

TRAITEMENT DE SEQUENCES ET DE METADONNEES DE COHORTES PUBLIQUES METAGENOMIQUES EN LIEN AVEC L'ETUDE DE L'IMPACT DU REGIME ALIMENTAIRE SUR LE MICROBIOTE INTESTINAL HUMAIN

Spécialité : Bioinformatique

Durée : 5-6 mois, début entre janvier à mars 2024, rémunéré par gratification.

Encadrement : Victoria Meslier (victoria.meslier@inrae.fr), Pauline Barbet (pauline.barbet@inrae.fr) et Nicolas Pons (nicolas.pons@inrae.fr)

Équipe d'accueil : MetaGenoPolis (US 1367), Bât. 325, INRAE, Centre de recherche de Jouy-en-Josas, Domaine de Vilvert, 78350 Jouy-en-Josas.

Plan : <https://www.google.fr/maps/place/MetaGenoPolis/>

CONTEXTE ET OBJECTIF DU STAGE

L'essor de la métagénomique a permis des avancées scientifiques majeures dans le domaine de la santé humaine en étudiant le rôle du microbiote intestinal dans des maladies aussi diverses que l'obésité, le diabète, le cancer ou les maladies cardio-vasculaires.

Au sein de l'unité MetaGenoPolis, l'équipe InfoBioStat (IBS) développe une expertise pointue dans la découverte, la caractérisation et la compréhension de la composition en espèces bactériennes présentes dans le microbiote intestinal et leur association à des pathologies. L'équipe IBS est aussi impliquée dans l'étude de l'impact du régime alimentaire, en particulier de la consommation de fibres, sur la composition du microbiote intestinal en lien avec la santé.

Parmi ces projets, l'un d'entre eux vise à créer un répertoire de cohortes métagénomiques publiques associées à l'alimentation, possédant des métadonnées disponibles et informatives et d'identifier certaines espèces associées aux fibres et potentiellement bénéfiques pour la santé.

Dans ce contexte, l'objectif de ce stage est de contribuer à la constitution de ce répertoire de cohortes métagénomiques publiques en prenant en main les outils de l'équipe IBS pour réaliser le téléchargement, le nettoyage de données et le pré-processing bioinformatique des séquences. L'étudiant sera également responsable du téléchargement et de la curation des métadonnées des échantillons associés, afin d'élaborer un jeu de données métagénomique curé et consolidé (FAIR).

VOS MISSIONS

- Réaliser un état de l'art bibliographique, notamment afin (i) d'identifier les bases de données/collection de données métagénomiques similaires à celles proposées dans ce stage, (ii) des méthodologies associées à la mise en place de ces bases de données
- A partir de cet état de l'art, identifier les forces/faiblesses de chaque méthode, en comparaison à notre approche
- Participer activement à l'identification des cohortes métagénomiques shotgun d'intérêt
- Réaliser le téléchargement et le nettoyage de ces cohortes (trimming et retrait des reads de l'hôte)
- Télécharger, curer et harmoniser les métadonnées associées aux données métagénomiques précédemment téléchargées
- Le cas échéant, réaliser les étapes de mapping et comptage des espèces microbiennes avec nos pipelines bioinformatiques et biostatistiques.

PROFIL SOUHAITE

- Formation M1 ou M2 en bioinformatique
- Connaissances en programmation (langage R, python 3)
- Connaissances en système de gestion de base de données seraient un plus (noSQL (MongoDB))
- Des connaissances générales en biologie et sur les problématiques liées aux données métagénomiques seraient un plus
- Anglais scientifique