menu highlighted.

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.

Le menu « Workflow » vous permet de lister l'ensemble des workflows disponibles. Un clic droit sur le nom d'un workflow permet de l'éditer ou de lancer le workflow sélectionné.

The menu 'Workflow' allows you to list all available workflows. Right click on the name of a workflow can edit or run it.

INRA

En résumédes datasets aux workflows.

- 1 – Télécharger vos fichiers de données
- 2 – Renommer votre historique :

History Options

history name 35.5 Mb

Map with BWA for Illumina (version 1.2.2)

Will you select a reference genome from your local filesystem? Use a built-in index

Select a reference genome: Arabidopsis thaliana

Is this library mate-paired?: Single-end

FASTQ file: 2: GM.fastasanger

- 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).

Galaxy
Your workflows

Étape 1: Importez vos fichiers de données

Tout d'abord, vous devez importer vos fichiers de données depuis le menu « Analyze data / Get Data ».
Ensuite, ces fichiers téléchargés sont automatiquement archivés dans « User / Saved Datasets ».

Étape 2: Sélectionnez vos outils et créez votre historique

Choisissez votre outil en fonction des traitements souhaités, dans la fenêtre « Tools », sur le côté gauche de l'interface Galaxy.
Toutes ces étapes sont automatiquement enregistrées dans un historique qui est aussi automatiquement archivé dans « User / Saved histoires ». Vous pouvez, si nécessaire, partager cet historique.

Étape 3: Exporter votre historique en workflow

Pour exécuter plusieurs tâches simultanément, vous pouvez enregistrer un workflow à partir d'un historique (**History panel / click Options → Extract Workflow**).

Vidéo "Comment extraire un workflow ? ":

<http://screencast.g2.bx.psu.edu/flash/WorkflowFromHistory.html>

Étape 4: Partager votre workflow

Solution 1 – Exporter puis importer un workflow

L'objectif est de télécharger un fichier contenant votre workflow sur votre PC.

Puis, un autre utilisateur peut importer ce fichier pour récupérer votre workflow.

Vidéo "Comment importer un workflow ? " : <http://screencast.g2.bx.psu.edu/mt-workflowImport/>

Solution 2 - Partagez votre workflow avec un autre utilisateur Galaxy

Pour partager votre workflow avec un autre utilisateur Galaxy, veuillez renseigner le mail de l'utilisateur. Attention, le mail au sens de Galaxy n'est pas le même que le mail INRA. Le mail Galaxy est de la forme : loginGalaxy@toulouse.inra.fr sachant que le login Galaxy est le login LDAP.

Présentation de la plateforme Galaxy.

Plateforme

Comment récupérer vos données (privées et publiques) ?

Vos données

Historique

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

Workflow

Bioinfo

Lancement de traitements bioinformatiques.

Vous



Guide pour les utilisateurs Galaxy.



Les principaux outils Galaxy

- 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, selectionner, 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

Comment citer Galaxy dans vos publications ?



Pour vos publications, citer:

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

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Extract features (version 1.0.0)

Select GFF data:

From:

Extract features:

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

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, et al. "Galaxy: a platform for interactive large-scale genome analysis." *Genome Research.* 2005 Oct; 15(10):1315-32.

Outils de traitement de fichiers



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_CTCT_R6 210 +
chr7 56736 56756 D17002_CTCT_R7 254 +
chr7 56761 56781 D17002_CTCT_R4 220 +
chr7 56792 56792 D17002_CTCT_R7 212 +
chr7 56775 56795 D17002_CTCT_R4 207 +
```

will produce:

```
chr7 56632 56652 D17003_CTCT_R6 210 +
chr7 56736 56756 D17002_CTCT_R7 254 +
```

The screenshot displays the Galaxy web interface. On the left is a navigation sidebar with buttons for 'Plateforme', 'Vos données', 'Historique', 'Workflow', 'Bioinfo', and 'Vous'. The main area is titled 'Outils bioinformatiques' and features a search bar with the text 'Mapper un FASTQ sur une référence avec BWA.' Below this is a 'NGS: Mapping' tool configuration panel. The 'Map with BWA for Illumina' option is selected and highlighted with an orange box. To the right of the configuration panel is a preview of the UCSC genome browser for chromosome 21, showing various tracks like RefSeq Genes, Human mRNAs, and HSK27Fc Mark.

Dans Galaxy, l'outil « [NGS:Mapping](#) → [Mapwith BWA for Illumina](#) » permet d'indexer et d'aligner les séquences. Les deux étapes d'indexation et d'alignement ne sont pas séparées.

L'indexation des séquences de référence peut prendre une demi-journée, il est donc préférable d'utiliser un fichier de référence déjà indexé. Pour cela, si vous ne trouvez pas un génome de référence dans le menu déroulant, veuillez demander à l'administrateur de Galaxy si un built-in index de référence indexé est disponible et si, par conséquent, il peut être ajouté à la liste.

In Galaxy, tool "NGS: Mapping → Map with BWA for Illumina" tool allows you to index and align sequences. The two-step indexing and alignment are not separated.

Index a reference sequences can take half a day, so it is best to use a reference file already indexed.

Therefore, if you can not find a reference genome from the dropdown menu, please ask the administrator to Galaxy if a built-in index indexed reference is available and, therefore, it can be added to the list.

En résumé ... 

De nombreux outils disponibles :

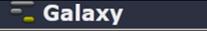
- Outils de traitement de fichiers
- BWA, FastQC, SAM Tools, Picard Tools ...

Facilité 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.

Plateforme

Comment récupérer vos données (privées et publiques) ?

Vos données

Historique

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

Workflow

Bioinfo

Lancement de traitements bioinformatiques.

Vous



Guide pour les utilisateurs Galaxy.

FAQ et formation en ligne 

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,**

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



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
- Demandes via des tickets Redmine ou mail à sigenae-support@listes.inra.fr



En conclusion ...



GALAXY

- ✓ Simplicité d'utilisation (sans Linux).
- ✓ Partage de vos datasets, historiques et workflows.
- ✓ Présentation schématique de vos traitements grâce aux workflows.
- ✓ Possibilité d'ajout de nouveaux outils, selon vos besoins.





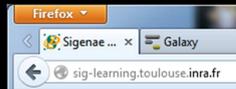
Votre auto-formation
continue en ligne avec
« sig-learning »



Votre accès à sig-learning



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



- 2 Authentification

Login

Pass

Plateforme

Vos formations

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

Homepage **Trainings** Profile Agenda Reporting Administration

Trainings

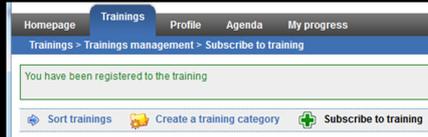
- 1 - Linux & Unix
UNIX1 – SIGENAE Team
- 2 - Cluster (en construction)
CLUSTER – SIGENAE Team
- 3 - Management of large files on Unix and Galaxy
UNIX2 – SIGENAE Team



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 |

Page d'accueil de votre formation



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



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.

Plateforme

Vos
formations

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

