



## Formation Galaxy

Janvier 2017



## 1 – Premiers Pas...

### 1-1 Connexion

□ Connectez-vous sur la plateforme Galaxy SouthGreen à l'adresse suivante :

<http://galaxy.southgreen.fr/galaxy/>

Utiliser votre adresse email et votre mot de passe pour vous identifier. Si vous n'avez pas de compte, utilisez pour aujourd'hui le compte formationN/formationN

### 1-2 Import de fichier

□ Dans l'onglet Tools à gauche de l'interface allez dans « Get Data » puis « Upload File ».

*Galaxy permet d'importer un fichier via plusieurs moyens :*

- Importer un fichier stocké localement sur votre ordinateur en cliquant sur « choisissez un fichier »
- Importer un fichier à partir d'une url en copiant l'adresse dans le cadre « URL/Text »
- Copier directement le fichier dans le cadre « URL/Text »

Il est ensuite possible de visualiser, supprimer ou éditer les attributs d'un fichier.

Testez les 3 types d'import possibles de données. Vous observerez dans l'onglet « History » à droite de l'interface la progression de l'importation.

Galaxy offre un suivi de l'état de chaque job avec un code couleur :

- Bleu : le job a été soumis
- Jaune : le job est en cour de traitement
- Vert : le job s'est terminé avec succès
- Rouge : le job est en erreur

## Import des données de la librairie partagée

- Accédez aux données partagées (Shared data => Data libraries)
- Cliquez sur la library Formation
- Cochez les fichiers input.fastq et Reference.fasta
- Cliquez sur le bouton "To history" pour importer les données.

**Galaxy** Analyze Data Workflow Shared Data Visualization Admin Help User Using 54.3 GB

### Data Libraries

search dataset name, info, message, dbkey Q

[Advanced Search](#)

Data library name:	Data lib
<a href="#">454 Coffea</a>	
<a href="#">Amalia project</a>	Coffea Amalia project
<a href="#">Arcad</a>	
<a href="#">Arcad Cafe</a>	
<a href="#">Arcad Fonio</a>	
<a href="#">Arcad Mil</a>	
<a href="#">Arcad Monococcum</a>	
<a href="#">Arcad Vigne</a>	
<a href="#">Arcad Yam</a>	
<a href="#">Arcad Riz</a>	
<a href="#">Azucena</a>	
<a href="#">Banana</a>	
<a href="#">CIAT RICE</a>	
<a href="#">Clotault</a>	Personal data
<a href="#">cocos_chloroplast</a>	
<a href="#">cocos_fasta_files</a>	
<a href="#">cocos_fasta_files</a>	
<a href="#">coffea</a>	chromosome annotation
<a href="#">Coffea NBS</a>	
<a href="#">coffea_454</a>	coffea 454
<a href="#">Coffee Genome SNP</a>	

### 1-3 Exécution de job

Nous allons maintenant exécuter plusieurs opérations à partir de ces fichiers fastq.

*Pour trouver facilement un outil vous pouvez entrer son nom dans la case de recherche « search tools ».*

Exécutez les programmes listés ci-dessous :

▣ **FastQc** : Permet de contrôler la qualité des reads issus d'un séquençage NGS.

Sélectionnez le fichier input.fastq dans « FASTQ reads » puis cliquez sur « Execute ».

▣ **Cutadapt** : Supprime les séquences d'adaptateurs des reads. Sélectionnez le fichier fastq dans « Input File », cliquez sur 5' or 3' (Anywhere) Adapters puis sélectionnez l'adaptateur "Illumina".

▣ **Map with BWA** : Utilisez le fichier Reference.fasta de l'historique comme référence.

▣ **IdxStats** : Compte le nombre de séquences mappées sur chaque séquence de la référence.

