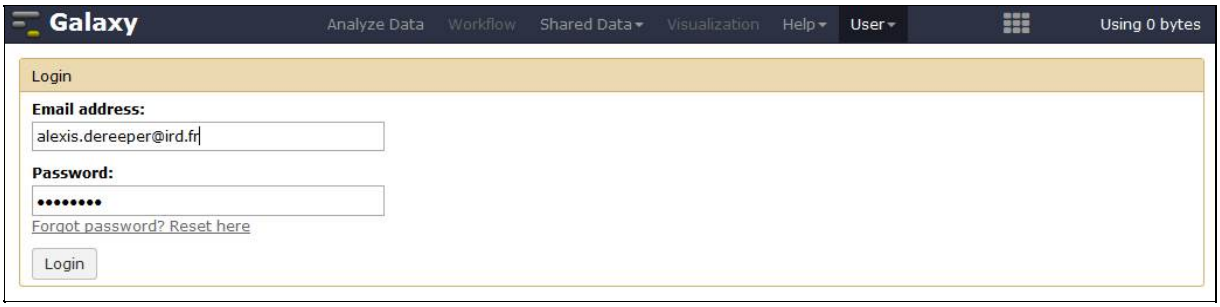
Formation Galaxy

13 Novembre 2014

1 – Premiers Pas…

1-1 Connexion

☐ **Connectez-vous sur la plateforme Galaxy SouthGreen à l'adresse suivante :** http://gohelle.cirad.fr/galaxy/



☐ Utiliser votre adresse email et votre mot de passe pour vous identifier.

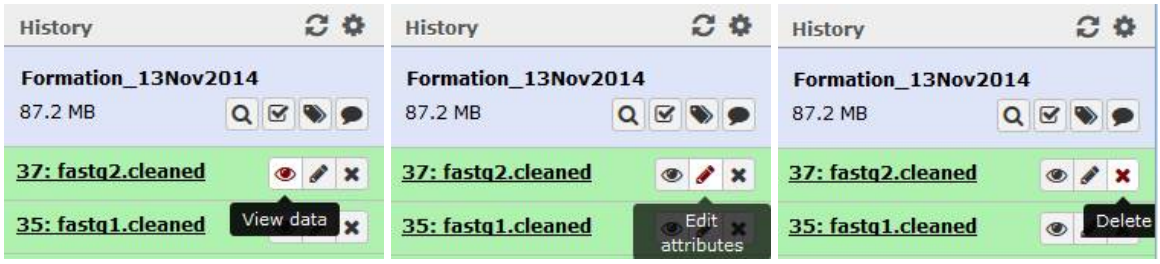
Si vous n'avez pas de compte, utiliser pour aujourd'hui le compte formationN/formationN

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/coller directement le contenu du fichier dans le cadre « URL/Text »*



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

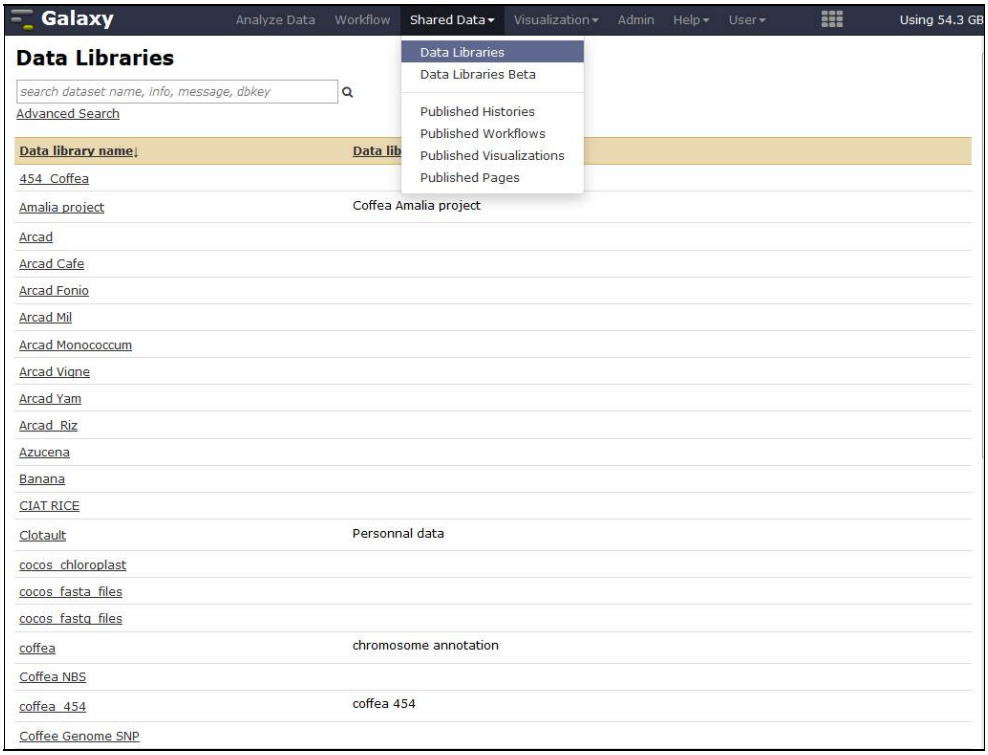
Testez les 3 types d'import possibles de données. Vous observerez dans l'onglet « History » à droite de l'interface la progression de l'importation.

