

## OFFRE de STAGE M1 – ANSES – SPAAD

Durée : minimum de 2 mois (stage rémunéré)

Localisation : ANSES, service SPAAD, Maisons-Alfort (94)

Contact : [virginie.chesnais@anses.fr](mailto:virginie.chesnais@anses.fr)

### Impact des outils d'assemblage *de novo* sur l'annotation fonctionnelle des génomes assemblés

En cas de toxi-infection alimentaire (TIAC) ou plus généralement d'épidémie bactérienne en lien avec l'alimentation et/ou les animaux de rentes, les équipes de bioinformatiques de l'ANSES ont développé un workflow d'analyse de données de séquençage haut débit Illumina (reads pairés) pour fournir l'ensemble des résultats attendus pour évaluer permettant de valider la qualité des données de séquençage, assembler les génomes bactériens et ainsi typer les agents pathogènes. Par cela, un choix d'outil a été fait, notamment pour la partie assemblage des données en se basant sur i. des études bibliographiques et ii. des benchmarks réalisés en interne. Un travail en cours de publication a récemment démontré l'impact des différentes méthodes d'assemblage sur les résultats de typage par cgMLST, du fait d'une fragmentation plus ou moins importante des génomes ainsi obtenus.

Nous souhaitons poursuivre ce travail afin d'évaluer l'impact de ces différentes qualités d'assemblage sur la caractérisation fonctionnelle des génomes obtenus. Ainsi, à partir des différents assemblages d'agents pathogènes obtenus dans notre précédent benchmark, l'étudiant devra réaliser des annotations fonctionnelles en utilisant un pipeline développé en interne : annotation de gènes *de novo*, recherche de gènes de virulence et de résistance, recherche de phages et de plasmides... Les résultats permettront de i. vérifier la répétabilité des résultats d'annotation fonctionnelles, ii. évaluer la reproductibilité des analyses sur des répliques de séquençage et iii. appréhender l'impact des méthodes d'assemblage de génomes bactériens sur la qualité de l'annotation.