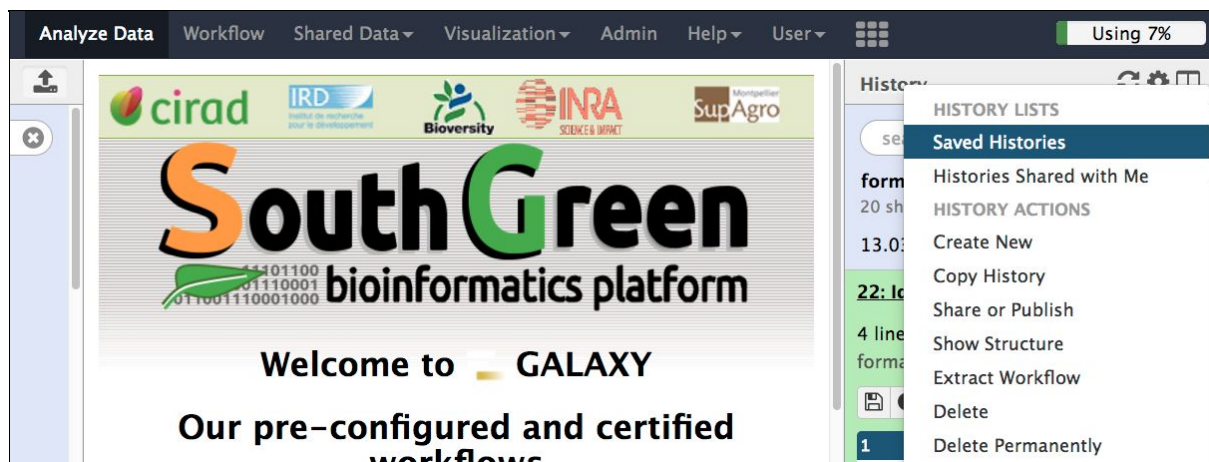
▣ **BAM-to-SAM** : Sélectionnez le fichier BAM issu du mapping dans « BAM File to convert »

### 1-4 Utilisation des historiques

*En cliquant sur Option => Saved histories Galaxy ouvre une interface de gestion des historiques.*

▣ Donnez un nom à votre historique courant en cliquant sur « Unnamed history ».

▣ Créez un nouvel historique.



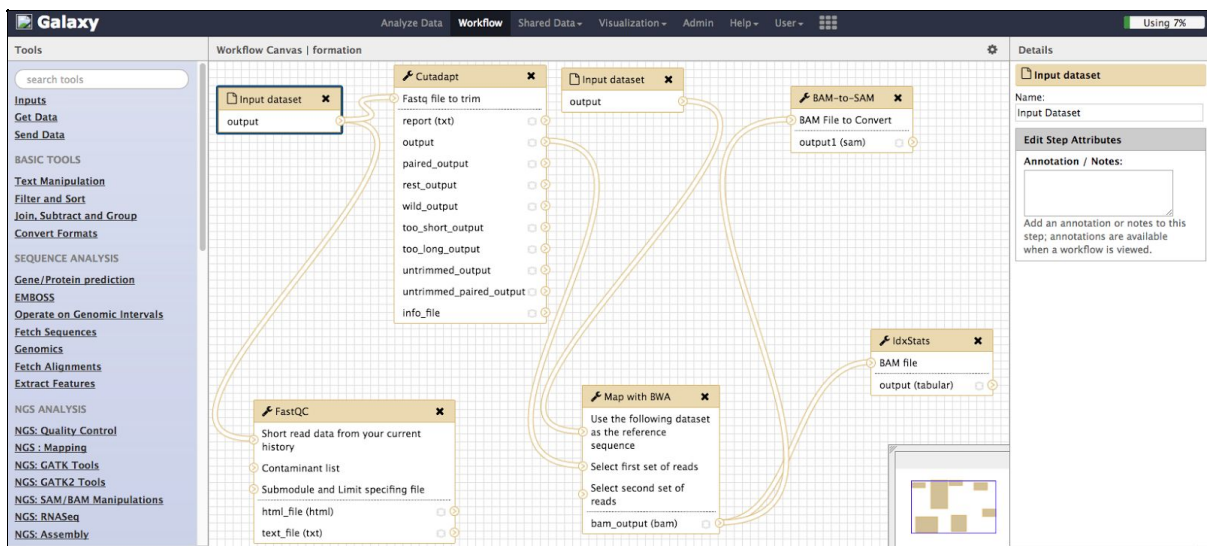
## 1-5 Création de workflow

Le but ici est de créer un workflow pour reproduire l'enchaînement des briques que nous venons d'exécuter.

