



Proposition de stage de Master en Bioinformatique

ECOSEQ : *Exploration du Côté Obscur du SEQuençage métagénomique dédié au diagnostic et à la recherche clinique.*

Lieu : HCL GENEPII (Lyon, Croix rouge)

Durée : 6 mois, dès janvier-février 2025

Co-Encadrement : **Laurence Josset** (CIRI, HCL) laurence.josset@chu-lyon.fr ; **Hussein Anani** (CIRI, HCL) hussein.anani@chu-lyon.fr

Mots clés : bioinformatique, données massives, métagénomique, virome, échantillons contrôlés

1. Contexte scientifique et local du projet

Le séquençage métagénomique et métatranscriptomique permet de séquencer tous les ADN et ARN d'un prélèvement sans a priori. Il est utilisé en diagnostic pour identifier de nouveaux pathogènes ou des pathogènes rares. Il offre aussi la possibilité d'analyser en profondeur les communautés du microbiome (virome, bactériome, etc.) (Wahl et al., 2022).

Néanmoins, le séquençage sans a priori présente des défis techniques significatifs, notamment lorsqu'il s'agit d'explorer le virome humain, en particulier dans le cadre respiratoire. En effet, la contamination par des acides nucléiques à divers stades du processus de séquençage peut biaiser l'analyse de la diversité microbienne et fausser l'interprétation des données obtenues. Ce risque de contamination est amplifié par la multiplicité des étapes impliquées dans le processus de préparation des échantillons. Afin de détecter des virus faiblement abondants, une augmentation de la profondeur de séquençage est souvent nécessaire, mais cela entraîne également une hausse du risque de détection de contaminants à faible concentration. Les sources de contamination sont variées (Jurasz et al., 2021) : i) les contaminations croisées provenant d'échantillons voisins lors de la préparation (cross-contamination ou « splashome »), ii) les contaminations liées aux réactifs employés au cours des différentes étapes (kitome), et iii) les contaminations environnementales ou provenant d'autres microbiotes associés au prélèvement.

L'impact de ces contaminants dépend fortement du type de prélèvement analysé. Dans les échantillons riches en microorganismes, tels que les prélèvements fécaux, les crachats ou les échantillons cutanés et nasopharyngés, la contamination peut être relativement faible et ne pas affecter significativement la diversité microbienne observée. En revanche, pour des échantillons de faible biomasse, voire stériles, le risque d'identifier de faux microorganismes est plus élevé, ce qui peut conduire à des interprétations erronées et masquer la véritable diversité sous-jacente (Porter et al. 2021). Ces contaminations ont d'ailleurs conduit à des conclusions inexactes sur la stérilité de certains systèmes biologiques.

Ainsi, pour garantir des résultats cliniquement interprétables et fiables, il est indispensable de bien identifier et caractériser ces contaminants, notamment ceux liés aux réactifs. Cela permettra de renforcer la validité des analyses et d'améliorer la caractérisation des agents pathogènes responsables des maladies diagnostiquées, contribuant ainsi à des décisions thérapeutiques mieux informées.

2. Objectifs scientifiques du projet

L'objectif du projet de master est de caractériser de manière exhaustive l'ensemble des contaminants associés aux échantillons contrôlés des runs de séquençage clinique par une approche métagénomique ou métatranscriptomique. Cette caractérisation comprendra le benchmarking de logicielles d'assemblages de novo et l'annotation de ces contaminants (Beller et al.2022). Ces nouvelles stratégies d'assemblage pourront également être testées sur des prélèvements respiratoires humains afin de mieux annoter le virome respiratoire, en comparaison des analyses réalisées en routine sur la plateforme de séquençage GENEPII.

Le-la candidat-e travaillera en collaboration étroite avec les responsables scientifiques du stage pour développer ses compétences à l'interface de la virologie, de la bioinformatique et de la métagénomique/métatranscriptomique. Il.elle pourra bénéficier d'un support technique de la plateforme GENEPII (<https://teamhcl.chu-lyon.fr/genepii>) pour être guidé au mieux dans les choix méthodologiques et leur mise en œuvre (bonnes pratiques de développement, infrastructure de calcul). Les bases de données, les jeux de données, l'infrastructure informatique et les outils bioinformatiques nécessaires à la bonne réalisation des tâches listées ci-dessus seront disponibles dès le début du projet.

3. Valorisation scientifique

Les résultats obtenus issus de ce master peuvent être facilement valorisé dans un mini article scientifique (type *Genome Announcement*).

4. Compétences recherchées

Nous recherchons un-e candidat-e de niveau bac + 4 en **bioinformatique**. Il-elle devra être capable d'interagir avec différents chercheurs et ingénieurs du projet pour développer et mettre en place des méthodes innovantes en sciences de la vie.

Compétences requises :

- R, Python, algorithmes bioinformatiques
- Traitement des données de séquençage (lectures de séquençage brutes)
- Comprendre les technologies de séquençage
- Versioning sous git
- Fortes capacités relationnelles et d'organisation, autonomie
- Capacités rédactionnelles, esprit d'analyse et de synthèse

5. Informations pratiques

Le stagiaire sera hébergé sur la plateforme GENEPII afin de favoriser les échanges interdisciplinaires. Ce stage bénéficiera d'une gratification de 6 mois à partir de février 2025 (552 euros/mois). Pour tout renseignement ou candidature, merci d'adresser un CV et une lettre de motivation le plus rapidement possible pour démarrage début 2025 à l'attention de : Laurence Josset (CIRI, HCL) laurence.josset@chu-lyon.fr ; Hussein Anani (CIRI, HCL) hussein.anani@chu-lyon.fr.

6. Bibliographie

- Wahl, A., Huptas, C. & Neuhaus, K. Comparison of rRNA depletion methods for efficient bacterial mRNA sequencing. *Sci. Rep.* 12, 5765 (2022).
- Jurasz, Henryk et al. "Contamination Issue in Viral Metagenomics: Problems, Solutions, and Clinical Perspectives." *Frontiers in microbiology* vol. 12 745076.
- Porter, Ashleigh F et al. "Metagenomic Identification of Viral Sequences in Laboratory Reagents." *Viruses* vol. 13,11 2122.
- Beller, Leen et al. "The virota and its transkingdom interactions in the healthy infant gut." *Proceedings of the National Academy of Sciences of the United States of America* vol. 119,13 (2022): e2114619119.