

## 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 ▾



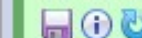
TP FastQC



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



# Vos traitements bioinformatiques avec GALAXY

***Sarah Maman, Ibouniyamine Nabihoudine  
Céline Noirot, Olivier Rué***



# Galaxy project

- Center for Comparative Genomics and Bioinformatics
- Département « Biology » et « Mathematics and Computer Science » (Emory)
- Publication :

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

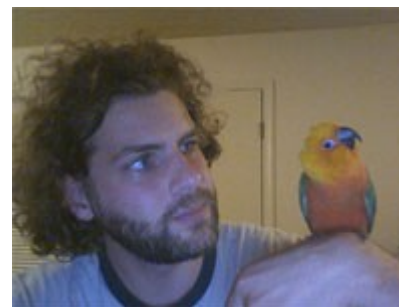
Jeremy Goecks<sup>1</sup>, Anton Nekrutenko<sup>2</sup>, James Taylor<sup>1</sup> and The Galaxy Team



Anton Nekrutenko  
Penn State



Nate Coraor  
Penn State



James Taylor  
Emory



EMORY  
UNIVERSITY



# La communauté

- Nationale et internationale très active
- Liste de diffusion / Wiki / Twitter
- « Galaxy Tour de France »
- Groupe de travail Galaxy IFB
  - Documentation collaborative
  - Formations (mise en commun agenda)
  - Architecture
  - Intégration d'outils (ToolShed)
  - <http://www.ifb.galaxy.org>





# Une « Galaxy » parmi tant d'autres

- **Serveur public**

- <https://main.g2.bx.psu.edu>
- Quotas limité, petits jeux de données
- Impossible d'ajouter des banques, génomes ou outils
- Données non protégées

- **L'instance locale Sigenae de Galaxy**

- Maintenu par l'équipe Sigenae
- Intégration d'outils/scripts/génome etc.

- ...



# Une « Galaxy » parmi tant d'autres

- Chaque instance possède ses propres outils
- Le Toolshed permet de partager certains outils entre plusieurs instances



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



# Avantages

- Exécution sur le cluster sans passer par la ligne de commande
- Traitements parallèles
- Partage des résultats et des chaînes de traitement
- Possibilité de demander l'ajout d'outils ([sigenae-support@liste.inra.fr](mailto:sigenae-support@liste.inra.fr))
- Pas d'utilisation du disque dur de votre PC



# Limites

- Les limites et règles d'utilisation sont celles de genotoul
- Le quota est vite atteint !
- L'organisation des données doit être rigoureuse !



# L'interface de Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help

Welcome smaman, you are working in /

**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



**History** Options ▾

TP FastQC 54.0 Mb

8: <u>FastQC data 5.html</u>	👁	✍	✕
6: <u>GM.fastqsanger</u>	👁	✍	✕
5: <u>h1.fastqsanger</u>	👁	✍	✕
4: <u>FastQC data 18.html</u>	👁	✍	✕
3: <u>FASTQ Summary Statistics on data 18</u>	👁	✍	✕
2: <u>FASTQ Summary Statistics on data 18</u>	👁	✍	✕
1: <u>GM.fastqsanger</u>	👁	✍	✕



# Exercice n°1 : Prise en main de l'interface



Analyze Data Workflow Shared Data Visualization Admin Help

Welcome smaman, you are working in /

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



History Options

TP FastQC 54.0 Mb

8: <u>FastQC data 5.html</u>	👁️ ✎ ✕
6: <u>GM.fastqsanger</u>	👁️ ✎ ✕
5: <u>h1.fastqsanger</u>	👁️ ✎ ✕
4: <u>FastQC data 18.html</u>	👁️ ✎ ✕
3: <u>FASTQ Summary Statistics on data 18</u>	👁️ ✎ ✕
2: <u>FASTQ Summary Statistics on data 18</u>	👁️ ✎ ✕
1: <u>GM.fastqsanger</u>	👁️ ✎ ✕

