

## 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 Sigenae 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 ▾



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 |

# Présentation de GALAXY

Sarah Maman - Sigenae

# Galaxy Project

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.

Une communauté active autour de cet outil.

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

Jeremy Goecks<sup>1</sup>, Anton Nekrutenko<sup>2,3</sup>, James Taylor<sup>1,3</sup> and The Galaxy Team



EMORY  
UNIVERSITY

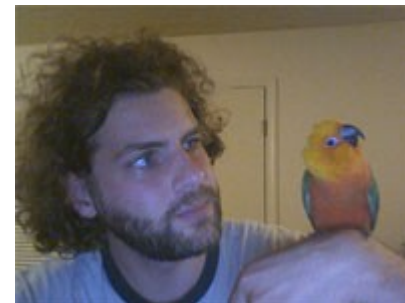
PENNSTATE



Anton Nekrutenko  
Penn State

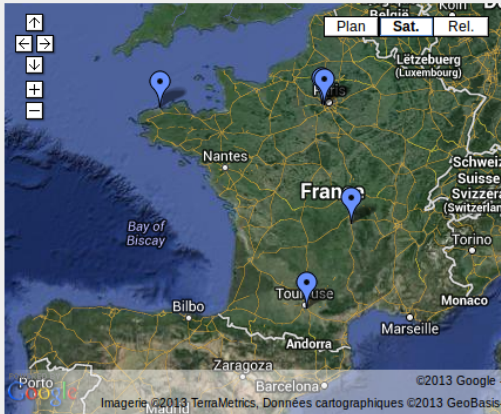


Nate Coraor  
Penn State



James Taylor  
Emory

# Une Galaxy parmi tant d'autres



Afficher [Galaxy IFB France](#) sur une carte plus grande

## Liste des instances

|                              |                                                          |                                                                                                   |                                                                             |
|------------------------------|----------------------------------------------------------|---------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>ABIMS Roscoff</b>         | Initiation, NGS Cleaning, RNASeq Differential Expression | <a href="http://galaxy.sb-roscoff.fr/">http://galaxy.sb-roscoff.fr/</a>                           | Christophe Caron - Alexandre Cormier - Gildas Lecorguille - Pierre Pericard |
| <b>Institut Curie</b>        | ChIP-Seq Analysis                                        | <a href="http://nebula.curie.fr/">http://nebula.curie.fr/</a>                                     | Alban Lermine                                                               |
| <b>Genotoul / Sigenae</b>    | Initiation to Galaxy, SNP calling, RNASeq, sRNASeq       | <a href="http://galaxy-workbench.toulouse.inra.fr/">http://galaxy-workbench.toulouse.inra.fr/</a> | Sarah Maman                                                                 |
| <b>INRA URGI</b>             | Differential expression analysis, Variant detection      | <a href="http://urgi.versailles.inra.fr/galaxy2">http://urgi.versailles.inra.fr/galaxy2</a>       | Olivier Inizan                                                              |
| <b>INRA MIGALE</b>           | Initiation to Galaxy, NGS Galaxy                         | <a href="http://migale.jouy.inra.fr/galaxy/">http://migale.jouy.inra.fr/galaxy/</a>               | Sandra Derozier - Franck Samson                                             |
| <b>Southgreen</b>            | Generalist platform, and crop breeding                   | <a href="http://gohelle.cirad.fr/galaxy/root/">gohelle.cirad.fr/galaxy/root/</a>                  | Jean-Francois Dufayard                                                      |
| <b>INRA PFEM / MetaboHUB</b> | Metabolomics data analysis                               | <a href="https://pfem-galaxy/">https://pfem-galaxy/</a>                                           | Franck Giacomoni                                                            |

## Groupe de travail Galaxy IFB

- ✓ Documentation collaborative (wiki)
- ✓ Formations (mise en commun agenda PF)
- ✓ Architecture
- ✓ Intégration d'outils (Tool Shed)

<http://www.ifb-galaxy.org>

# Une interface simplifiée

Interface divisée en 4 parties :

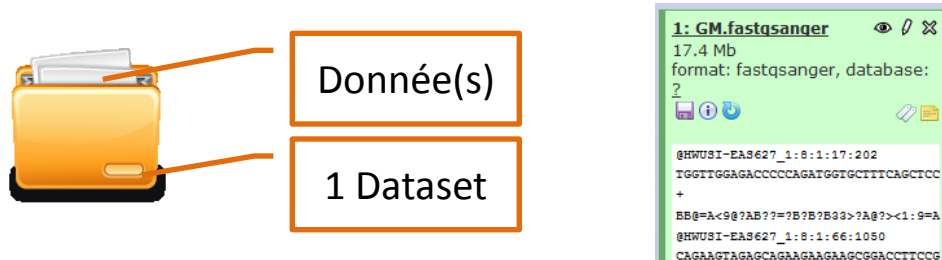
- 1 - Liste des outils disponibles.
- 2 - Visualisation de l'outil utilisé, historique, dataset ou workflow.
- 3 - Historique ou workflow détaillé.
- 4 - Menu .

The screenshot shows a web interface with a dark navigation bar at the top containing the following menu items: **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Admin**, and **Help**. A large orange number '4' is placed over the 'Shared Data' menu item. Below the navigation bar, the interface is divided into three main sections:

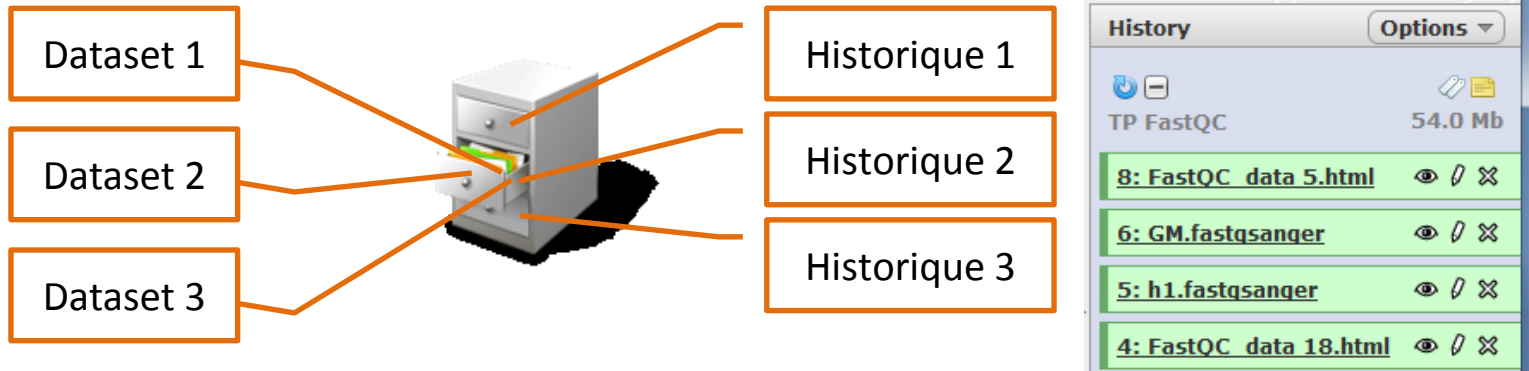
- Section 1 (Left):** A sidebar menu with a large orange number '1' overlaid. It contains a 'User' section with fields for 'Your user name: smaman' and 'Your file path: /work/smaman/'. Below this are several tool categories: '1 - UPLOAD YOUR DATA', 'Get Data', '2 - FILES MANIPULATION', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Form', '3 - SEQUENCE MANIPULATION', 'FASTA manipulation', and 'FASTQ manipulation'.
- Section 2 (Center):** A main workspace area with a large orange number '2' overlaid. It features a green checkmark icon in the top left and a central graphic of a circular arrangement of green and blue icons representing various bioinformatics tools.
- Section 3 (Right):** A 'History' panel with a large orange number '3' overlaid. It displays a list of recent operations, each with a green background and icons for visibility, edit, and delete. The operations listed are: '8: FastQC data 5...', '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'.

