

Galaxy: ressource pour le calcul en bioinformatique

Plateau Bioinformatique du CIRAD
pour la plateforme Southgreen

- Introduction
- Présentation générale de galaxy
- Lancer une analyse avec galaxy
- Les workflows
- Exercices
- FAQ

- **OBJECTIFS:**

- Savoir faire une analyse via galaxy
- Savoir gérer les résultats d'une analyse
- Savoir partager des résultats d'analyse
- Savoir créer et partager un protocole d'analyse

Galaxy en bref

- Conviviale : Outil web
- Accessible : aucune expérience en programmation n'est requise
- Reproductible : les utilisateurs peuvent facilement refaire une analyse (ou suivre un protocole)
- Echange (partage) : Les utilisateurs peuvent partager toutes leurs créations (données, protocole d'analyse, Pages explicatives)
- Evolutif : Ajout régulier de nouveaux outils, amélioration régulière de l'interface

Galaxy

Accès direct au cluster

Interface et compte utilisateur indépendants du poste client

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Using 0 bytes

Tools

Meta genomic analyses

- Fetch taxonomic representation
- Summarize taxonomy
- Draw phylogeny
- Find diagnostic hits
- Find lowest diagnostic rank
- Poisson two-sample test

FASTA manipulation

- Compute sequence length
- Filter sequences by length
- Concatenate FASTA alignment by species
- FASTA-to-Tabular converter
- Tabular-to-FASTA converts tabular file to FASTA format
- FASTA Width formatter
- RNA/DNA converter
- Collapse sequences

NGS: QC and manipulation

- FASTQC: FASTQ/SAM/BAM
- Fastqc: Fastqc QC using FastQC from Bahraham

South Green bioinformatics platform

Using 0 bytes

History

0 bytes

1: Galaxy2-[IqB6_GCCAAT_L003_R2.fastq].fastqsanger
empty
format: txt, database: ?
Info: The uploaded file is empty

Centralisation et partage des données et des analyses

Requests for making new tools available shall be addressed to admin.bioinfo@cirad.fr

The GALAXY project is supported in part by NSF NHGRI and the Huck Institutes of the Life Sciences.

Utilisable depuis n'importe quel poste connecté à internet

Galaxy, première analyse

Galaxy - Vue générale

Boîte à outils

Menu de navigation

The screenshot shows the Galaxy bioinformatics platform interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The left sidebar contains a 'Tools' menu with categories like 'Metagenomic analyses' and 'FASTA manipulation'. The main content area displays the 'South Green bioinformatics platform' logo and a 'Welcome to GALAXY' message. A red box highlights the main content area, labeled 'Panneau d'affichage (Configuration de programme ou visualisation des résultats)'. A dashed box in the main content area contains the text: 'Requests for making new tools available shall be addressed to admin.bioinfo@cirad.fr'. The right sidebar shows a 'History' section with a list of jobs, including one with the name 'Galaxy2-[IqB6_GCCAAT_L003_R2.fastq].fastqsanger'.

Panneau d'affichage
(Configuration de programme ou
visualisation des résultats)

Données/Résultats

Galaxy - Boîte à outils

The screenshot shows the Galaxy interface with a search bar at the top containing 'search tools'. Below it, there are sections for 'Recently Used', 'Get Data', and 'Send Data'. A dropdown menu is open, showing 'Show Tool Search' and 'Hide Recently Used'. The 'TOOLS' section is highlighted with a red circle and a red banner. Below it, various tool categories are listed, including 'Convert Formats', 'Evolution', 'ESTtik', 'Filter and Sort', 'Gene/Protein prediction', 'SAT', 'NGS: Quality Control', 'NGS: Mapping', 'NGS: SAM/BAM Manipulations', 'NGS: SNP Detection', 'Protein Structures', and 'Sequence comparisons'. The 'UNTESTED TOOLS' section is also highlighted with a red circle and a red banner. The South Green bioinformatics platform logo is visible at the bottom left.