- **Boite a outils** : liste des outils disponibles



Analyze Data Workflow Shared Data Visualization Admin Help

Welcome smaman, you are working in /

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

History Options

TP FastQC 54.0 Mb

8: FastQC data 5.html eye edit delete

6: GM.fastqsanger eye edit delete

5: h1.fastqsanger eye edit delete

4: FastQC data 18.html eye edit delete

3: FASTQ Summary  
Statistics on data 18 eye edit delete

2: FASTQ Summary  
Statistics on data 18 eye edit delete

1: GM.fastqsanger eye edit delete

- Boite a outils : liste des outils disponibles
- **Espace de travail** : visualisation et configuration d'un outils en particulier, visualisation des données générées

Analyze Data Workflow Shared Data Visualization Admin Help

Welcome smaman, you are working in /

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

History Options

TP FastQC 54.0 Mb

8: FastQC data 5.html eye edit delete

6: GM.fastqsanger eye edit delete

5: h1.fastqsanger eye edit delete

4: FastQC data 18.html eye edit delete

3: FASTQ Summary Statistics on data 18 eye edit delete

2: FASTQ Summary Statistics on data 18 eye edit delete

1: GM.fastqsanger eye edit delete

- Boite a outils : liste des outils disponibles
- Espace de travail : visualisation et configuration d'un outils en particulier, visualisation des données générées
- **Historique** : liste de l'ensemble des résultats (bon ou mauvais) générés ainsi que l'ensemble des traitements effectuées

Analyze Data Workflow Shared Data Visualization Admin Help

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

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

1: GM.fastqsanger

- Boite a outils : liste des outils disponibles
- Espace de travail : visualisation et configuration d'un outils en particulier, visualisation des données générées
- Historique : liste de l'ensemble des résultats (bon ou mauvais) générés ainsi que l'ensemble des traitements effectuées
- **Menu**



# Vocabulaire spécifique

- **TOOL**: outil présent dans la boîte à outils d'une instance galaxy

▪ [Upload File](#) from your computer



# Vocabulaire spécifique

- **TOOL**: outil présent dans la boîte à outils d'une instance galaxy
- Un **DATASET** est une donnée résultant de l'exécution d'un tool

▪ [Upload File](#) from your computer





# Vocabulaire spécifique

- **TOOL**: outil présent dans la boîte à outils d'une instance galaxy
- Un **DATASET** est une donnée résultant de l'exécution d'un tool

▪ [Upload File](#) from your computer

Analyse en attente



# Vocabulaire spécifique

- **TOOL**: outil présent dans la boîte à outils d'une instance galaxy
- Un **DATASET** est une donnée résultant de l'exécution d'un tool

▪ [Upload File](#) from your computer

Analyse en attente

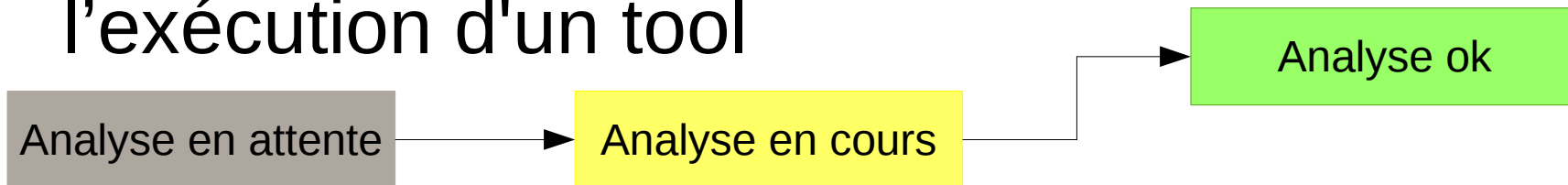
Analyse en cours



# Vocabulaire spécifique

- **TOOL**: outil présent dans la boîte à outils d'une instance galaxy
- Un **DATASET** est une donnée résultant de l'exécution d'un tool