# Un vocabulaire spécifique

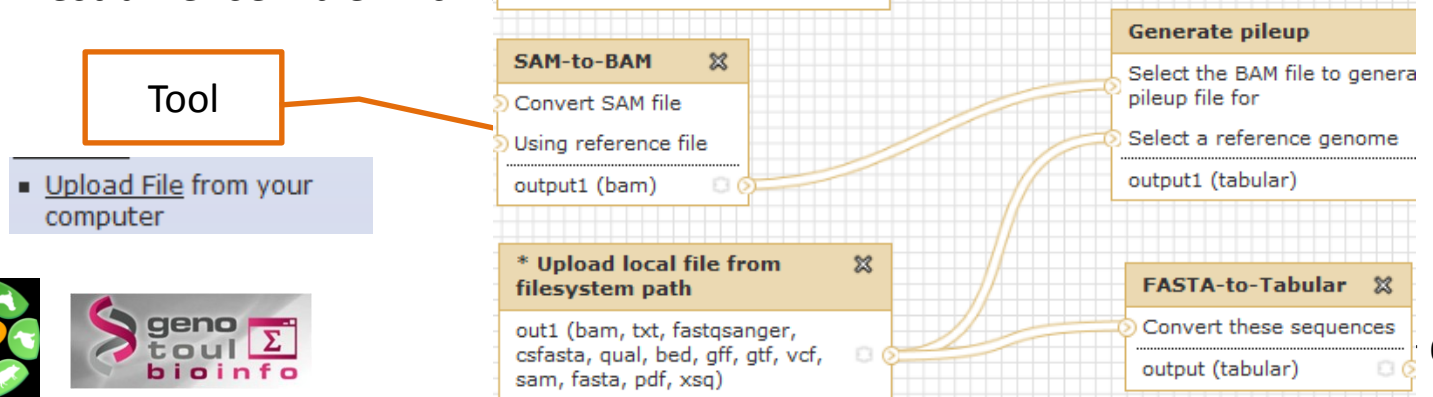
Un **DATASET** est un fichier de données (fichiers d'entrée, fichiers résultats) :



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



Votre **WORKFLOW** est un ensemble : fichiers. outils. traitements.





# Suivi des jobs

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 Shared, Accessible |
| test TP miRNA     | 36       | 1    | 0 Tags                    |
| Unnamed history   |          |      | 0 Tags                    |

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



Rapidement, beaucoup de données sont générées

→ D'où l'importance de bien nommer ses historiques / datasets / workflows pour les trier et les organiser au mieux.

# Installation d'une instance GALAXY

Sarah Maman



# Instance publique du Galaxy Project

Sur le serveur public du Galaxy Project : <https://usegalaxy.org/>

The screenshot shows the Galaxy Project web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Cloud', 'Help', and 'User'. A 'Using 0%' indicator is in the top right. The left sidebar shows a 'Tools' section with a search bar and a list of tool categories like 'Get Data', 'Send Data', 'ENCODE Tools', 'Lift-Over', 'Text Manipulation', 'Convert Formats', 'FASTA manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Extract Features', 'Fetch Sequences', 'Fetch Alignments', 'Get Genomic Scores', 'Operate on Genomic Intervals', 'Statistics', 'Graph/Display Data', 'Regional Variation', 'Multiple regression', 'Multivariate Analysis', 'Evolution', and 'Motif Tools'. The main content area displays a welcome message: 'Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).' Below the text is a word cloud with 'Galaxy is hiring' as the central theme, surrounded by terms like 'workflow', 'sequencing', 'reproducible', 'biology', 'data', 'genomics', 'research', 'accessible', 'analysis', 'science', 'NGS', 'transcriptomics', 'visualization', 'team', 'computational', and 'transparent'. The right sidebar shows a 'History' section with 'Unnamed history' and '0 bytes', and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.

De nombreuses autres instances publiques sont disponibles.

# Instance SlipStream du Galaxy Project

- Edité par Galaxy project
- Solution plug-and-play
- Inutile d'installer, de configurer et d'administrer une instance, ou de gérer les mises à jour code et outils.
- Solution adaptée pour les biologistes non informaticiens.
- Quelques infos sur la performance :

| TOOLS       | TASK                             | DATA                                        | RUN-TIME              |
|-------------|----------------------------------|---------------------------------------------|-----------------------|
| Bowtie 2    | Mapping whole human genome       | 204 million paired-end 100bp Illumina reads | 2 Hours<br>44 Minutes |
| SAMTools    | SAM-BAM conversion               | 127GB SAM (41GB resulting BAM)              | 2 Hours<br>7 Minutes  |
| TopHat 2    | RNA-Seq mapping                  | 24 million 100bp Illumina reads             | 1 Hours<br>24 Minutes |
| Cufflinks 2 | Differential Expression Analysis | 4.3 GB SAM File                             | 0 Hours<br>11 Minutes |

# Instance sur un VM ou en local

**VirtualBox** est un logiciel de virtualisation de systèmes d'exploitation.

En utilisant les ressources matérielles de l'ordinateur (*système hôte*), VirtualBox permet la création d'un ou de plusieurs ordinateurs virtuels dans lesquels s'installent d'autres systèmes d'exploitation (*systèmes invités*). Source : <http://doc.ubuntu-fr.org/virtualbox>

## Voici les principales étapes d'installation de Galaxy sur votre vm

1. Préparer votre espace de travail

```
mkdir ~/galaxy-python
ln -s /path/to/python2.7 ~/galaxy-python/python
export PATH=~/galaxy-python:$PATH
```
2. Télécharger les sources de Galaxy (Mercurial ?)

```
hg clone https://bitbucket.org/galaxy/galaxy-dist/
cd galaxy-dist
hg update stable
```
3. Lancer Galaxy

```
sh run.sh
```
4. Paramétrer Galaxy : universe ini file