Les outils sont nombreux, et classés par catégories

TOOLS Outils testés et maintenus par Southgreen

- La version original de Galaxy est livrée avec des dizaines d'outils.
- La plateforme Southgreen a ajouté des dizaines d'outils supplémentaires.
- Un outil de recherche par mot clef est disponible.

UNTESTED TOOLS Outils installés sur le Galaxy original

Galaxy - Vue générale

Galaxy Analyze Data Workflow Shared Data Visualization Help User

South Green Using 0 bytes

Tools

Meta-genomic analyses

- Fetch taxonomic representation
- Summarize taxonomy
- Draw phylogeny
- Find diagnostic hits
- Find lowest diagnostic rank
- Poisson two-sample test

FASTA manipulation

- Compute sequence length
- Filter sequences by length
- Concatenate FASTA alignment by species
- FASTA-to-Tabular converter
- Tabular-to-FASTA converts tabular file to FASTA format
- FASTA Width formatter
- RNA/DNA converter
- Collapse sequences

NGS: QC and manipulation

FASTQC: FASTQ/SAM/BAM

- Fastqc: Fastqc QC using FastQC from Bahraham

South Green bioinformatics platform

History

0 bytes

1: Galaxy2-[IqB6_GCCAAT_L003_R2.fastq].fastqsanger
empty
format: txt, database: ?
Info: The uploaded file is empty

**Panneau d'affichage
(Configuration de programme ou
visualisation des résultats)**

Requests for making new tools available shall be addressed to admin.bioinfo@cirad.fr

The GALAXY project is supported in part by NSF NHGRI and the Huck Institutes of the Life Sciences.

Galaxy - Vue générale

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Tools

Meta-genomic analyses

- Fetch taxonomic representation
- Summarize taxonomy
- Draw phylogeny
- Find diagnostic hits
- Find lowest diagnostic rank
- Poisson two-sample test

FASTA manipulation

- Compute sequence length
- Filter sequences by length
- Concatenate FASTA alignment by species
- FASTA-to-Tabular converter
- Tabular-to-FASTA converts tabular file to FASTA format
- FASTA Width formatter
- RNA/DNA converter
- Collapse sequences

NGS: QC and manipulation

- FASTQC: FASTQ/SAM/BAM
- Fastqc: Fastqc QC using FastQC from Bahraham

South Green bioinformatics platform

IRD Institut de recherche pour le Développement

Bioversity

INRA

Montpellier SupAgro

Welcome to GALAXY

... at your disposal as part of the services provided by **SouthGreen**

In order to **figure out which tools were made available by our team**, please activate the "tool search" functionality from the Options drop-down and type "southgreen" in the lookup filter.

Requests for **making new tools available** shall be addressed to admin.bioinfo@cirad.fr

The GALAXY project is supported in part by [NSF NHGRI](#) and [the Huck Institutes of the Life Sciences](#).