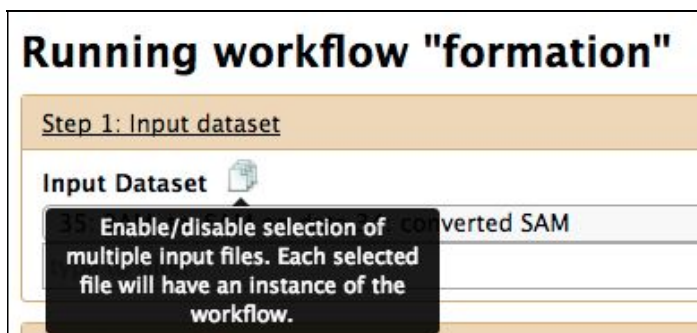
- ▣ Rendez vous dans la rubrique « Workflow » de Galaxy.
- ▣ Cliquez sur « Create new workflow » et attribuez lui un nom. Cliquez ensuite sur votre workflow puis sur « edit » pour ouvrir l'interface de création.

*En cliquant sur une brique à gauche de l'interface une petite fenêtre apparaît sur le canevas. Cette fenêtre indique les fichiers d'entrée et de sortie de la brique. Pour enchaîner deux briques il suffit de relier le fichier de sortie de la première brique vers le fichier d'entrée de la seconde à l'aide des petites flèches.*

- ▣ A l'aide des briques disponibles à gauche de l'interface, réaliser un workflow qui reproduit l'enchaînement des briques de l'exercice précédent.
- ▣ Sauvegarder votre workflow en cliquant sur « Option » puis « Save ».
- ▣ Lancez votre workflow en partant du fichier récupéré dans les "shared data" et envoyez les résultats dans un nouvel historique.



- ▣ Galaxy offre la possibilité de lancer un même workflow sur plusieurs fichiers à la fois, en cliquant sur la petite icône à droite de « Input dataset ».



## 1-6 Publication de workflow

*Galaxy permet de partager un workflow facilement avec la communauté.*

*Il y a 3 moyens de partager un workflow :*

- \* Via un lien : crée un lien de partage pour permettre à vos contacts d'importer le workflow.*
- \* Par publication : rend le workflow public et accessible en le publiant dans la section «Published Workflows».*
- \* Par email : partage le workflow avec un autre utilisateur de galaxy.*

Partagez votre workflow par email avec l'utilisateur de votre choix.

## 2 – Présentation et utilisation des workflows South Green

### 2-1 Démo du workflow “Cleaning and Mapping”

### 2-2 Workflow “SNP Calling”

Chargez les fichiers bam de la librairie partagée “Galaxy\_trainings\_2015 / NGS” ainsi que la séquence de référence.

Visualiser un des alignements à l'aide du plugin Trackster. Cliquez sur l'icône de visualisation “View in Trackster”.



Cliquez ensuite sur le bouton “View in new visualisation”.

#### **View Data in a New or Saved Visualization?**

You can add this dataset as:

- a new track to one of your existing, saved Trackster sessions if they share the genome build: %3F
- or create a new session with this dataset as the only track

Cancel

View in saved visualization

View in new visualization

Puis ajouter une nouvelle référence de visualisation correspondant au 3 séquences. “Add a Custom Build”.

## Add a Custom Build

**New Build**

**Name (eg: Hamster):**

**Key (eg: hamster\_v1):**

**Definition:**

FASTA [Len File](#) [Len Entry](#)

▾

Recommencez la manip et choisissez la bonne référence.

Importez le workflow “SNP Calling” puis lancez le sur l’ensemble des fichiers bam.

Observer le fichier de SNP au format VCF.

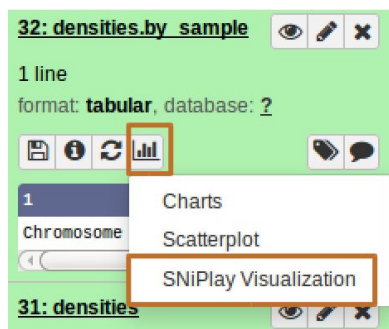
Visualiser les variations obtenues encore à l’aide du plugin de visualisation “Trackster”.