## Intérêts d'un vm

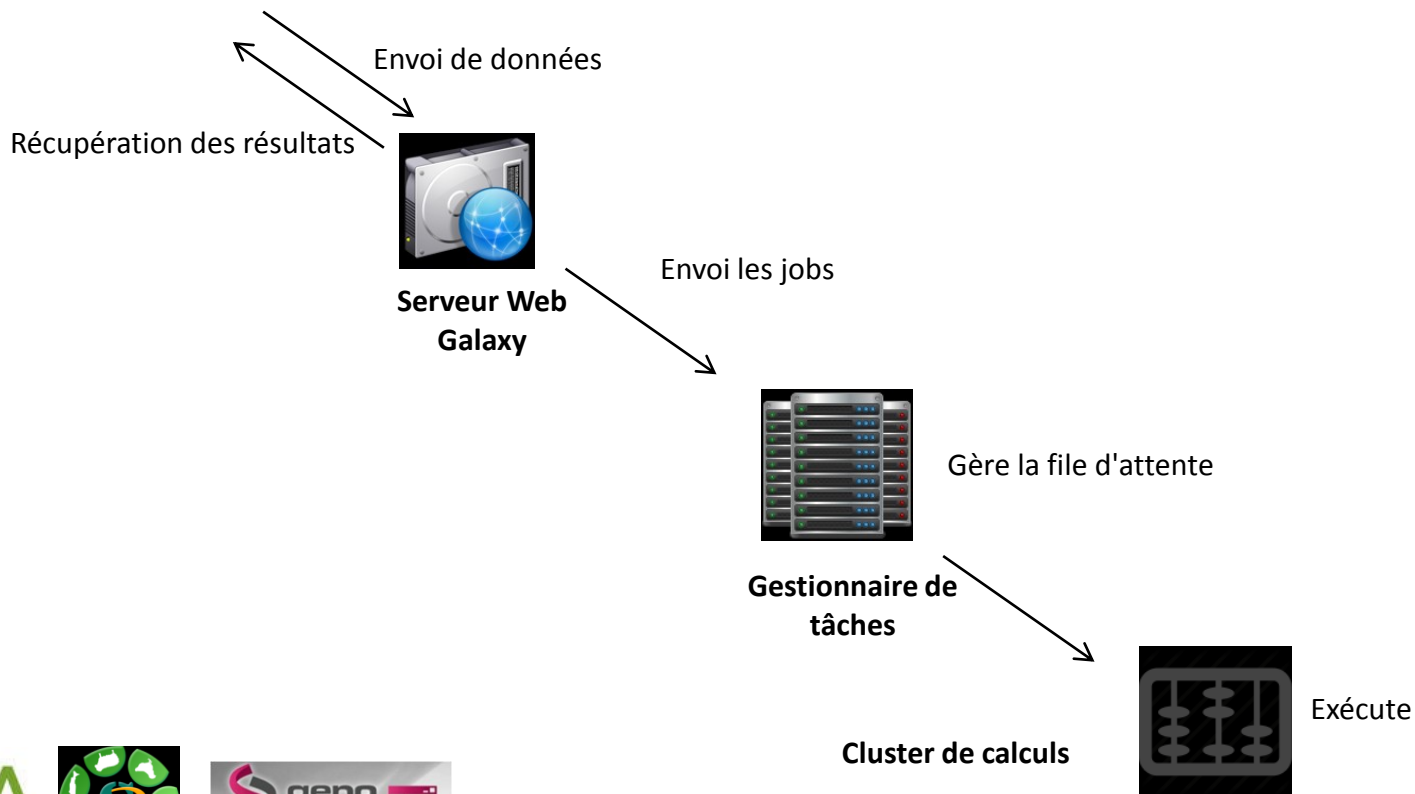
1. Faire fonctionner plus d'un système d'exploitation en même temps en toute sécurité
2. Possibilité de cloner une vm donc de partager des machines Galaxy
3. vm sauvegardée donc restaurable.

# Architecture de l'instance toulousaine

Galaxy est installée sur une machine virtuelle qui envoie les calculs à un cluster.



Utilisateur de Galaxy (biologiste)



# Comparaison des différentes instances

En fonction du contexte d'utilisation et des ressources IT disponibles :

|                   | NO WAIT TIMES | NO STORAGE QUOTAS | NO JOB SUBMISSION LIMITS | NO DATA TRANSFER BOTTLENECKS | NO IT EXPERIENCE REQUIRED | NO REQUIRED INFRASTRUCTURE |
|-------------------|---------------|-------------------|--------------------------|------------------------------|---------------------------|----------------------------|
| GALAXY MAIN       | ✗             | ✗                 | ✗                        | ✗                            | ✓                         | ✓                          |
| LOCAL GALAXY      | ?             | ?                 | ?                        | ✓                            | ✗                         | ✗                          |
| CLOUD GALAXY      | ✓             | ✓                 | ✓                        | ✗                            | ✗                         | ✓                          |
| SLIPSTREAM GALAXY | ✓             | ✓                 | ✓                        | ✓                            | ✓                         | ✓                          |

Source : <http://bioteam.net/slipstream/galaxy-edition/>

# Configuration et administration d'une instance GALAXY

Ibouniyamine Nabihoudine - Sarah Maman  
Sigenae

# Configuration à l'aide du fichier universe\_wsgi.ini

- Fichier de configuration .ini :
  - Déploiement
  - Répertoire de travail
  - Base de donnée de travail, etc.
- Ce fichier est organisé en sections :
  - **[server:main]** : configuration du serveur de déploiement
  - **[app:main]** : configuration de l'application Galaxy
  - Autres thèmes : Files and directories, Logging and Debugging, Job execution, Users and Security
- Configuration de l'adresse de déploiement de l'instance **[server:main]**
  - Section **[server:main]**

```
port = 8090
```
  - Paramètres **host** et **port**

```
host = 127.0.0.1
```

... déploiement de l'instance Galaxy à l'adresse **http://127.0.0.1:8090**



# Autres configurations

## Configuration des fichiers et répertoires de travail

**Job\_working\_directory** : Scripts sh générés par <galaxy et lancés sur le cluster.

**datatypes\_conf\_file** : Fichier permettant de configurer les types de données

**job\_working\_directory** : Répertoires/Fichiers de travail de Galaxy

**tool\_config\_file** : Fichier de configuration et description des outils disponibles

## Gestion des utilisateurs

Plusieurs méthodes de sécurisation d'une instance :

- Gestion des utilisateurs interne à Galaxy (universe.ini)
- Gestion externe des utilisateurs (LDAP)

Les paramètres à modifier pour la gestion interne à Galaxy :

- **require\_login** : **True** pour limiter l'accès aux utilisateurs ayant un compte galaxy
- **allow\_user\_creation** : **True** pour autoriser un utilisateur à se créer lui-même un compte

# Intégrer un outil (bio)informatique dans une instance GALAXY

Sarah Maman  
Sigenae

# Une galaxy de poupées russes

Mr Galaxy sur le web

- Instance locale
- Partage ToolShed

Wrapper xml

- I/F web
- Avec ou sans Cheetah

Wrapper perl

- Facultatif
- Adaptable à chaque instance

(Bio)Info tool ou ligne de commande

- Localisation Path
- Evolution



# Etapes pour l'ajout d'un wrapper

Demande explicite de wrapper

Exécution direct :  
tools/ini\_tool.XML  
Avec Cheeta

Exécution indirecte avec un wrapper : tools/ini\_tool.XML +  
Perl/py/..