History

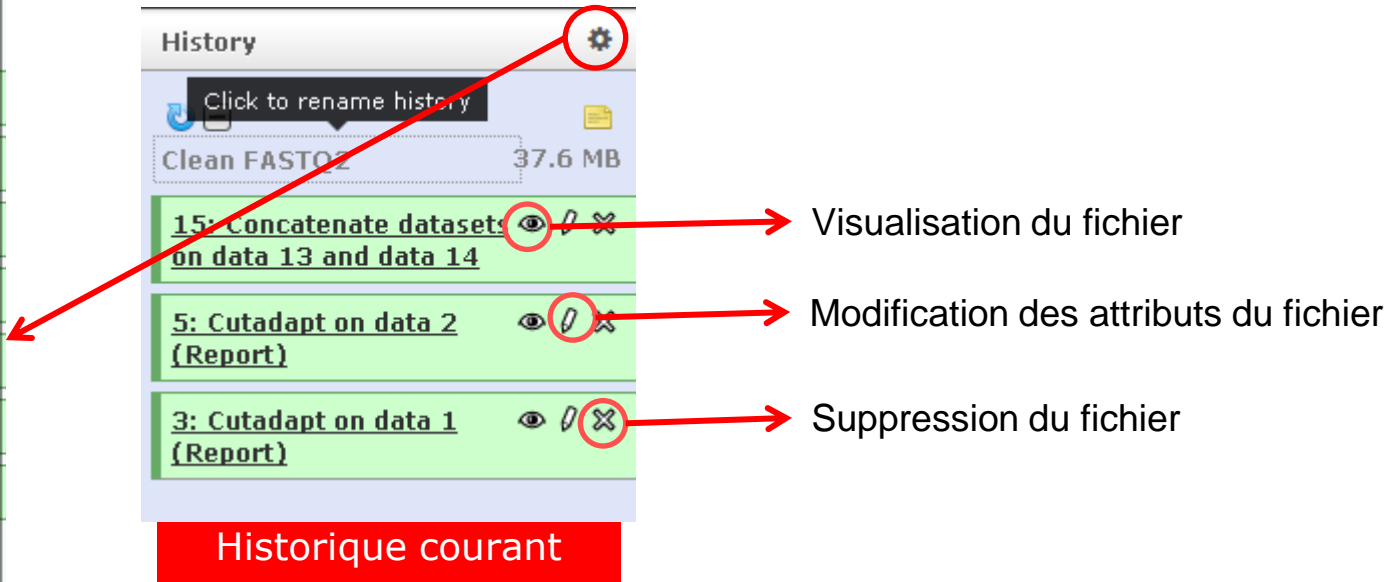
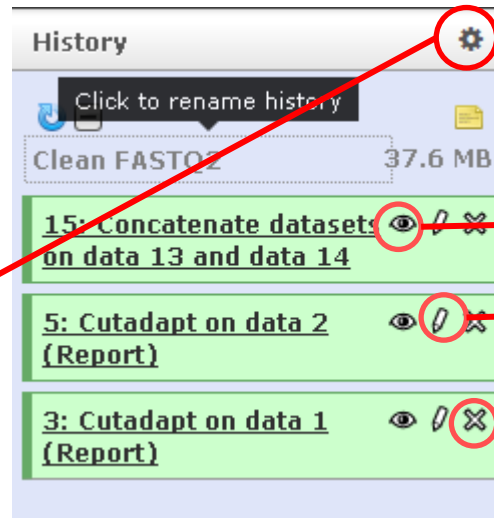
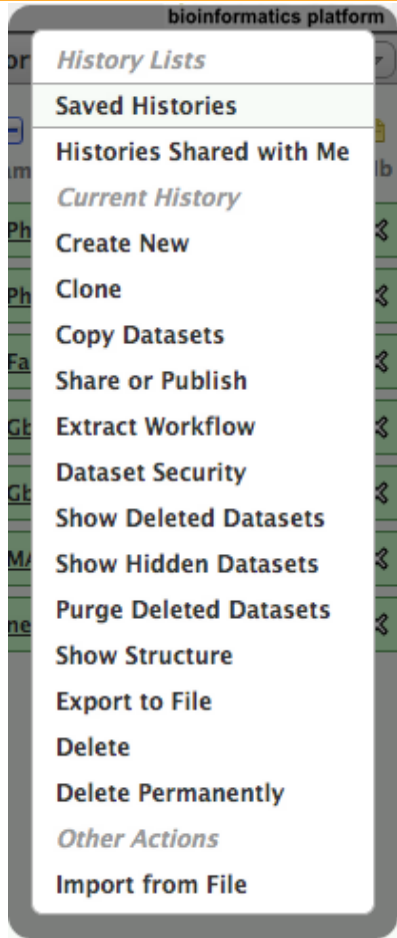
Using 0 bytes

bytes

1: Galaxy2-[IqB6_GCCAAT_L003_R2.fastq].fastqsanger
empty
format: txt, database: ?
Info: The uploaded file is empty

Données/Résultats

Galaxy - Historiques des analyses



La notion d'historiques multiples est très importante pour gérer ses résultats

Ne jamais utiliser de caractères accentués dans le nom des fichiers ou historiques ainsi que dans les annotations.

