

Sujet de stage de M2 « Bioinformatique et biologie des systèmes » - UPS Toulouse III
(janvier-juin 2019)

INRA - GenPhySE – SIGENAE
UMR INRA-ENVT IHAP 1225 – Équipe Interactions hôtes virus & vaccinologie

Développement d'outils de bioinformatique destinés au traitement de données NGS issues de métagénomique virale

Responsable scientifique du projet : Jean-Luc Guérin - INRA-ENVT IHAP 1225

Co-encadrants du stage :

- **Bioinformatique :** Sarah Maman - INRA - GenPhySE - SIGENAE
- **Biologie :** Guillaume Croville - INRA-ENVT IHAP 1225

Lieu du stage : École Nationale Vétérinaire de Toulouse, 23 chemin des capelles, 31076 Toulouse

1. Objectifs et enjeux scientifiques

Les maladies respiratoires sont la résultante de coinfections dont les agents infectieux partenaires potentiels sont nombreux et mal caractérisés. Les méthodes de diagnostic moléculaires disponibles génèrent des informations de première importance telles que la présence d'un agent infectieux, sa charge (PCR temps réel) ou sa séquence génétique (NGS), permettant alors d'identifier finement le virus ou la bactérie et d'éventuels marqueurs de pathogénicité.

La métagénomique clinique est une des applications majeures des NGS dans le laboratoire d'accueil. La description du métagénome de la sphère respiratoire chez les animaux de rente représente par exemple un enjeu capital dans la description des populations microbiennes à des fins de réduction d'intrants médicamenteux (antibiotiques). L'analyse de la variabilité et de la distribution des génomes viraux pour une espèce virale donnée (ex : virus influenza) est également réalisée au moyen des outils NGS. Les facteurs d'émergence de sous-populations virale peuvent être étudiés grâce à la masse de données produites par les NGS.

L'un des verrous rencontrés en métagénomique est la pauvreté relative en séquences d'intérêt, masquées par l'abondance du génome de l'hôte et du microbiote. Une mission parallèle, confiée à un étudiant de master biotechnologies, sera de développer une stratégie de pré-enrichissement en amont du séquençage NGS afin d'accroître la quantité d'acides nucléiques des agents pathogènes ciblés potentiellement présents.

La mission du stagiaire de bioinformatique sera de développer, en binôme avec l'équipe SIGENAE, un processus d'analyse de données NGS appliqué aux besoins spécifiques de l'équipe de virologie. L'objectif de ce stage repose sur l'adaptation d'outils existant ou le développement d'outils innovants d'affiliation taxonomique pour l'analyse de données NGS Oxford Nanopore ainsi que leur intégration dans Galaxy.

2. Projet de stage : lots de travail

Lot #1 : analyse bibliographique et définition d'une maquette de processus d'analyse bioinformatique.

Cette phase consistera, par l'analyse d'une bibliographie fournie en partie par l'équipe de virologie, à identifier et étudier les outils bioinformatiques les plus pertinents déjà disponibles. L'étudiant proposera plusieurs maquettes de processus d'analyse à tester.

Délivrable du lot : présentation des maquettes devant les équipes pour définir collectivement le lot 2.
Echéance : semaine 3

Lot #2 : Test des premiers algorithmes sélectionnés en lot#1

Les algorithmes sélectionnés sur la base de la bibliographie seront testés individuellement sur des jeux de données publics disponibles afin de juger de leur pertinence. Dans ces jeux de données seront ajoutées des séquences d'intérêt sélectionnées par l'équipe de virologie.

Délivrable du lot : Présentation des résultats des premiers tests devant les équipes.
Echéance : semaine 10

Lot #3 : Test des algorithmes sélectionnés en lot#2

Les algorithmes sélectionnés en lot#2 seront testés sur des jeux de données générés par l'équipe de virologie.

Délivrable du lot : Présentation des résultats des premiers tests devant les équipes.
Echéance : semaine 14

Lot #4 : Création du processus d'analyse

Les outils bioinformatique sélectionnés en lot#3 seront assemblés pour créer le processus d'analyse. Le pipeline sera mis à disposition sous galaxy.

Délivrable du lot : Présentation et description du processus d'analyse devant les équipes.
Echéance : semaine 20

Lot #5 : rédaction du rapport de stage et présentation au laboratoire

Délivrable du lot : présentation des résultats devant l'équipe d'accueil ; rapport de stage.

Echéance : semaine 24