ToolShed / tool  
web  
Récupération du  
wrapper (wget)

Vérifier l'accès aux bioinfo tools

Configurer job\_conf.xml et tool\_conf.xml

Restart vm-galaxy-test ou reload xml depuis l'interface web d'administration de Galaxy

Tester les outils

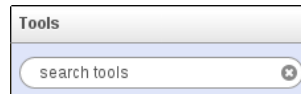
Partage de code sous SVN

Mise en production sur vm-galaxy-dev

# Etales pour l'ajout d'un wrapper

## 1. Vérifier si le wrapper existe déjà ?

1. Dans votre instance galaxy ?
2. dans le ToolShed Galaxy Project (<http://wiki.galaxyproject.org/Tool%20Shed>) ?
3. ou sur le site web de l'outil ?



## Categories of valid repositories

search repository name, description



2. Si le wrapper est dans le ToolShed Galaxy, récupérer le (wget), le paramétrer, puis accéder directement au point 5.
3. Créer le fichier xml dans tools/vosinitiales\_myTool/
4. Vérifier les accès aux éventuels outils (bio)informatiques sous jacents.
5. Ajouter l'outil dans le tool\_conf.xml
6. Configuration du mode d'exécution
7. Redémarrer vm-galaxy-test pour tester votre wrapper
8. Tester et corriger -> retour à l'étape 7.
9. Transmettre votre code dans le SVN Galaxy pour un passage en production

# Qu'est-ce qu'une demande explicite de wrapper ?

1. Tout d'abord, vérifier si l'outil n'existe pas déjà dans l'instance Galaxy.
2. Demande correspondant à un besoin exprimé par plusieurs utilisateurs ou utile à plusieurs utilisateurs.
3. Demande accompagnée des informations utiles listées ci-dessous :

## Informations à transmettre pour une récupération d'un outil dans le ToolShed Galaxy

- Intitulé exact de l'outil
- Sa version et date de création
- Auteur
- Lien URL (si possible) ou nom du Tool Shed (international, local, IFB , autre ?).
- Thématique / section correspondante dans Galaxy.

## Informations à transmettre pour l'intégration d'un script (maison ou pas)

- Si le script fait appel à des outils bioinfo, veuillez préciser lesquels ainsi que leur version.
- Un exemple de ligne de commande avec l'ensemble des options souhaitées et des paramètres.
- Des fichiers de tests (les plus légers possibles).
- Une Liste détaillée des inputs et outputs avec, pour chaque fichier: son format, son intitulé exact, si le fichier doit être ou pas récupérer par l'utilisateur Galaxy, si l'output sera ou pas réutilisé par un autre script ensuite.
- Les commentaires de l'outil (documentation en anglais seulement), à destination des utilisateurs Galaxy:
  - . Exemple d'utilisation,
  - . Description détaillée des fichiers entrants et sortants,
  - . Usage détaillé du script et autres informations qui vous semblent utiles,
  - . Description détaillée des paramètres et des options.
- Si plusieurs scripts constituent un workflow, préciser : L'ordre des scripts, les liens entre ces scripts (output de l'un = input de l'autre ?), vérifier aussi la compatibilité des formats).

# Premiers pas dans le ww (wrapper world)

## 1. Choisir son éditeur favori

1. Dans votre instance galaxy ?
2. dans le ToolShed Galaxy Project (<http://wiki.galaxyproject.org/Tool%20Shed>) ?
3. ou sur le site web de l'outil ?

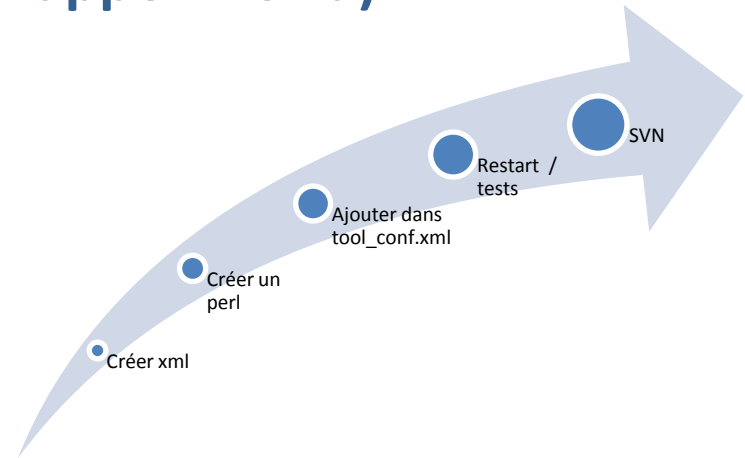
- Ou utiliser une interface dédiée : <http://cli-mate.lumc.nl/>

## 1. Créer le fichier xml dans tools/vosinitiales\_myTool/

## 2. Vérifier les accès aux éventuels outils (bio)informatiques sous jacents.

## 3. Récupérer les paths d'accès à ces outils avec la commande :

\$ which tool



The screenshot shows the CLI-mate web interface. At the top, there is a logo and the text 'CLI-mate Interface generator for command line programs'. Below this, there are navigation tabs for 'Tool Definition', 'Interface Generation', and 'About Us', along with the version 'Version 0.3'. The main content area has a 'Load a RDF definition' input field and an 'Upload' button. Below that, there are three tabs: 'General Information', 'Parameters', and 'Runtime Requirements'. The 'General Information' tab is active, showing fields for 'Name \*', 'Command \*', 'Description', 'Owner', 'Owner Email', and 'Version'. To the right of these fields is a text area for 'Enter the help manual below.' with a help icon. A note at the bottom of the form states '\* = required fields.' Below the form, there is a 'Test your definition' section with options to generate RDF definitions in 'turtle' format and to generate interfaces using a 'galaxy.xml' template. At the bottom, there is a 'Save and Share' section with options to download RDF definitions in 'turtle' format and a 'Submit' button.