Galaxy - Historiques des analyses

bioinformatics platform

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

Listes des historiques personnels

Saved Histories

search history names and tags

Advanced Search

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated
Formation fevrier	28	0 Tags		0 bytes	Feb 09, 2012	less than a minute ago
Analyse NSLTps	11	0 Tags		98 bytes	Aug 22, 2011	1 minute ago
Unnamed history	7	0 Tags		4.0 Mb	Apr 18, 2012	Apr 23, 2012
Clone of 'Unnamed history' shared by 'dominique.this@supagro.inra.fr' (active items only)	150	1	0 Tags	56.1 Mb	Apr 23, 2012	Apr 23, 2012
Unnamed history	3	0 Tags		0 bytes	Feb 09, 2012	Feb 09, 2012

Listes des historiques publiés

Galaxy | Analyze Data | Workflow | Shared Data | Admin

Published Histories

search name, annotation, owner, and tag

Advanced Search

Data Libraries

Published Histories

Published Workflows

Name	Annotation	Owner	Community Rating	Community
RC2		rogeriomercres	★★★★★	
RC2_Formation28		reynaldo	★★★★★	
imported: RC4_workflow		mruiZ	★★★★★	
RC3_alexis		mruiZ	★★★★★	
Unnamed history		formation12	★★★★★	

Galaxy - Historiques des analyses

Galaxy utilise des codes couleurs pour la gestion des fichiers

The screenshot displays the Galaxy web interface with a 'Saved Histories' table. The table has columns for Name, Datasets, Tags, Sharing, Size on Disk, Created, Last Updated, and Status. The 'Datasets' column contains color-coded boxes with numbers: green (13), yellow (2), white (2), and red (7). Red callout boxes with arrows point to these colors, labeling them as 'Analyse OK', 'Analyse en cours', and 'Analyse en erreur' respectively. The right sidebar shows a 'History' panel with details for various analysis jobs, including '79: Output file, snp information', '78: Output file, snp genotype MAF', '77: Output file, snp information', '76: Output file, snp genotype MAF', and '75: Output file, snp information'.

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated	Status
Unnamed history	13	0 Tags		65.2 MB	Dec 21, 2012	~ 23 hours ago	OK
arcad_workflow_big_d	2	0 Tags		3.7 GB	May 07, 2012	Dec 21, 2012	In Progress
gwas_test	2	0 Tags		39.9 MB	Dec 13, 2012	Dec 21, 2012	OK
galaxy_test	20	0 Tags		52.1 MB	May 04, 2012	Nov 13, 2012	OK
Unnamed history	7	0 Tags		0 bytes	Oct 01, 2012	Oct 01, 2012	Error

Galaxy - Historiques des analyses

South Green
bioinformatics platform

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

Saved Histories

search history names
Advanced Search

Cliquez ici pour afficher
le panneau de gestion
des historiques

<input type="checkbox"/>	Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated	Status
<input type="checkbox"/>			0 Tags		0 bytes	less than a minute ago	less than a minute ago	current history
<input type="checkbox"/>			0 Tags		0 bytes	3 days ago	~ 4 hours ago	
<input type="checkbox"/>	IqC3 combine ▾	4	0 Tags		590.0 Mb	~ 6 hours ago	~ 5 hours ago	
<input type="checkbox"/>	IqO10 Clean Data ▾	4	1 0 Tags		38.6 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO9 Clean Data ▾	4	1 0 Tags		38.0 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO8 Clean Data ▾	4	1 0 Tags		28.9 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO7 Clean Data ▾	4	1 0 Tags		25.4 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO6 Clean Data ▾	5	0 Tags		26.9 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO5 Clean Data ▾	4	1 0 Tags		31.9 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO4 Clean Data ▾	4	1 0 Tags		35.2 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO3 Clean Data ▾	4	1 0 Tags		27.7 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO2 Clean Data ▾	4	1 0 Tags		28.6 Gb	Jun 04, 2012	3 days ago	
<input type="checkbox"/>	IqO1 Clean Data ▾	4	1 0 Tags		30.3 Gb	May 31, 2012	3 days ago	
<input type="checkbox"/>	arcad_test_data ▾	3	1 0 Tags		2.6 Gb	Apr 24, 2012	Jun 11, 2012	
<input type="checkbox"/>	Cleaning_NGS_felix ▾	4	1 0 Tags			May 07, 2012	May 07, 2012	
<input type="checkbox"/>	galaxy_test ▾	17	1 0 Tags			May 04, 2012	May 07, 2012	