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) et importez un fichier dans votre historique courant. (répertoire "Formation_Galaxy").



- Cliquez sur la library Formation
- Ouvrez le repertoire Preprocessing and Mapping
- Cochez les fichiers **input1.fastq**, **input2.fastq** et **Solexa_mRNA_primers.txt**
- Ouvrez le repertoire SNP
- Cochez le fichier **reference.fas.txt**
- En bas de page cliquez sur le bouton GO avec l option "import to current history"

1-3 Exécution de jobs

Nous allons maintenant exécuter plusieurs opérations à partir de ce fichier fastq.

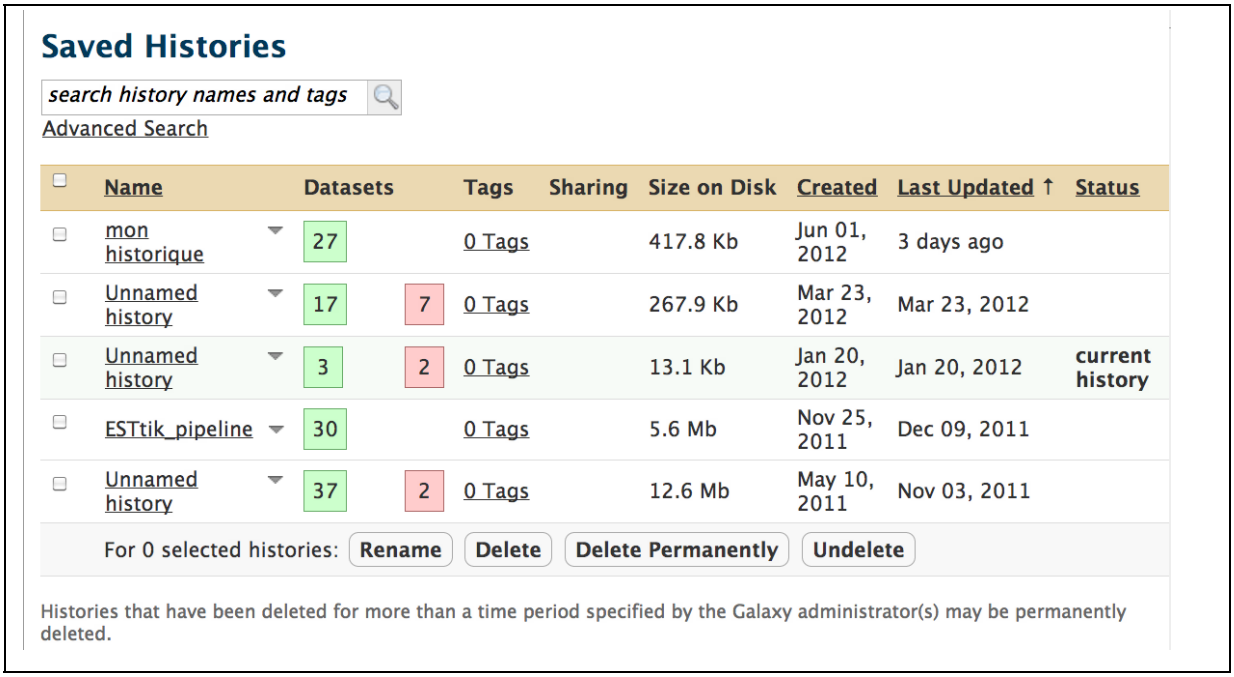
Pour trouver facilement un outil vous pouvez entrer son nom dans la case de recherche « search tools ».