# Fichier xml

Ce fichier XML est :

- un formulaire de saisie,
- visible depuis l'interface web Galaxy.

```
<tool id="fa_gc_content_1" name="Compute GC content">
  <description>for each sequence in a file</description>
  <command interpreter="perl">toolExample.pl $input $output</command>
  <inputs>
    <param format="fasta" name="input" type="data" label="Source file" />
  </inputs>
  <outputs>
    <data format="tabular" name="output" />
  </outputs>

  <help>
This tool computes GC content from a FASTA file.
  </help>

</tool>
```

Nom du tool affiché sur l'interface web,  
dans le menu de gauche.

Nom interne à Galaxy

Label affiché sur l'interface web

Help :  
reStructuredText  
Markup Specification :  
<http://docutils.sourceforge.net/docs/ref/rst/restructuredtext.html>

# Tags de base de votre XML

```
<tool id = "id_outil" name = "nom_outil " version = "version_outil">  
  <description> description de l'outil </description>  
  <command> perl ou cheeta </command>  
  <inputs>  
    <param ... />  
  </inputs>  
  <outputs>  
    <data ... />  
  </outputs>  
  <help> commentaires, contact, citation </help>  
</tool>
```

# Tag <tool> et <description>

```
<tool id="sm_mothur_preprocess_V-1-0_date" name="* Nom de l'outil visible dans l'interface galaxy">
  <description>plus de détails</description>
  <command interpreter="perl">sm_wrapper_version_date.pl
    $input
    $param
    "$int "
    $output
  </command>
  <inputs> .. </inputs>
  <outputs> ... </outputs>
  <help> ... </help>
</tool>
```

- L'id du tool doit être unique (integrated\_tool\_conf.xml)
- Le fichier sm\_wrapper\_version\_date.pl est appelé dans la balise <command> du xml.
- Le xml et le wrapper perl (ou autre langage) doivent être dans le même répertoire : tools/vosInitiales/fwrapper.pl et file.xml
- Les outils « maison » de l'instance Sigenae de Galaxy sont identifiés avec une \*
- Les outils Galaxy sont en anglais.

# Tag <command>

## Exécution directe :

- Ligne de commande :

```
<command>sed -r '$pattern' $input > $outfile </command>
```

- CHEETA (<http://www.cheetahtemplate.org/>): langage puissant mais limité. D'où le passage par un wrapper.

```
<command interpreter="python">gatk_wrapper.py  
  --max_jvm_heap_fraction "4"  
  --stdout "${output_log}"  
  #if str( $ref ) != "None":  
    -d "gatk_input"  
  #end if </command>
```

## Exécution indirecte avec un wrapper :

- un script ini\_tool\_wrapper.pl est appelé dans la balise <command>
- Nous appellerons ce script : wrapper.
- Peu importe le langage de ce wrapper. Ils sont tous supportés par Galaxy.
- Il est cependant nécessaire de préciser le langage utilisé dans « interpreter ».

```
<command interpreter="perl">ini_tool_wrapper.pl ... </command>
```

# Fichier xml : principaux composants dans <inputs></inputs>

|                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                      |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Concatenate Dataset:</b></p> <p>264: Hierarchical classif report</p>                                                                                                                             | <p>Fichier entrant :</p> <pre>&lt;param name = "nom_interne" type = "data" format = "fasta" label = "affichage" /</pre>                                                                                                                                                                                                                                                                              |
| <p><b>Add this value:</b></p> <p>1</p>                                                                                                                                                                 | <p>Champs de saisie :</p> <pre>&lt;param name = "nom_interne" type = "integer " value = "10.0" label = "affichage"/&gt;</pre>                                                                                                                                                                                                                                                                        |
| <p><b>Iterate?:</b></p> <p>NO</p> <p>NO</p> <p>YES</p>                                                                                                                                                 | <p>Liste :</p> <pre>&lt;param name = "nom_interne" type = "select " label = « Iterate ?"&gt;   &lt;option value = "T"&gt;NO&lt;/option&gt;   &lt;option value = "F"&gt;YES&lt;/option&gt; &lt;/param&gt;</pre>                                                                                                                                                                                       |
| <p><b>Type of BLAST:</b></p> <p><input checked="" type="radio"/> blastn</p> <p><input type="radio"/> blastn-short</p> <p><input type="radio"/> dc-megablast</p> <p><input type="radio"/> megablast</p> | <p>Boutons radio :</p> <pre>&lt;param name = "nom_interne" type = "select " display = "radio" label = "affichage"&gt;   &lt;option value = "megablast"&gt;megablast&lt;/option&gt;   &lt;option value = "blastn"&gt;blastn&lt;/option&gt;   &lt;option value = "blastn-short"&gt;blastn-short&lt;/option&gt;   &lt;option value = "dc-megablast"&gt;dc-megablast&lt;/option&gt; &lt;/param&gt;</pre> |

# Fichier xml : principaux composants dans <inputs></inputs>

## Select a reference genome (

Arabisopsis thaliana ▼

- Arabisopsis thaliana
- Arabisopsis lyrata
- Bos taurus
- Drosophila melanogaster
- Homo sapiens
- Mus musculus
- Rattus norvegicus
- V4\_454Scaffolds
- V4\_454Scaffolds\_filter
- Yeast
- Sus scrofa

1 - XML avec le composant select pour le "Génome de référence "

```
<param name="input_ref_genome" type="select" label="Select a reference genome (if your genome of interest is not listed, please contact Sigenae)">
  <options from_file="mirdeep2_indexes.loc">
    <column name="name" index="0"/>
    <column name="value" index="1"/>
  </options>
</param>
```

2 - **mirdeep2\_indexes.loc** dans tool-data/ :

more mirdeep2\_indexes.loc

Arabisopsis thaliana /save/galaxy-dev/bank/mirdeep2/Arabisopsis\_thaliana

Arabisopsis lyrata /save/galaxy-dev/bank/mirdeep2/ensembl\_arabisopsis\_lyrata\_genome

Bos taurus /save/galaxy-dev/bank/mirdeep2/ensembl\_bos\_taurus\_genome

## Statistic(s) chosen:

Select All

Unselect All

mean

sd

variance

median

quartile

decile

Cases à cocher:

```
<param name="stat_chosen" type="select" display="checkboxes" multiple="True" label="Statistic(s) chosen">
  <option value="mean">mean</option>
  <option value="sd">sd</option>
  <option value="variance">variance</option>
  <option value="median">median</option>
  <option value="quartile">quartile</option>
  <option value="decile">decile</option>
  <validator type="empty_field" message="Please choose a statistic representation" />
</param>
```



# Fichier xml : principaux composants dans <inputs></inputs>

Your first htseq count file:

First htseq count file name:

Datasets

Add new Dataset

1 - Répéter un paramètre dans le tag <inputs> du XML :

```
<param format="tab" name="input_htseqcount" type="data" label="Your first htseq count file"/>
<param name="name1" size="20" type="text" value="" label="First htseq count file name"/>
  <repeat name="queries" title="Dataset">
    <param name="inputs_count" type="data" format="tab" label="Other htseq count files" />
    <param name="names" size="20" type="text" value="" label="htseq count file name"/>
  </repeat>
```

