

# Traitement de données métagénomiques

IUT Génie Biologique, Option Bioinformatique (2e année)

Bérénice Batut

Février 2016

## Introduction

Les séances de TP vont s'organiser autour d'un mini-projet qui vous permettra de vous familiariser avec le traitement de séquences métagénomiques issues d'un échantillon d'intestin humain.

## Principe du projet

L'objectif du projet est d'extraire de l'information sur la composition (organismes et fonctions réalisées) d'un échantillon du microbiote intestinale.

Tous les jeux de données sont issus d'un [projet](#) de comparaison des microbiotes intestinaux de jumeaux maigres et obèses. Les résultats de cette études ont été publiés et sont disponibles (Turnbaugh et al., 2009).

## Organisation

Pour ce projet, vous vous mettez par binôme. Les deux membres du binôme doivent participer au projet, pour les analyses mais aussi la rédaction du compte-rendu. D'ailleurs, il serait bien qu'au moins un des membres du binôme amène son ordinateur personnel. De plus, la personne qui "analyse" et la personne qui "rédige" ne doivent pas être toujours les même.

La note du projet portera sur le compte-rendu et le suivi des instructions (2 points).

## Instructions

Comme pour le projet précédent, l'organisation est importante. Ainsi, tous les fichiers liés à ce projet sont rassemblés dans un dossier comprenant

- Un dossier **data** pour toutes les données
- Un dossier **results** pour les résultats des analyses (graphiques, fichiers générés, ...)
- Un dossier **doc** pour les notes et en particulier le compte-rendu du TP

Le compte-rendu dans le cadre de ce projet correspond à un "cahier de notes" avec les différentes étapes de l'analyse (graphiques, résultats, échecs, ...). En effet, quand on travaille sur un projet, tenir un tel "cahier" permet de garder une trace de ce qui a été fait, ce qui a marché pour pouvoir expliquer la démarche, faciliter la reprise du travail par quelqu'un, ... Dans ce "cahier", vous noterez les différentes étapes, une description des différents outils utilisés (et pourquoi), les graphiques, les liens vers les fichiers utiles ainsi que les sites. Pour faciliter la rédaction, un modèle de cahier en [markdown](#) doit être utilisé. Il est disponible sur [bebatut-edu.github.io](http://bebatut-edu.github.io).

## Récupération du jeu de données

Chaque binôme se voit attribuer un jeu de données (accessible depuis [bebatut-edu.github.io](http://bebatut-edu.github.io)). Chaque lien renvoie sur la page de l'*EBI metagenomics* consacrée au jeu de données. Cette page rassemble la description du jeu de données, les données téléchargeables ainsi que les résultats des analyses effectuées par l'*EBI metagenomics*.

Pour vous familiariser avec le jeu de données, analysez les métadonnées disponibles sur la page de l'*EBI metagenomics* et expliquez comment les séquences ont été obtenues, à quoi elles correspondent, quel type de séquençage a été utilisé (les avantages et les limites de ce type de séquençage), ...

## Traitement des données métagénomiques

### Analyses des résultats de l'*EBI metagenomics*

Ce jeu de données a été analysé par l'*EBI metagenomics* pour extraire des informations sur la composition et les fonctions effectuées dans l'échantillon. Quelles sont les différentes étapes d'analyses effectuées par l'*EBI metagenomics*? Quels sont les outils utilisés?

Pour chaque étape, quels sont les résultats obtenus? Pour les différentes étapes du contrôle qualité, combien et quelle proportion de séquences sont conservées? Quels sont les taxons majoritaires trouvés par l'analyse taxonomique? Quelles sont les principales fonctions réalisées par les organismes présents d'après l'analyse fonctionnelle?

### Analyses du jeu de données avec ASaiM

**ASaiM** est un framework développé pour traiter les données métagénomiques dans un environnement Galaxy dédié. Cet environnement est accessible depuis un navigateur Web à l'adresse : <http://172.22.15.14:8080>.

Pour analyser le jeu de données, suivez le [tutoriel d'ASaiM](#) (Section "Process and analyze the dataset"). Pensez à enregistrer les fichiers générés à chaque étape pour pouvoir reprendre les analyses aux différentes séances de TP.

Pour chaque étape du tutoriel, expliquez les outils choisis, les paramètres, les résultats obtenus. Lorsque c'est possible, comparez les résultats obtenus avec ASaiM avec ceux obtenus avec le pipeline d'*EBI metagenomics*.

## Discussion

Faites un schéma des analyses effectuées et leur enchainements (avec les entrées et sorties des différentes étapes).

Que pensez-vous de ce type d'analyses? Quelles sont les avantages et les limites?

Discutez les affiliations taxonomiques et fonctionnelles obtenues. Remettez ces résultats dans le contexte des résultats de l'analyse initiale (Turnbaugh et al., 2009).

## Références

Turnbaugh, P. J., Hamady, M., Yatsunencko, T., Cantarel, B. L., Duncan, A., Ley, R. E., ... Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, 457(7228), 480–484. <http://doi.org/10.1038/nature07540>