### 2-3 Workflow “SNIPlay” pour l’analyse de SNPs et GWAS

Dans un nouvel historique, chargez depuis la librairie partagée “Galaxy\_trainings\_2015 / GWAS” le fichier VCF “riceGBS.vcf” et le fichier de phentypage “phenotyping\_DRB.csv”.

Importez le workflow “SNIPlay3\_without\_annotation” puis lancez le.

Observez les sorties à l’aide de la visualisation “SNIPlay visualization”.



Par exemple, observez les sorties suivantes:

- snp\_density.snpden: densité des SNPs le long des chromosomes
- densities.by\_sample: densité des SNPs pour chaque individu
- All Outputs: analyse de structure de la population par sNMF. Coefficients d’admixture pour chaque individu.
- analyse.mds\_plot.txt: MDS plot des individus sur la base des données génotypiques
- Newick viewer: arbre de distance des individus analysés sur la base des SNPs

Importez ensuite le workflow GWAS pour étudier l’association des marqueurs avec le contrôle du poids racinaire. Observez le résultat du fichier “Tassel output” sous forme de Manhattan plot.

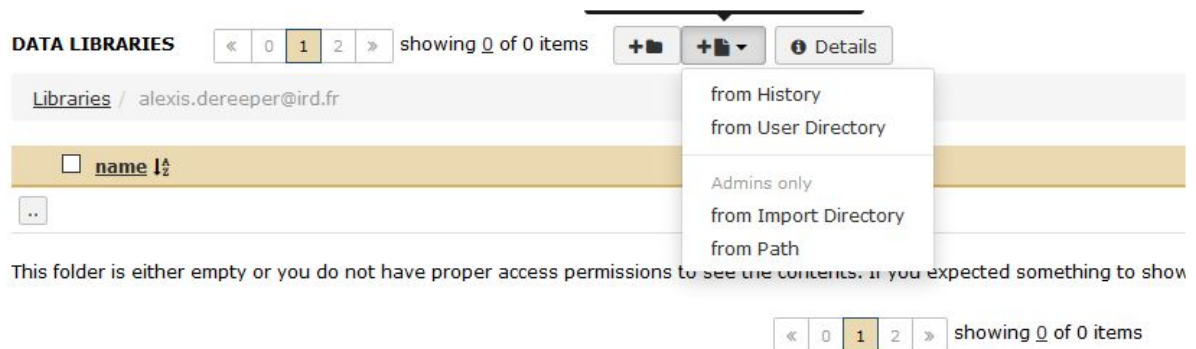
## 2-4 Autres workflows

Testez les workflows de votre choix en utilisant les données fournies en exemple.

## 3 – En roues libres...

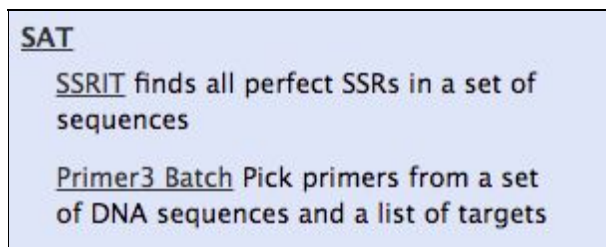
Se connecter sur l'instance Galaxy auquel vous êtes rattaché et utilisez les outils et workflows correspondant à vos problématiques de recherche.

Pour le chargement de données lourdes dans votre librairie Galaxy personnelle, demandez-nous, nous ferons la manipulation nécessaire en lien avec votre compte. Dans un 2e temps, importer les données chargées dans votre espace, en cliquant sur "from User Directory" et sélectionnez les fichiers à charger.



Autres exemples d'analyses possibles que ceux vus précédemment:

### - Recherche de microsatellites avec définition des primers : Suite d'outils SAT



### - Réaliser un alignement blast d'un fichier contenant plusieurs séquences contre sa propre banque de séquences:

- 1 - Importer sa propre banque de séquences
- 2 - Créer les fichiers index nécessaires pour réaliser une recherche de similarité avec blast avec l'outil **MakeBlastDB**
- 3 - Réaliser la recherche de similarité avec **Blast**.

### - Expression Différentielle (HT-seq count, EdgeR et DESeq) ...