1 - Répéter un paramètre dans le tag <command> du XML :

```
<command interpreter="perl">sm_htseqcount_merge.pl
... #for $q in $queries
    ${q.inputs_count}
    ${q.names}
  #end for
</command>
```

Plus d'infos :

<https://wiki.galaxyproject.org/Admin/Tools/ToolConfigSyntax?action=show&redirect=Admin%2FTools%2FTool+Config+Syntax>



# Fichier xml : principaux composants dans <inputs></inputs>

1 – Une condition dans le tag <inputs> du XML :

```
<conditional name="stat_cond">
  <param name="stat" type="select" help="Possible values" label="Stats T or F">
    <option value="T">T</option>
    <option value="F">F</option>
  </param>
  <when value="F" />
  <when value="T">
    <param name="stat_chosen" type="select" display="checkboxes" multiple="True" label="Statistic(s) chosen">
      <option value="mean">mean</option>
      .....
      <validator type="empty_field" message="Please choose a statistic representation" />
    </param>
  </when>
</conditional>

<conditional name="ploting_cond">
  <param name="ploting" type="select" help="Ploting" label="Ploting T or F">
    <option value="T">T</option>
    <option value="F">F</option>
  </param>
  <when value="F" />
  <when value="T">
    <param name="plot_chosen" type="select" help="" display="checkboxes" multiple="True" label="Plot(s) chosen">
      <option value="boxplot">boxplot</option>
      <option value="histogram">histogram</option>
      <option value="density">density</option>
      <option value="pairsplot">pairsplot</option>
      <option value="MAplot">MAplot</option>
    </param>
  </when>
</conditional>
```

Stats T or F:

F ▾

Possible values

Ploting T or F:

T ▾

Ploting

Plot(s) chosen:

Select All

Unselect All

boxplot

histogram

density

pairsplot

MAplot

2 – Une condition dans le tag <command> du XML :

```
<command interpreter="perl">Graphics_desc.pl
...
  $stat_cond.stat
  $stat_cond.stat_chosen
  $ploting_cond.ploting
  $ploting_cond.plot_chosen
...
</command>
```

# Fichier xml : principaux composants dans <outputs></outputs>

<outputs>

```
<data format = "pdf" name = " result" label " result of ${tool.name} on ${input.name} />
```

```
<filter>pdf is True</filter>
```

```
<!--Ne pas afficher cet output s'il n'est pas généré -->
```

```
<data name="blast_out" format="tabular" label="BLAST Report">
```

```
<change_format>
```

```
<-- format variable en fonction des conditions d'exécution -->
```

```
<when input="view_options" value="0" format="txt"/>
```

```
<when input="view_options" value="7" format="blastxml"/>
```

```
<when input="view_options" value="8" format="tabular"/>
```

```
<when input="html_output" value="T" format="html"/>
```

```
</change_format>
```

```
</data>
```

```
</outputs>
```

239: Hierarchical  
classif report

error  
An error occurred with this

242: Hierarchical  
classif report

358 bytes  
format: html, database: ?  
Group member file NO hclustfun(  
count.file = "/work/galaxy-  
dev/database/files  
/001/dataset\_1961.dat",  
group.member.file = NULL,  
format.image.out = "jpeg",  
transformation.method = "none",  
sample.clustering = TRUE, se



HTML file

# Fichier xml : Commentaires de l'outil avec la balise <help></help>

**\* Sigcufflinks (version 1.0.0)**

Your accepted hits bam file:

Your gtf file:

G or g ?:  
quantitate against reference transcript annot. ▾

Execute

Cufflinks code has been modified in Sigcufflinks by the  
*OPTIONS :*

-p/--num-threads : number of threads used during ana  
-G ou (exclusif) -g :  
-G/--GTF quantitate against reference transcript annot  
-g/--GTF-guide use reference transcript annotation to

## Cufflinks Overview

Cufflinks assembles transcripts, estimates their abunda  
reads and assembles the alignments into a parsimonious  
many reads support each one. Please cite: Trapnell C,  
assembly and abundance estimation from RNA-Seq reve  
doi:10.1038/nbt.1621

## Know what you are doing

⚠ There is no such thing (yet) as an automated gears

<help>

**\*\*Titre en gras\*\***

Pour plus de détails, cliquer ici\_...\_ ici: <http://www.google.fr>

.. class:: warningmark  
Warning

**How to cite**

</help>

# Wrapper type

## Ajouter la licence

```
Copyright (C) 2009 INRA #  
This program is free software: you can redistribute it and/or modify  
it under the terms of the GNU General Public License as published by  
the Free Software Foundation, either version 3 of the License, or  
(at your option) any later version. #  
This program is distributed in the hope that it will be useful,  
but WITHOUT ANY WARRANTY; without even the implied warranty of  
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
GNU General Public License for more details. #  
You should have received a copy of the GNU General Public License  
along with this program. If not, see <http://www.gnu.org/licenses/>
```

## Preciser la version de l'outil informatique ou bioinfo sous-jacent

```
<version_command>tophat -version</version_command>
```

# Wrapper type

## Exemple de code du xml

Le nom du wrapper commence par une \* lorsqu'il s'agit d'un tool ajouté par Sigeane ou PF BioInfo à l'instance Galaxy. L'objectif étant de distinguer facilement les tools de base de Galaxy Project des tools ajoutés dans l'instance.

```
<tool id="unique_wrapper_id" name="* nom du wrapper">
  <description></description>
  <command interpreter="perl">vosInitiales_nomWrapper.pl
  --input_1 $input1
  --param $param
  --output1 $output1
  --selector $conditional.selector
  #if $conditional.selector == "1":
    --conditionalparam1 $conditional.param1
    --conditionalparam2 $conditional.param2
  #else
    --conditionalparam3 $conditional.$param3
    --conditionalparam4 $conditional.$param4
  #end if
  </command>
  <inputs>
    <param format="txt" name="input1" type="data" label="First input file"/>
    <param name="param" type="select" label="Param">
      <option value="fastq">fastq</option>
      <option value="fq">fq</option>
    </param>
    <conditional name="conditional">
      <param name="selector" type="select" label="Question for biologists ?">
        <option value="0">No</option>
        <option value="1">Yes</option>
      </param>
      <when value="1">
        <param name="param1" type="data" format="fasta" label="Path to file 1"/>
        <param name="param2" type="data" format="fasta" label="Path to file 2"/>
      </when>
      <when value="0">
        <param name="param3" size="20" type="text" value="0" label="param 3"/>
        <param name="param4" size="20" type="text" value="0" label="param 4"/>
      </when>
    </conditional>
  </inputs>
  <outputs>
    <data format="fasta" name="output1" label="file.fasta"/> <!-- choisir un label le plus court possible -->
  </outputs>
  <help>
    Voir la section du wiki Galaxy consacrée à la rédaction des tools Galaxy.
  </help>
</tool>
```