Exécutez les programmes listés ci-dessous :

- ☐ **FastQC** : Permet de contrôler la qualité des reads issus d'un séquençage NGS.
  Sélectionnez un des fichiers fastq dans « FASTQ reads » puis cliquez sur « Execute ».
- ☐ **Fastq Groomer** : Permet de recoder les scores de qualité de séquençage.
  Réalisez un recodage de la qualité des fichiers fastq (Illumina 1.5) en qualité Sanger.
- ☐ **Cutadapt** : Permet le nettoyage des fichiers fastq bruts (élimination des régions de mauvaises qualité, et des adaptateurs).
  Réalisez le nettoyage des fichiers fastq 1 et 2, en renseignant la séquence de 3 adaptateurs contenus dans le fichier "Solexa_mRNA_primers.txt" et en fixant les paramètres suivants: qualité minimale à 20, taille minimale de l'overlap avec un adaptateur égale à 7, taille minimale d'un read égale à 20.
- ☐ **FastQ Interlacer** : Permet de dissocier les reads singles et pairés.
  Le nettoyage a parfois supprimé certains reads forward et gardé le reverse correspondant et vice-versa. Utilisez cet outil sur chaque fastq nettoyé pour obtenir un fichier fastq ne contenant que les reads en paires.
- ☐ **FastQ de-Interlacer** : Permet de séparer les reads forward et reverse.
  Utilisez cet outil pour regénérer 2 fichiers fastq gauche et droite.
- ☐ **BWA** : Permet le mapping de reads Illumina sur une séquence de référence.
  Réalisez le mapping des reads sur le fichier "reference.fas.txt" (3 gènes de Riz).

1-4 Utilisation des historiques

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



Saved Histories

search history names and tags

Advanced Search

	Name	Datasets		Tags	Sharing	Size on Disk	Created	Last Updated ↑	Status
☐	mon historique	27		0 Tags		417.8 Kb	Jun 01, 2012	3 days ago	
☐	Unnamed history	17	7	0 Tags		267.9 Kb	Mar 23, 2012	Mar 23, 2012	
☐	Unnamed history	3	2	0 Tags		13.1 Kb	Jan 20, 2012	Jan 20, 2012	current history
☐	ESTtik_pipeline	30		0 Tags		5.6 Mb	Nov 25, 2011	Dec 09, 2011	
☐	Unnamed history	37	2	0 Tags		12.6 Mb	May 10, 2011	Nov 03, 2011	

For 0 selected histories: Rename Delete Delete Permanently Undelete

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.

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

☐ Créez un nouvel historique.