Selectionnez les
historiques à
supprimer

Suppression
temporaire

Suppression
définitive

For 0 selected histories:

Galaxy - Historiques des analyses

Saved Histories

Close Advanced Search

name:

tags:

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

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

**Panneau de gestion
des historiques**

**Cliquez ici pour voir les
historiques temporairement
supprimés**

<input type="checkbox"/>	Name		Size on Disk	Created	Last Updated	Status
<input type="checkbox"/>	Unnamed hist		bytes	less than a minute ago	less than a minute ago	current history
<input type="checkbox"/>	Unnamed history	2	0 bytes	3 days ago	~ 4 hours ago	
<input type="checkbox"/>	IGC3 combine	4	590.0 Mb	~ 6 hours ago	~ 5 hours ago	
<input type="checkbox"/>	IqO10 Clean Data	4	38.6 Gb	Jun 04, 2012	3 days ago	1
<input type="checkbox"/>	IqO9 Clean Data	4	38.0 Gb	Jun 04, 2012	3 days ago	1
<input type="checkbox"/>	IqO8 Clean Data	4	28.9 Gb	Jun 04, 2012	3 days ago	1
<input type="checkbox"/>	IqO7 Clean Data	4	25.4 Gb	Jun 04, 2012	3 days ago	1
<input type="checkbox"/>	IqO6 Clean Data	5	26.9 Gb	Jun 04, 2012	3 days ago	0 Tags
<input type="checkbox"/>	IqO5 Clean Data	4	31.9 Gb	Jun 04, 2012	3 days ago	1
<input type="checkbox"/>	IqO4 Clean Data	4	35.2 Gb	Jun 04, 2012	3 days ago	1
<input type="checkbox"/>	IqO3 Clean Data	4	27.7 Gb	Jun 04, 2012	3 days ago	1
<input type="checkbox"/>	IqO2 Clean Data	4	28.6 Gb	Jun 04, 2012	3 days ago	1
<input type="checkbox"/>	IqO1 Clean Data	4	30.3 Gb	May 31, 2012	3 days ago	1
<input type="checkbox"/>	arcad_test_data	3	2.6 Gb	Apr 24, 2012	Jun 11, 2012	1
<input type="checkbox"/>	Cleaning_NGS_felix	4	41.6 Mb	May 07, 2012	May 07, 2012	0 Tags
<input type="checkbox"/>	galaxy_test	17	3.2 Mb	May 04, 2012	May 07, 2012	0 Tags