▪ [Upload File](#) from your computer

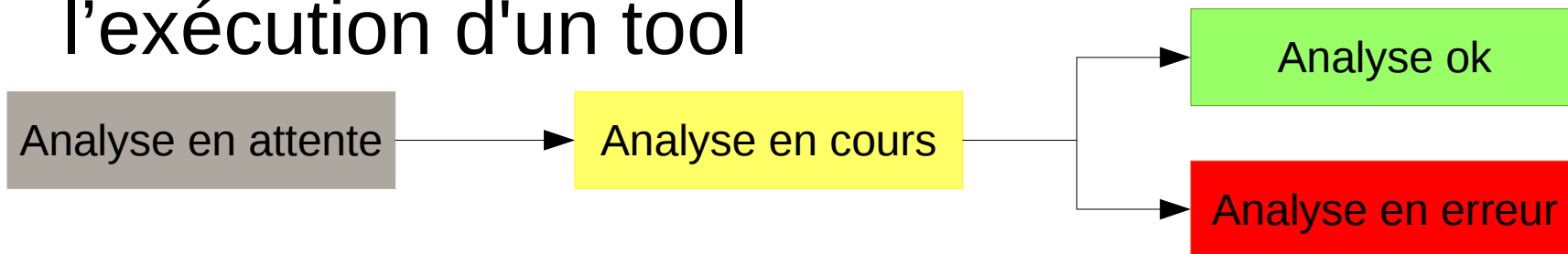




# Vocabulaire spécifique

- **TOOL**: outil présent dans la boîte à outils d'une instance galaxy
- Un **DATASET** est une donnée résultant de l'exécution d'un tool