# Exemple de wrapper & bonnes pratiques

```
#!/usr/bin/perl -w

# usage : perl script.pl <input 1> <output 1>
# Date - Wrapper dans le cadre du projet .... (Pour 'nom du/des biologistes')

use strict;
use File::Basename;
use Getopt::Long;

my $input1;
my $param;
my $output1;
my $selector;
my $conditionalparam1;
my $conditionalparam2;
my $conditionalparam3; ##conditional.param3
my $conditionalparam4; ##conditional.param4

Getopt::Long::Configure( 'no_ignorecase', 'bundling' );
GetOptions (
    'input1=s' => \$input1,
    'param=s' => \$param,
    'output1=s' => \$output1,
    'selector=i' => \$selector,
    'conditionalparam1=i' => \$conditionalparam1,
    'conditionalparam2=i' => \$conditionalparam2,
    'conditionalparam3=s' => \$conditionalparam3,
    'conditionalparam4=s' => \$conditionalparam4
) or die "Usage: Error in command line arguments\n";

my $cmd1 = "perl script.pl $input1 $param $output1 $selector";
my $cmd2 = "perl script.pl $input1 $param $output1 $selector";

if ($selector == "1"){
    ..traitements divers...
    #Info pour les biologistes
    print STDOUT "Step 1) - Commenter les traitements effectués : \n\n $cmd1 \n\n ";
}
else
{
    ..traitements divers...
    #Info pour les biologistes
    print STDOUT "Step 2) - Commenter les traitements effectués : \n\n $cmd2 \n\n ";
}

#Si besoin :
#TEST 1 : command ligne on vm-galaxy
#TEST 2 perl Galaxy file : perl script.pl p
$cmd3 = "(cd $dirresults; sh $BIN parametes
system $cmd3;

#Recuperation des fichiers par Galaxy
if (! -e "$dirresults/file.fasta"){print STDERR "file.fasta not found. \n\n";}else{'cp -a $dirresults/file.fasta $output1';}
..etc...
```

Path d'accès aux outils  
(bio)informatiques ou scripts wrappés.

Récupération d'un identifiant unique  
Création d'un répertoire de travail pour  
maîtriser le chemin d'accès aux fichiers  
(ou unlink/move)

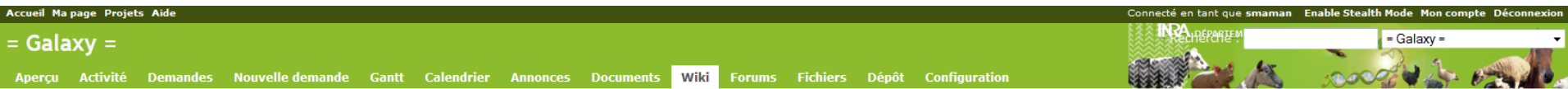
Si besoin, modification de l'extension  
des inputs (.dat par défaut)

Création d'un fichier de logs  
supplémentaire





# Wrapper type disponibles depuis le wiki Galaxy



- simple\_template.pl : [https://forge-dga.jouy.inra.fr/attachments/download/3191/simple\\_template.pl](https://forge-dga.jouy.inra.fr/attachments/download/3191/simple_template.pl)
- simple\_tool\_template.xml: [https://forge-dga.jouy.inra.fr/attachments/download/3192/simple\\_tool\\_template.xml](https://forge-dga.jouy.inra.fr/attachments/download/3192/simple_tool_template.xml)
- template.pl : <https://forge-dga.jouy.inra.fr/attachments/download/3193/template.pl>
- template\_podusage.pl : [https://forge-dga.jouy.inra.fr/attachments/download/3194/template\\_podusage.pl](https://forge-dga.jouy.inra.fr/attachments/download/3194/template_podusage.pl)

## Groupe de travail sur l'instance Sigena de Galaxy

Bonnes pratiques :

- Dev ou modification de wrappers en local sur vm-galaxy-test ---> par les personnes qui souhaitent développer des wrappers (accès actuels : Ibou, Olivier, Frederic, Maria...)
- Puis mise en production (sur vm-galaxy et vm-galaxy-dev) dans un second temps, une fois que les modifications se sont bien déroulées en local (par les admins : Sarah, Patrice, Marie-Stéphane, Didier).

### Comment tester vos wrappers sur vm-galaxy-test ?

SVN Galaxy

Voir aussi la section "wrapper type" de ce wiki.

\*[[Slides : comment wrapper ?]]

# Ajouter un outil

1. Créer un sous répertoire vosInitiales\_tool/ dans galaxy-dist/tools/
2. Placer les fichiers file.xml et wrapper.pl (ou .py, ou .R ..) dans tools/ vosInitiales\_tool/
3. Si votre tool utilise :
  - Un outil (bio)infomatique : préciser son nom, sa version et s'il est disponible sur le cluster Genotoul.
  - Un script : donner son chemin d'accès et le copier dans /save/galaxy\*/scripts/

# Description du fichier tool\_conf.xml

Ce fichier décrit la toolbox : le menu de gauche de l'interface Galxy.

```
<?xml version="1.0"?>
<toolbox>

  <Label text="Label1" id="feat" />

  <section name="Get Data" id="get_data">
    <tool file="data_source/upload.xml" />
    <tool file="sm_upload/genotoul_upload.xml" />
    <tool file="data_source/ucsc_tablebrowser.xml" />
    <tool file="data_source/ucsc_tablebrowser_test.xml" />
    <tool file="data_source/ucsc_tablebrowser_archaea.xml" />
    <tool file="data_source/microbial_import.xml" />
    <tool file="data_source/biomart.xml" />
    <!--<tool file="sm_save/save.xml"/>-->
  </section>

  <section name="Text Manipulation" id="textutil">
    <tool file="stats/column_maker.xml" />
    <tool file="filters/fixValueColumn.xml" />
    <tool file="filters/catWrapper.xml" />
    <tool file="filters/cutWrapper.xml" />
    <tool file="filters/mergeCols.xml" />
    <tool file="filters/convert_characters.xml" />
    <tool file="filters/CreateInterval.xml" />
    <tool file="filters/cutWrapper.xml" />
    <tool file="filters/changeCase.xml" />
    <tool file="filters/pasteWrapper.xml" />
    <tool file="filters/remove_beginning.xml" />
    <tool file="filters/randomlines.xml" />
    <tool file="filters/headWrapper.xml" />
    <tool file="filters/tailWrapper.xml" />
    <tool file="filters/trimmer.xml" />
    <tool file="filters/wc_gnu.xml" />
    <tool file="filters/secure_hash_message_digest.xml" />
    <tool file="new_operations/tables_arithmetic_operations.xml" />
  </section>

</toolbox>
```

Label descriptif

L'outil est placé dans une section.  
Le path est relatif.