Galaxy - accès aux données

Les données peuvent être chargées depuis votre ordinateur local

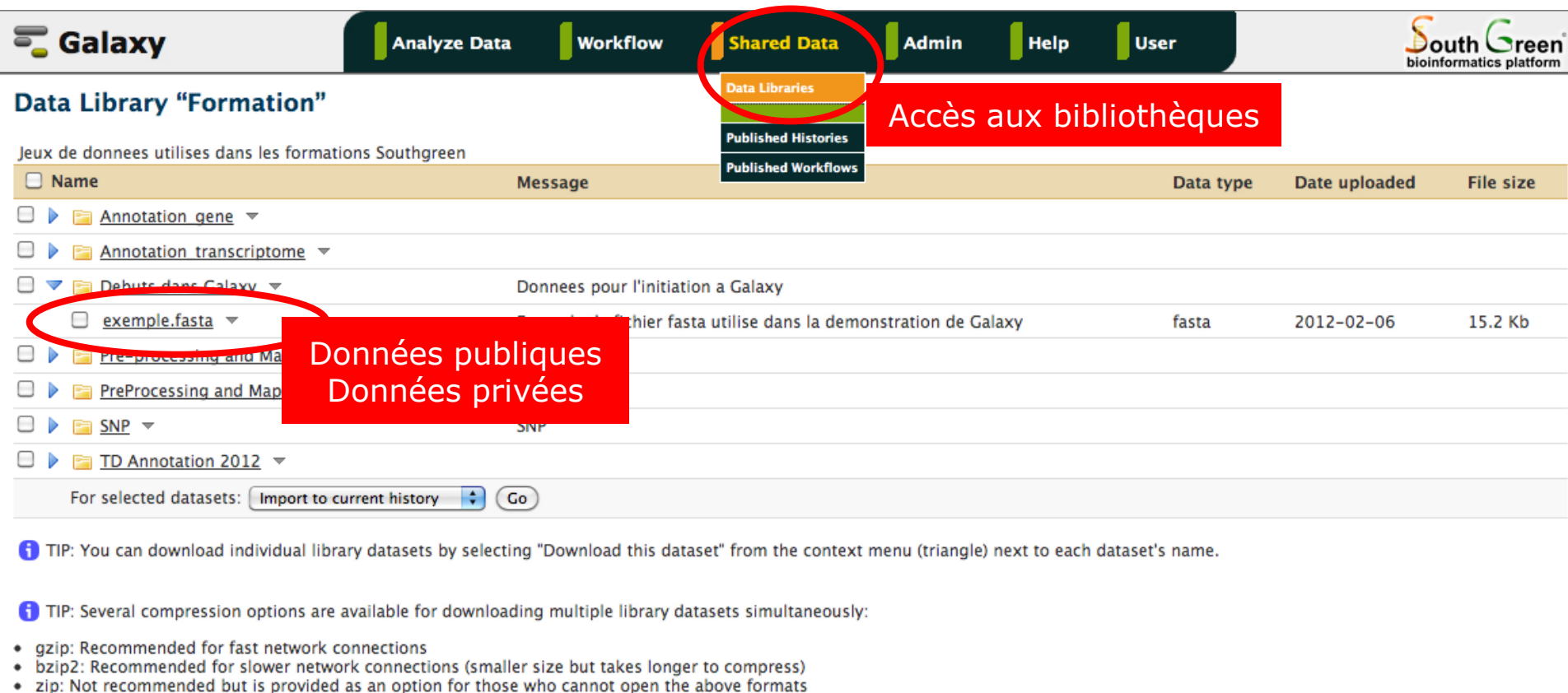
Pour importer depuis votre ordinateur local

Pour importer des données depuis une autre page web ou par copier/coller

Explication sur les différents formats

Galaxy - accès aux données

Les données peuvent être chargées depuis des bibliothèques partagées



The screenshot shows the Galaxy bioinformatics platform interface. At the top, there is a navigation bar with the following items: Galaxy logo, Analyze Data, Workflow, Shared Data (circled in red), Admin, Help, and User. The 'Shared Data' menu is open, showing three options: Data Libraries (highlighted in orange), Published Histories, and Published Workflows. A red box with the text 'Accès aux bibliothèques' is positioned over the 'Data Libraries' option.

Below the navigation bar, the page title is 'Data Library "Formation"'. The main content area displays a table of data libraries. The table has columns for Name, Message, Data type, Date uploaded, and File size. The 'Name' column contains a list of folders and files, including 'Annotation_gene', 'Annotation_transcriptome', 'Debuts dans Galaxy', 'exemple.fasta' (circled in red), 'Pre-processing and Mapping', 'PreProcessing and Mapping', 'SNP', and 'TD Annotation 2012'. A red box with the text 'Données publiques' and 'Données privées' is positioned over the 'exemple.fasta' entry.

