



# **Formation sRNAseq Analyse des miRNAs sous Galaxy**

## **- EXERCICES -**



cutadapt

BWA

SAMtools

“**FastQC** is a quality control tool for high throughput sequence data.”

<http://www.bioinformatics.bsrc.ac.uk/>

A tool that removes adapter sequences from DNA sequencing reads.

**cutadapt** removes adapter sequences from high-throughput sequencing data.

<https://code.google.com/p/cutadapt/>

“**Burrows-Wheeler Aligner (BWA)** is an efficient program that aligns relatively short nucleotide sequences against a long reference sequence such as the human genome.” <http://bio-bwa.sourceforge.net>

“**SAM** (Sequence Alignment/Map) format is a generic format for storing large nucleotide sequence alignments.”

<http://samtools.sourceforge.net>



## Objectifs :

Cette formation a pour objectif de vous aider à traiter les séquences issues de projet de sRNAseq (miRNA). Vous y découvrirez les problématiques spécifiques de l'analyse des petits ARNs non codant, les outils liés et les mettrez en œuvre afin de détecter, annoter, prédire, quantifier, ... les miRNA.

Pour vous connecter à l'instance Sigena de Galaxy :

- <http://sigena-workbench.toulouse.inra.fr>
- Si besoin, vous avez la possibilité de demander un compte depuis <http://bioinfo.genotoul.fr/index.php?id=74> (avec une adresse académique uniquement).
- Comptes disponibles pour votre session de formation :
  - Logins : anemone , aster, bleuet, iris, muguet, narcisse, pensee, rose, tulipe, violette, lilas, pervenche, laurier, lavande, lis, capucine, coquelicot, geranium, liseron, arome, chardon.
  - Password : f1o2r3!

## Exercice n°1 : Création d'un historique – importation des données

Nous allons travailler à partir de jeux de données Illumina représentant 2 tissus « s2.fastq » et « s1.fastq ».

Pour récupérer vos jeux de données, veuillez importer cet historique : « tp-2-srnaseq-29-mai-2013 ».

Nommer votre historique (Ex : « miRNAs »).

## Exercice n°2 : Analyse de la qualité et nettoyage (suppression des adaptateurs – suppression de la redondance intra « fastq »)

### 1 - Analyse de la qualité des données : « FastQC »

A partir du jeu de données s2, veuillez utiliser l'outil Galaxy « [Fastqc: Fastqc QC](#) using FastQC from Babraham » pour générer les graphiques d'analyse de la qualité.

En cliquant sur l'icône « œil », les statistiques s'affichent dans la fenêtre centrale :

Il vous est possible de récupérer :

- L'ensemble des résultats de FastQC en cliquant sur l'icône « disquette » du dataset.
- Un graphique par un clic droit sur son intitulé, en fin de page :

#### Files created by FastQC

[duplication\\_levels.png \(19.9 KB\)](#)

[fastqc\\_data.txt \(24.2 KB\)](#)

[fastqc\\_report.html \(42.1 KB\)](#)

[kmer\\_profiles.png \(105.8 KB\)](#)

[per\\_base\\_gc\\_content.png \(31.4 KB\)](#)

## 2 - Nettoyage des données :

### 2.1 - Suppression des adaptateurs

La suppression des adaptateurs s'effectue grâce au script « cutadapt » disponible dans Galaxy « \* Remove adaptators with cutadapt ».

Lancer ce traitement sur « s2 » avec les paramètres suivants :

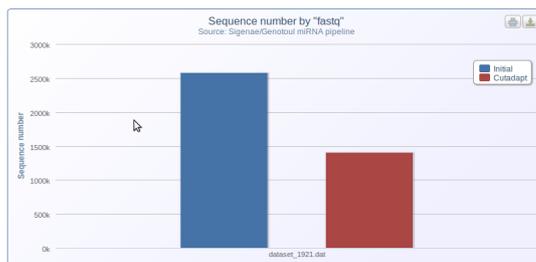
- Adaptateur : ATCTCGTATGCCGTCTTCTGCTTG (champs pré-rempli dans Galaxy).
- Taille minimum : 18pb - Taille maximum : 25pb

Sorties de « cutadapt » :

- Fichier « fastq » contenant les séquences nettoyées de leurs adaptateurs
- Fichier « log » contenant les informations de traitement