# Configuration de l'exécution des tools avec job\_conf.xml

```
<?xml version="1.0"?>
<job_conf>

  <!-- Plateformes d'execution des programmes -->
  <plugins>
    <plugin id="local" type="runner" load="galaxy.jobs.runners.local:LocalJobRunner" workers="4"/>
    <plugin id="sge" type="runner" load="galaxy.jobs.runners.drmaa:DRMAAJobRunner" workers="4"/>
  </plugins>

  <handlers>
    <handler id="main"/>
  </handlers>

  <!-- Chaque destination doit correspondre a une plateforme d'execution-->
  <destinations default="real_user_cluster">
    <destination id="local" runner="local"/>
    <destination id="real_user_cluster" runner="sge">
      <param id="galaxy_external_runjob_script">scripts/drmaa_external_runner.py</param>
      <param id="galaxy_external_killjob_script">scripts/drmaa_external_killer.py</param>
      <param id="galaxy_external_chown_script">scripts/external_chown_script.py</param>
    </destination>
  </destinations>

  <!-- Forcer des outils s'executer sur unse seule destination -->
  <tools>
    <tool id="genotoul_upload" destination="local"/>
  </tools>

</job_conf>
```

Plateformes d'exécution

Configuration des destinations

1outil = 1 destination unique

Le fichier job\_conf.xml permet de spécifier les paramètres du qsub finement :

```
<destination id="gatk_variant_select_job" runner="drmaa">
  <param id="galaxy_external_runjob_script">scripts/drmaa_external_runner.py</param>
  <param id="galaxy_external_killjob_script">scripts/drmaa_external_killer.py</param>
  <param id="galaxy_external_chown_script">scripts/external_chown_script.py</param>
  <param id="nativeSpecification">-clear -q workq -l mem=10G -l h_vmem=50G -pe parallel_smp 4</param>
</destination>
```

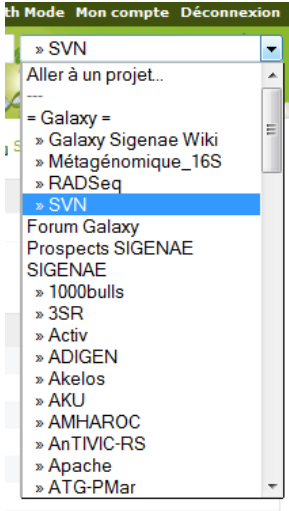
# Tests avant la mise en production

Avant de partager vos wrappers via le SVN en vu d'une mise en production :

1. Tester vos outils sur vm-galaxy-test avec un reload de l'outil via l'interface web et, si besoin, un restart de la machine.
2. Si votre xml n'apparaît pas dans l'interface web, parcourez de nouveau les tags de ce fichier xml puis re-tester.
3. Si vous changez le code du wrapper.pl , il n'est pas nécessaire de faire un reload ni un restart, les modifications sont automatiquement prises en compte.

# SVN

[https://forge-dga.jouy.inra.fr/svn/svn\\_galaxy\\_sigenae](https://forge-dga.jouy.inra.fr/svn/svn_galaxy_sigenae)



Accueil Ma page Projets Aide

= Galaxy = >> SVN

Aperçu Activité Annonces Documents Fichiers **Dépôt** Configuration

root

|   | Nom               | Taille | Révision | Âge               | Auteur        |                        |
|---|-------------------|--------|----------|-------------------|---------------|------------------------|
| ▣ | Metagenomique_16S |        | 12       | environ 17 heures | Maria Bernard | mv sm_mothur_preproces |

## SVN Galaxy

### Import du svn

```
svn co https://forge-dga.jouy.inra.fr/svn/svn_galaxy_sigenae/
```

### Ajout d'un dossier

```
svn import dossier_a_importer https://forge-dga.jouy.inra.fr/svn/svn_galaxy_sigenae/NOUVEAU_DOSSIER_DANS_SVN -m "MESSAGE"
```

### Modification des fichiers et propagation au SVN

```
svn add/del fichier  
svn commit
```

### Mise à jours

```
svn status  
svn update
```

# Administration de Galaxy via l'interface web

## Administration

### Security

- [Manage users](#)
- [Manage groups](#)
- [Manage roles](#)

### Data

- [Manage quotas](#)
- [Manage data libraries](#)

### Server

- [Tool versions](#)
- [Reload a tool's configuration](#)
- [Profile memory usage](#)
- [Manage jobs](#)

### Tool sheds

- [Search and browse tool sheds](#)

### Form Definitions

- [Manage form definitions](#)

### Sample Tracking

- [Manage sequencers and external services](#)
- [Manage request types](#)
- [Sequencing requests](#)
- [Find samples](#)

## Users

### Manage users

- Liste des utilisateurs

### Manage groups

- Associer un utilisateur à un groupe

### Manage roles

- Associer un utilisateur à un rôle

## Data

### Manage data libraries

- Définir les permissions afin de permettre à certains utilisateurs d'accéder à certaines données

## Server

### Reload a tool's configuration

- Pour rechargé un outil qui a été commenté dans le tool\_conf.xml (préférer le terminal)

### Profile memory usage

- Estimer la mémoire utilisée par certaines fonctions de Galaxy

### Manage jobs

- Lister les jobs trop longs (en vu d'un kill)

# Bonnes pratiques users

## Choix du navigateur

Il est préférable d'utiliser Firefox pour accéder à Galaxy.

## Organiser votre espace de travail pour maîtriser votre quota

•Le nombre de fichiers uploadés, d'outils lancés, d'historiques et de workflows créés avant de concevoir un traitement finalisé, peut être important.

L'aspect positif est que l'utilisateur peut manipuler et bien comprendre ses données.

L'aspect négatif est que cela génère beaucoup de fichiers (fichiers uploadés, fichiers intermédiaires, fichiers résultats) et surcharge ainsi le /work de galaxy.

•Télécharger vos données dans votre historique en utilisant, de préférence, dans la section "Get Data", l'outil " Upload local file from filesystem path" plutôt que l'outil "Upload File".

•Créer un nouvel historique par analyse pour classer vos traitements.

•Renommer et taguer vos fichiers résultats, vos historiques et vos workflows pour les retrouver facilement et surtout différencier rapidement les tests des fichiers définitifs. Ainsi, la suppression des tests sera facilitée.

•Supprimer définitivement (Purge Deleted ) et régulièrement les datasets, historiques et workflows inutiles.

## Des questions ?

•Via un ticket sur la forge DGA ou un mail ([sigenae-support@listes.inra.fr](mailto:sigenae-support@listes.inra.fr)).

•Une FAQ et un manuel utilisateur sont disponibles depuis la page d'accueil de l'instance Sigenae de Galaxy.

## Recommandations/ informations en fin de formation

•Les formations de la plateforme BIOINFO GENOTOUL sont disponibles sur <http://sig-learning.toulouse.inra.fr>

•Nettoyer votre compte de formation : "Delete permanently" de l'ensemble des "histories" créés.