

TP2

Analyse de séquences de gènes d'ARNr 16s (MiSeq Illumina)(DADA2)

Nous allons dans ce TP manipuler les données sortant d'un séquençage Illumina 16s.

Nous pouvons utiliser différents logiciels qui pourraient nous mener au même résultat mais pour des raisons de facilité nous allons utiliser **Dada2** et **Phyloseq** afin d'obtenir :

- Une table de variante de séquence d'amplicon (ASV)
- Une table OTU, afin d'obtenir le nombre de fois on a enregistré chaque variante de séquence d'amplicon dans chaque échantillon
- Attribution de taxonomie des séquences de sortie
- Analyser ainsi l'alpha diversité

Le tout se fera sur l'environnement R

En fonction de votre version de R vous allez vers le tutoriel adéquat à votre version

1. Ouvrir le tutoriel **DADA2 pipeline Tutorial.pdf**

Vous allez suivre ce tutoriel

2. Récupérer les séquences fastq et les fichiers (nécessaire pour l'analyse) contenus dans le dossier **MiSeq_SOP-1**.

Les fichiers fastq ont été obtenus via séquençage **Illumina MiSeq** des échantillons fécaux prélevés sur des souris après le sevrage.

L'objectif de ce tutoriel est de déterminer la population microbiennes contenue dans les différents échantillon fécaux.

⇒ Avertissement

Pour ce tutoriel les données ont déjà été :

- Démultipliés (divisés en fichiers de données par échantillon)
- Primer, adaptateur,.. déjà retirés

I. Téléchargement des librairies et initiation des chemin d'accès

Entamer par le point « **Getting ready** » afin de télécharger vos librairies et initier votre dossier de travail

Vous allez initier vos chemin d'accès là ou vous avez mis vos données fastq.

II. Inspection de la qualité des reads

Cette étape permet d'observer la qualité des reads de votre jeu-de données

- Gris => fréquence de chaque score de qualité à chaque position de base
- Vert => la note médiane de qualité à chaque position
- Orange => les quartiles

III. Taux d'erreur

Taux d'erreur pour chaque transition possible (A C, A G,)

- Ligne noire => taux d'erreur estimé par algorithme d'apprentissage automatique
- Ligne rouge => taux d'erreur attendu
- Point => l'erreur observée

IV. Dé répllication

Combinaison de toutes les reads identiques en une séquences uniques

V. fusionner les paires de reads

construction des séquences « contig »