▪ [Upload File](#) from your computer



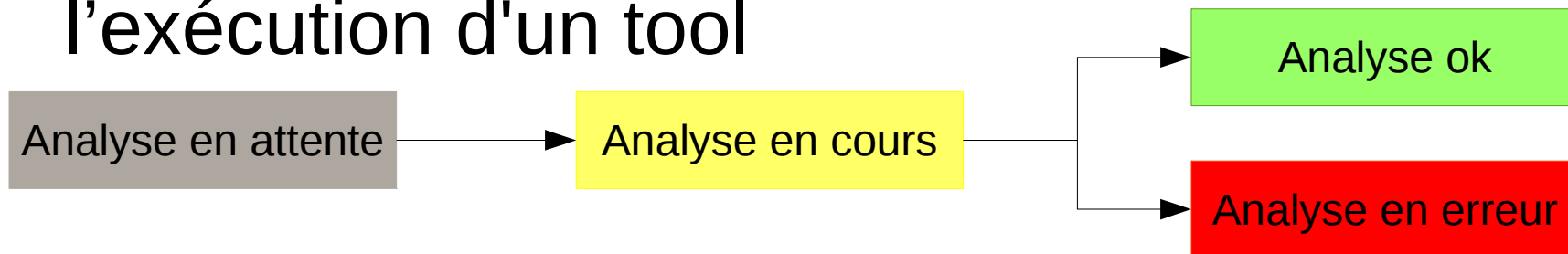


# Vocabulaire spécifique

- **TOOL**: outil présent dans la boîte à outils d'une instance galaxy

▪ [Upload File](#) from your computer

- Un **DATASET** est une donnée résultant de l'exécution d'un tool



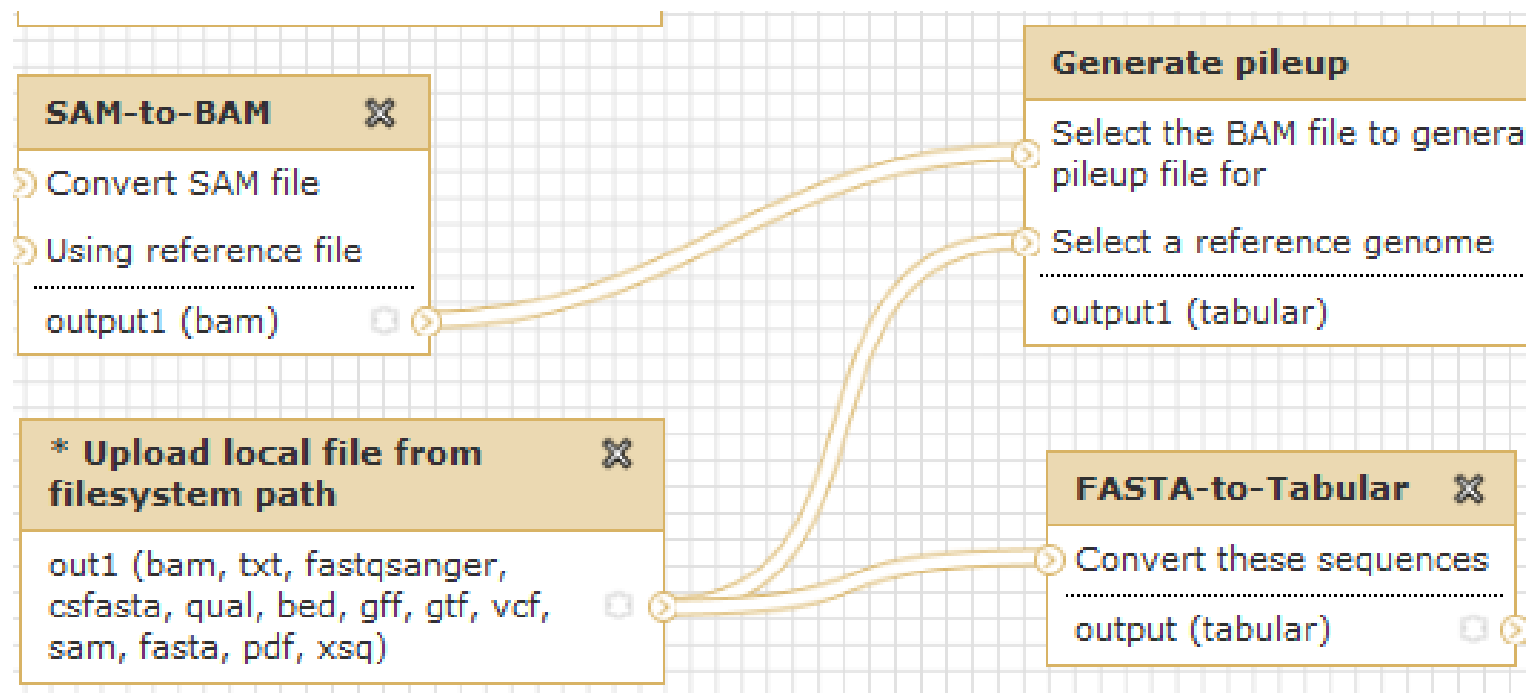
- L'**HISTORIQUE** liste de l'ensemble des résultats (bon ou mauvais) générés ainsi que l'ensemble des traitements effectués (cachés)





# Vocabulaire spécifique

- Un **Workflow** est une suite d'exécution d'outils



# Comment fonctionne Galaxy

Envoi de données  
Exécution d'outils

Exécution sur SGE



Serveur web  
galaxy

Cluster de calcul

Récupération de  
résultats

Fin d'exécution

Fin d'exécution





# Liste des historiques et exécution des jobs

The screenshot shows the Galaxy Sig interface with a 'Saved Histories' table. The table lists various history entries with columns for Name, Datasets, Tags, and Sharing. Annotations in colored boxes indicate the status of the analysis for each entry: 'Analyse OK' (green), 'Analyse en attente' (grey), 'Analyse en erreur' (red), and 'Analyse en cours' (yellow).

Name	Datasets	Tags	Sharing	Status
TP Galaxy project	2	1	0 Tags	Analyse en attente
miRNA tests	59	21	0 Tags	Analyse en erreur
TP SNPs calling	84	9	0 Tags	Analyse en erreur
TP RNAseq	88	1	0 Tags	Analyse en erreur
test TP miRNA	36	1	0 Tags	Analyse en cours
Unnamed history			0 Tags	Analyse en erreur

Annotations:

- Analyse OK** (green box) points to the 'Datasets' column of 'TP Galaxy project'.
- Analyse en attente** (grey box) points to the 'Tags' column of 'TP Galaxy project'.
- Analyse en erreur** (red box) points to the 'Tags' column of 'miRNA tests' and 'TP SNPs calling'.
- Analyse en cours** (yellow box) points to the 'Datasets' column of 'test TP miRNA'.