At the bottom of the table, there is a section for 'For selected datasets:' with a dropdown menu set to 'Import to current history' and a 'Go' button. Below this, there are two informational tips (TIP) and a list of compression options:

- TIP: You can download individual library datasets by selecting "Download this dataset" from the context menu (triangle) next to each dataset's name.
- TIP: Several compression options are available for downloading multiple library datasets simultaneously:
 - gzip: Recommended for fast network connections
 - bzip2: Recommended for slower network connections (smaller size but takes longer to compress)
 - zip: Not recommended but is provided as an option for those who cannot open the above formats

Gestion fine des droits d'accès aux données

Workflow

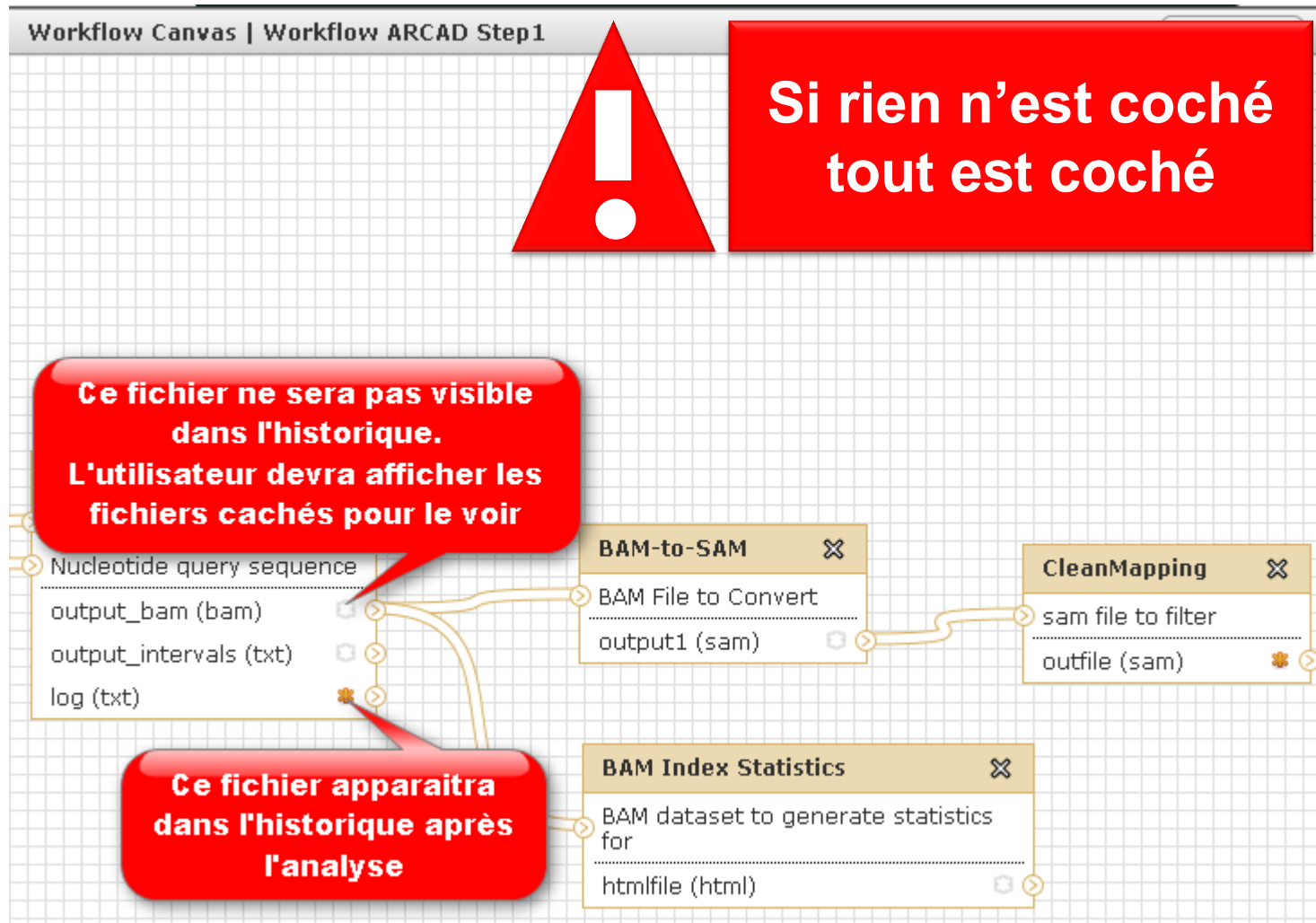
Galaxy - Workflows

Un workflow est un enchaînement d'outils paramétrés, ils sont partageables et publiables

