

Décryptage des fonctions du microbiote rhizosphérique : optimisation du traitement bioinformatique des données obtenues par spectrométrie de masse.

Encadrant :

[Willy Bienvenu](#), chargé de recherche CNRS,
Email : Willy.Bienvenu@cnr.fr; Willy.Bienvenu@inrae.fr
Tel. : 01 69 33 23 41

Laboratoire

[GQE – le Moulon](#), [IDEEV](#)
12 route 128,
91190 Gif sur Yvette

Descriptif du laboratoire :

L'UMR **Génétique Quantitative et Évolution-Le Moulon** est membre fondateur de l'Institut Diversité, Écologie et Évolution du Vivant ([IDEEV](#)). L'UMR accueille des scientifiques de différents instituts et universités : INRAE (UMR0320), CNRS (UMR8120) et Université Paris-Saclay (AgroParisTech et Faculté des Sciences).

Une spécificité de l'**UMR Génétique Quantitative et Evolution – Le Moulon** est de mêler approches expérimentales et théoriques pour produire des connaissances sur la génétique et l'évolution des caractères quantitatifs dans les populations, avec un intérêt particulier pour les plantes cultivées et des finalités en lien avec l'agriculture.

Nos recherches couvrent différents champs disciplinaires de la biologie : biologie théorique, biologie évolutive, génétique, génomique, sciences agronomiques, mathématiques (biostatistiques et modélisation mathématique), bio-informatique. Dans cet environnement multidisciplinaire, notre spécialité est la génétique et génomique des populations pour des caractères quantitatifs observés à différentes échelles d'intégration dans des environnements contrastés : phénotypes moléculaires, caractères de développement ou d'architecture, composantes du rendement, caractères adaptatifs.

Notre production scientifique est mondialement reconnue. Pour la transition agroécologique, nous contribuons à la valorisation de la diversité cultivée chez le blé (sélection participative), le maïs (sélection assistée par marqueurs, sélection génomique) et en concevant des mélanges de variétés ou d'espèces à cultiver sur des parcelles agricoles.

Sujet de stage

L'étude du microbiote intestinal a montré l'intérêt qu'il y a à mieux connaître les espèces bactérienne qui interagissent avec l'hôte et peuvent influencer sa santé¹. De façon similaire, les plantes ont, elles aussi, leur propre microbiote autour de leur système racinaire : le rhizobiote. La rhizosphère, cette zone riche en microorganismes entourant les racines, constitue un véritable écosystème dynamique où se jouent des interactions complexes prennent place. Grâce à la sécrétion de molécules spécifiques, les racines façonnent leur environnement microbien en sélectionnant des espèces et des fonctions spécifiquement adaptées à leurs besoins. Décrypter les mécanismes offre des perspectives à de nouvelles

stratégies pour optimiser la croissance des plantes et augmenter leur résistance aux stress biotiques et abiotiques ².

La caractérisation des protéines par des approches de protéomique repose sur la connaissance préalable de la séquence de ces protéines souvent dérivées de génomes déjà séquencés. Cependant, dans le cadre des études de métaprotéomique, les génomes des organismes bactériens présents dans des environnements complexes tels que les selles ou le sol, sont rarement disponibles et/ou connus. Bien que la diversité taxonomique des communautés microbiennes de la rhizosphère soit relativement restreinte et principalement associée aux *Actinobacteria*, *Bacteroidetes*, *Firmicutes* et *Proteobacteria* ³, la majorité des bactéries peuplant le sol sont non-cultivables, et de génome inconnu.

Pour pallier cette difficulté, il est crucial de coupler la métaprotéomique avec des approches de métagénomique ou de métabarcoding, afin de générer une banque de séquences protéiques composite. Celle-ci pourra ensuite être utilisée pour identifier les protéines et peptides extraits du microbiote rhizosphérique. Toutefois, les données issues du séquençage métagénomique présentent souvent des défis majeurs, avec des assemblages complexes et une qualité parfois discutable. La construction d'une banque protéique exhaustive est donc essentielle et influence directement la qualité des résultats obtenus. Un équilibre subtil doit être trouvé : une banque trop large risque d'augmenter les faux positifs et le temps de traitement, tandis qu'une banque trop restreinte pourrait passer à côté de certaines protéines importantes présentes dans l'échantillon.

Récemment, plusieurs approches innovantes de retraitement des données de spectrométrie de masse ont été publiées, et deux d'entre elles se démarquent particulièrement. Jouffret *et al.* ⁴ ont exploré différentes stratégies de construction de la base de données nécessaire à l'identification des peptides/protéines en combinant des séquences issues de métagénomique et de génome bactériens disponibles dans les bases de données publiques. De son côté, Lee *et al.* ⁵ propose une approche novatrice basée sur le séquençage peptidique *de novo* grâce à un modèle d'apprentissage: Kaiko. Cette approche a l'avantage de pouvoir identifier les organismes présents dans des extraits de sol sans nécessiter de métagénomique ou de métabarcoding.

L'objectif de ce stage sera de comparer ces deux méthodes en utilisant les données de métaprotéomique que nous avons générées à partir d'un échantillon de sol agricole. En collaboration avec les bio-informaticiens de l'équipe, vous aurez pour mission de mettre en place les outils nécessaires à cette comparaison, de formater les données expérimentales pour qu'elles puissent être traitées par les deux approches, et d'évaluer la performance de chaque méthode en termes de temps d'exécution, de qualité des résultats et de capacité à représenter fidèlement le rhizobiote, en comparaison aux approches de séquençage par métabarcoding et métagénomique.

Profil recherché :

Le/La candidat/e devra être titulaire d'une formation en bio-informatique/informatique et/ou biostatistique et posséder une bonne maîtrise en programmation en R et Python. Une formation (ou complément de formation) en chimie, biologie et/ou biochimie sera particulièrement apprécié.

Informations complémentaires

L'offre de stage s'adresse aux étudiants en Master 2 pour l'année 2024-25, pour une durée de 5 à 6 mois avec gratification (~580 euros par mois, remboursement partiel des titres de transport). Les dates de début et fin sont flexibles.

Bibliographie :

- (1) Henry, C.; Bassignani, A.; Berland, M.; Langella, O.; Sokol, H.; Juste, C. *Cells* **2022**, *11*.
- (2) Vincze, E. B.; Becze, A.; Laslo, E.; Mara, G. *Agriculture-Basel* **2024**, *14*.
- (3) Bulgarelli, D.; Schlaeppi, K.; Spaepen, S.; Ver Loren van Themaat, E.; Schulze-Lefert, P. *Annu Rev Plant Biol* **2013**, *64*, 807-838.
- (4) Jouffret, V.; Miotello, G.; Culotta, K.; Ayrault, S.; Pible, O.; Armengaud, J. *Microbiome* **2021**, *9*, 195.
- (5) Lee, J. Y.; Mitchell, H. D.; Burnet, M. C.; Wu, R.; Jenson, S. C.; Merkley, E. D.; Nakayasu, E. S.; Nicora, C. D.; Jansson, J. K.; Burnum-Johnson, K. E.; Payne, S. H. *J Proteome Res* **2022**, *21*, 2023-2035.