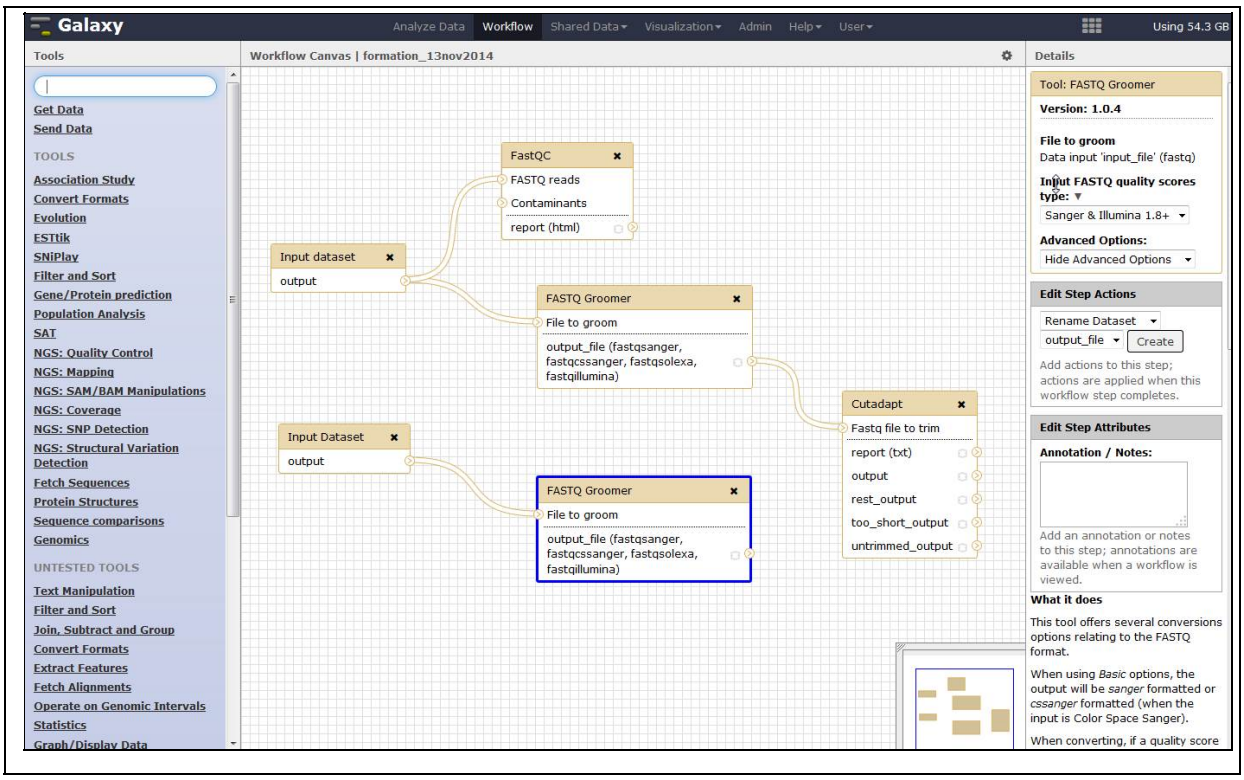
1-5 <u>Création de workflow</u>

Le but ici est de créer un workflow pour reproduire l'enchainement des briques du paragraphe 1-3.

☐ 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 enchainer 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.



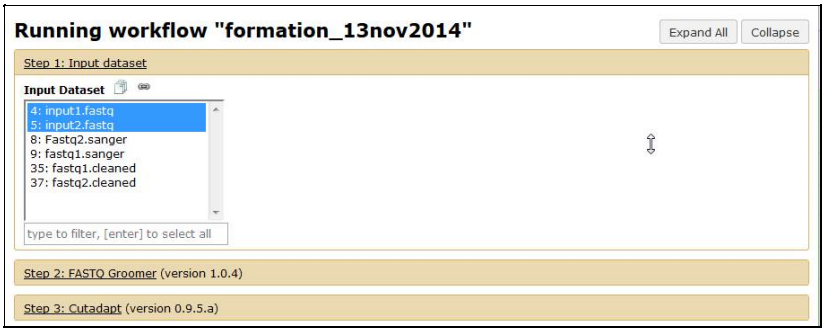
☐ A l'aide des briques disponibles à gauche de l'interface, réaliser un workflow qui reproduit l'enchainement des briques de l'exercice précédant.

☐ 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.

Workflow sur fichiers multiples:

☐ Galaxy offre la possibilité de lancer un même workflow sur plusieurs fichiers à la fois, en cliquant sur la petite icone à droite de "Input dataset".



Lancer le workflow sur les échantillons RC1, RC2 et RC3 importés depuis la librairie partagée.

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 en le publiant dans la section «Published Workflows ».*

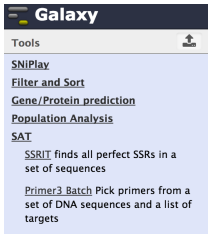
** Par email : partage le workflow avec un autre utilisateur de galaxy.*

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

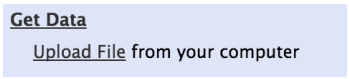
2 – En roues libres…

Exemples d'analyses possibles :

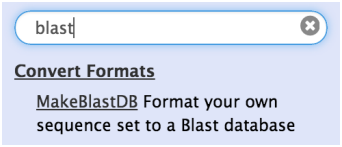
- **Recherche de microsatellites avec définition des primers : Suite d'outils SAT**



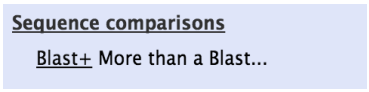
- Réaliser un alignement blast d'un fichier contenant plusieurs séquences contre sa propre banque de séquences:
    1. Importer sa propre banque de séquences



    2. Créer les fichiers index nécessaires pour réaliser une recherche de similarité avec blast avec MakeBlastDB



    3. Réaliser la recherche de similarité



- Recherche de SNP (à partir d'un fichier de type bam) :
    a. Marquer les duplicats techniques **Picard *Mark Duplicate reads***
    b. Réaligner les reads à l'aide des outils de la suite **GATK *Realigner Target Creator*** et ***Indel Realigner***
    c. Détecter les polymorphismes avec l'outil **GATK *Haplotype Caller***

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