### 2.2 - Statistiques après suppression des adaptateurs

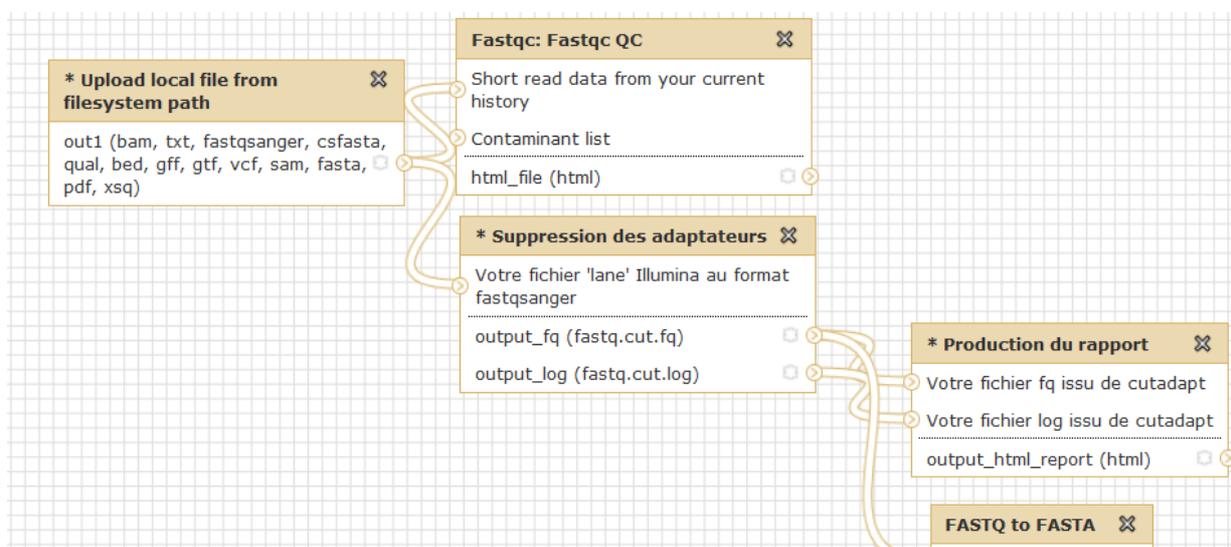
Lancer l'outil « \* Cutadapt report after adaptators removing », sur « cutadapt\_on\_{s2}\_give\_logs\_file » et « cutadapt\_on\_{s2}\_give\_fastq\_file », pour produire d'un rapport de traitement (au format html).



Pour visualiser ces statistiques, veuillez ouvrir ou enregistrer le fichier sur votre PC et ouvrir la page « html » dans votre navigateur.

## 3- Exporter votre historique en workflow

Créer un workflow sur la base des traitements précédent et l'exécuter sur le deuxième fastq.







### Exercice - Étape 1 - Annotation fonctionnelle du « fasta » issu du mapper :

Annoter le « fasta » issu de Mapper avec l'outil « \* Bam alignment bam, sort and filter on a fasta file » contre les banques mirBase (hairpin\_T), Rfam et tRNA. Chaque alignement doit être filtré selon ces paramètres : Flag 0 ou 16 - 1 mismatch autorisé.

Puis lancer l'outil « \* Annotations comparison with Venn diagram », à partir des 3 fichiers au format « filter2 » générés par l'alignement. L'objectif est de construire un diagramme de Venn.

Puis lancer l'outil « \* Annotation matrix built in csv format », à partir des 3 fichiers au format « filter1 » générés par l'alignement.

### Exercice - Étape 2 - Annotation fonctionnelle du « fasta » issu du mirDeep2core :

Convertir le fichier bed, issu de miRDeep2core, en fichier fasta grâce à l'outil « \* miRDeep2core - bed to fasta file bed file from mirdeep2core ».

Annoter le « fasta » issu de miRDeep2core avec l'outil « \* Bam alignment bam, sort and filter on a fasta file » contre les banques mirBase (hairpin\_T), Rfam et tRNA. Chaque alignement doit être filtré selon ces paramètres : Flag 0 ou 16 - 1 mismatch autorisé.

Puis lancer l'outil « \* Annotations comparison with Venn diagram », à partir des 3 fichiers au format « filter2 » générés par l'alignement. L'objectif est de construire un diagramme de Venn.

Puis lancer l'outil « \* Annotation matrix built in csv format », à partir des 3 fichiers au format « filter1 » générés par l'alignement.

### Exercice - Étape 3 - Comparaison des diagrammes de Venn :

Comparer les diagrammes de Venn selon que les traitements aient été lancés à partir du « fasta » issu du Mapper ou du « fasta » issu de miRDeep2core.