Right sidebar details for 'TP Galaxy project':

- 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





# Bonnes pratiques

- Quota utilisateur à surveiller
- Suppression des datasets et historiques inutiles
- La suppression peut être temporaire ou permanente



# Bonnes pratiques

The screenshot displays the Galaxy web interface. At the top, there is a navigation bar with tabs: Workflow, Shared Data, Visualization, Admin, Help, User, and a welcome message for 'smaman' with a 'Using 30%' indicator. Below the navigation bar is a table listing various workflows. The table has columns for workflow name, a green box with a number, a red box with a number, '0 Tags', and a size. A context menu is open over the table, listing various actions. The 'Purge Deleted Datasets' option is circled in red. At the bottom of the table, there are buttons for 'Rename', 'Delete', 'Delete Permanently', and 'Undelete'.

Workflow	Green Box	Red Box	Tags	Size
crippmark	1	3	0 Tags	328.0
: fichiers abs du	4		0 Tags	3.6 G
tation SNP	17		0 Tags	2.4 M
rted: anTargetCreator	6	10	0 Tags	3.6 G
Mirdeep2 sans ination de la ndance intra ni inter	51		0 Tags Accessible	12.8
ory archive	4		0 Tags Shared	0 byt
rted: Unnamed ry	6		0 Tags	742.0
iere session ation Galaxy	21		0 Tags	1.4 G
- GALAXY	16		0 Tags	1.6 G
NGS - norphisme	14	6	0 Tags Shared	0 byt
NGS RNA Analysis	4	2	0 Tags	41.1

selected histories: **Rename** **Delete** **Delete Permanently** **Undelete**

Context Menu Options:

- 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



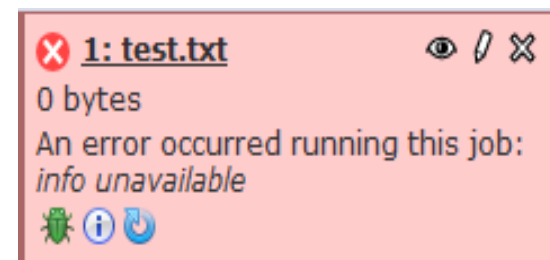
# Bonnes pratiques

- Renommer son historique
- Renommer les datasets
- Purger l'historique assez souvent



# Support

- Votre Dataset est en erreur, voici les informations a transmettre à [sigenae-support@liste.inra.fr](mailto:sigenae-support@liste.inra.fr)
  - Copie du message
  - Partager l'historique si besoin



## Dataset generation errors

Dataset 3: ContigLengthG1000ProfG8.res

Tool execution generated the following error message:

```

Traceback (most recent call last):
  File "/usr/local/bioinfo/src/galaxy/galaxy-dist/tools/sm_clones/scripts_module2/ace_statistics.py", line 98, in <module>
    pit.clif()
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/site-packages/matplotlib/pyplot.py", line 443, in clif
    gcf().clif()
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/site-packages/matplotlib/pyplot.py", line 369, in gcf
    return figure()
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/site-packages/matplotlib/pyplot.py", line 343, in figure
    **kwargs)
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/site-packages/matplotlib/backends/backend_tkagg.py", line 80, in new figure
    window = Tk.Tk()
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/lib-tk/Tkinter.py", line 1685, in init
    self.tk = tkinter.create(screenName, baseName, className, interactive, wantobjects, useTk, sync, use)
tkinter.TclError: no display name and no $DISPLAY environment variable

```

The tool produced the following additional output:

```

Votre repertoire de travail : /work/galaxy/database/files/workspace/38401
Epilog : job finished at mar. avril 1 15:47:29 CEST 2014

```

## Report this error to Sigene Team

Please create a ticket in [Redmine](#) or send a mail to [Galay administrator](#).

### Tool: Clone metagenomic

Name:	ContigLengthG1000ProfG8.res
Created:	Apr 01, 2014
Filesize:	134 bytes
Dbkey:	?
Format:	txt
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr

### Input Parameter

Other ace files  
Other ace files

### Inheritance Chain

ContigLengthG1000ProfG8.res



# Citer Galaxy



## How to cite Galaxy workbench ?

Depending on the help provided you can cite us in acknowledgements, references or both.

Examples :

Acknowledgements

We wish to thank the SIGENAE group for ....

References

X. SIGENAE [<http://www.sigenae.org/>]

## 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):1450-1456.

# 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,**



## 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 [Sigeneae e-learning platform](#)."

# Exercices