The screenshot displays the Galaxy Workflow Canvas interface. At the top, the navigation bar includes 'Analyze Data', 'Workflow' (highlighted with a red circle), 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The main workspace is a grid-based 'Workflow Canvas' titled 'Workflow Canvas | Workflow ARCHS Step1 (imported from uploaded file)'. It contains several tool steps connected by arrows, including 'Input dataset', 'FASTQ Groomer', 'Cutadapt', 'Filter FASTQ', 'Concatenate datasets', 'Reformat read identifier', and 'FASTQ de-interlacer'. A red banner with the text 'Canvas d'édition' is overlaid on the central part of the canvas. On the right side, a red banner with the text 'Paramétrage' is overlaid on the tool configuration panel for 'FASTQ Groomer'. This panel shows various settings such as 'File to groom', 'Input FASTQ quality scores', 'Advanced Options', 'Output FASTQ quality scores', 'Force Quality Score encoding', and 'Summarize input data'. A 'Tools' sidebar on the left lists various categories like 'Get Data', 'Send Data', 'Convert Formats', etc. The bottom left corner features the 'South Green bioinformatics platform' logo.

Galaxy - Workflows

Lors de la construction d'un workflow vous pouvez décider des fichiers qui seront visibles dans l'historique après l'analyse.



Series d'exercices

Exercices – niveau 0

1. Créer un nouvel historique puis le renommer **exemple1**
2. Importer le fichier **exemple.fastq** depuis la bibliothèque partagée **shared data -> formation -> galaxy** dans votre historique.
3. Importer le fichier **adapters.txt** depuis le lien
4. Créer le fichier **qualite.qual** à partir du fichier **exemple.fastq**
5. Cloner votre nouvel historique, que fais la commande clone ?
6. Supprimer le clone précédemment créé.
7. Créer un autre historique (**exemple2**), **copier** les données de l'historique **exemple1** dans l'historique **courant**.
8. Afficher la liste de vos historiques et changer l'historique courant.
9. Publier votre historique, afin que tous le monde puisse y avoir accès.
10. Supprimer l'**historique** que vous venez de **publier** de la liste des **historiques publiés**. Vérifier que ce dernier existe toujours dans la liste de vos historiques personnels.

Exercices – niveau 1

1. Créer un nouvel historique puis le renommer **analyse1**, copier le fichier **exemple.fastq** et **adapters.txt**.
2. Faites un résumé du contenu de votre fichier en utilisant l'outil fastqc.
3. Utilisez l'outil **cutadapt** pour supprimer les adaptateurs contenu dans le fichier **adapters.txt**
4. Utiliser l'outil **filter fastq (ARCAD)** pour **supprimer**, les séquences qui ont une **taille inférieur à 60 et une qualité moyenne de 35**.
5. Refaites la question 2 en utilisant la sortie de la question précédente, voyez-vous des différences ?
6. Transformer le fichier **fastq** en **fasta**, puis faites un **blastn** sur nt, la **sortie doit être au format XML**.
7. Appliquer l'outil **XML4Blast2GO** sur le résultat du blast puis lancer **Blast2GO** sur le **résultat**.

1. A partir des questions « niveau 1 » créer un workflow qui va réaliser toutes les étapes.
2. Lancer votre **workflow**, sur le fichier de départ de l'exercice analyse. (n'oubliez pas de créer un **nouvel historique** et de le renommer **analyse_en_chaine**)
3. Observez-vous une différence entre les deux méthodes ?
4. Partager votre workflow avec l'utilisateur '**felix.homa@cirad.fr**'
5. Attendez le signal de l'instructeur pour arrêter le partage.

Merci de votre attention

