Centre de Recherche Armand Frappier Santé Biotechnologie

Microbial-based predictive modeling of wheat yield and grain baking quality

Par

Mohammad Numan Ibne Asad

Thèse présentée pour l'obtention

Du grade de Philosophiæ Doctor (Ph.D.) en biologie

Jury d'évaluation

Président du jury et examinateur interne	Richard Villemur INRS – Armand Frappier Santé Biotechnologie
Examinateur externe	David Walsh Université Concordia
Examinatrice externe	Bobbi Helgason Université of Saskatchewan
Directeur de recherche	Étienne Yergeau INRS – Armand Frappier Santé Biotechnologie
Co-directeur de recherche	Philippe Constant INRS – Armand Frappier Santé Biotechnologie

ACKNOWLEDGEMENT

Firstly, I would like to express my gratitude and appreciation to my supervisor Étienne Yergeau for his outstanding and continuous support throughout this research project. I am grateful for his mentorship and supervision, especially in helping to execute research objectives toward goal-oriented achievement. His dynamic leadership skills and follow-up guidance helped me to complete my research project successfully. I am grateful to my supervisor Prof. Étienne Yergeau for giving me the opportunity to work in his lab on such an interesting project and for providing substantial funding support. I am also grateful to my co-supervisor Prof. Philippe Constant for friendly discussions about the project and its progress during the early stages of my Ph.D. study. I would also like to thank him for providing the laboratory facility and infrastructure to pursue some of my soil experiments. I would like to thank Prof. Richard Villemur for allowing me to perform quantitative PCR experiments in his lab. I would like to thank all the committee members who contributed to the thesis evaluation process with great patience.

I feel that without the support of my wonderful colleagues and lab mates, I would not have been able to complete my Ph.D. project on time. I would say, they are always a part of this journey, from personal to professional, and scientific knowledge development. I feel that I got a great research team that has helped blossom my hidden passion for science and played an important role in acquiring the necessary skills for my future scientific career. In short, it is difficult to express my gratitude for their tremendous support throughout this research project. At this point, I would like to mention some of the current and past faces of our lab, whose important contributions and support have helped complete this project. In particular, I would like to thank former postdoc colleagues Dr. Hamed Azarbad, Dr. Xiao-Bo Wang, PhD and MS colleagues (Jessica, Emmy, Pranav, Anne, Sara Correa Garcia, Robert McGee, Liliana, Asma, Itumeleng) for their tremendous support and assistance throughout my thesis. During my studies at the INRS-AFSB center, I met some other Ph.D. colleagues whom I would like to thank for their nice and friendly chats.

I would like to thank God (*Alhamdulillah!*), who has given me the strength and a lot of blessings to complete this project. I would like to thank my wife Mrs. Ayesha Siddika, who joined my life as part of this journey, gave me a lot of support and mental resilience during the pandemic and throughout my studies. I would like to remember my parents' sacrifices, blessings, and constant support from long distances throughout my whole studies. I feel truly blessed for my father Mr. Asaduzzaman, who has given me the courage and blessings to face any challenge and struggle in every step of my foreign life. Finally, I would like to thank all my well-wishers, friends, seniors, and neighbors living abroad and at home for their cordial talks and support during this entire graduate study.

ABSTRACT

It is very difficult to predict crop yield and produce quality based solely on soil physicochemical parameters, as the net effect these parameters is strongly affected by microbes. For instance, the form of soil nitrogen changes due to nitrification and denitrification activities, which will influence N mobility, availability, and energetic efficiency for plant growth. It is crucial to include microbial parameters to better predict crop yields and produce quality. However, microbial communities vary spatially and temporally, and are very complex, so it is uncertain if robust models could be derived from microbial data. My goal for this thesis was to create microbial-based models to predict wheat grain quality and yields across time and space. I used two sampling schemes: 1) early season sampling of 80 wheat fields across the province of Québec (Chapter 2) and 2) repeated sampling of a single wheat field across a growing season (Chapter 3). For both these experiments, I measured a wide array of microbial parameters: 16S rRNA gene and ITS region amplicon sequencing, qPCR quantification of key N-cycle genes, and microbial community level carbon usage. Grain baking quality and grain yields were measured at the end of the growing season. I used linear regression with stepwise forward/backward (Chapter 2) or LASSO selection (Chapter 3), limiting the models in most cases to less than 10 microbial indicators. In Chapter 2, I was able to explain observed variation of wheat grain quality and yields with an accuracy of up to 90% across all fields. Many of the inputs selected in the models had a link with soil nitrogen availability (e.g., ammonia-oxidizers and denitrifiers abundance). My microbial-based models also outperformed similar models based on commonly measured soil parameters (pH, total C, total N, C/N ratio, water content). However, in this Chapter, I had sampled the fields early in the growing season, and it was not certain that this was the best to create my predictive models. In Chapter 3, I then sampled our experimental field every two weeks across a single growing season to find the moment where the microbial predictive power was highest. My models highlighted a set of microbial parameters that were highly coherent with Chapter 2. I also found that the highest predictive power for wheat grain quality and yields was early in the season (May-June), which correlates with wheat growth stages that are critical for N nutrition. The results of my thesis show that it is possible to explain observed variation of wheat grain quality and yields from a few microbial parameters taken early in the season, and that these models are robust across a wide range of fields at the provincial scale. It also highlighted the key role of microorganisms involved in the nitrogen cycle for wheat quality and yields. In the longer term, these models could help farmers make management decisions for optimal grain quality and quantity, on our way toward a microbial-driven agriculture.

Keywords: amplicon sequencing, soil microbiome, nitrogen cycle, stepwise regression, Lasso regression, grain quality

RÉSUMÉ

Il est très difficile de prévoir le rendement des cultures et la qualité des produits en se basant uniquement sur les paramètres physicochimiques du sol, car l'effet net de ces paramètres est fortement influencé par les microbes. Par exemple, la forme de l'azote du sol change en raison des activités de nitrification et de dénitrification, ce qui influencera la mobilité, la disponibilité et l'efficacité énergétique de l'azote pour la croissance des plantes. Il est essentiel d'inclure les paramètres microbiens pour mieux prévoir les rendements des cultures et la qualité des produits. Cependant, les communautés microbiennes varient dans l'espace et dans le temps et sont très complexes, de sorte qu'il n'est pas certain que des modèles robustes puissent être dérivés des données microbiennes. L'objectif de cette thèse était de créer des modèles microbiens pour prédire la qualité des grains de blé et les rendements dans le temps et l'espace. J'ai utilisé deux schémas d'échantillonnage : 1) l'échantillonnage en début de saison de 80 champs de blé dans la province de Québec (chapitre 2) et 2) l'échantillonnage répété d'un seul champ de blé au cours d'une saison de croissance (chapitre 3). Pour ces deux expériences, j'ai mesuré un large éventail de paramètres microbiens : séquençage d'amplicons du gène de l'ARNr 16S et de la région ITS, quantification par qPCR des gènes clés du cycle de l'azote et mesure de l'utilisation du carbone au niveau de la communauté microbienne. La qualité boulangère des grains et les rendements en grains ont été mesurés à la fin de la saison de croissance. J'ai utilisé la régression linéaire avec une sélection « stepwise backward /forward » (chapitre 2) ou LASSO (chapitre 3), en limitant les modèles à moins de 10 indicateurs microbiens dans la plupart des cas. Dans le chapitre 2, j'ai pu prédire la qualité des grains de blé et les rendements avec une précision allant jusqu'à 90 % dans tous les champs. De nombreux intrants sélectionnés dans les modèles avaient un lien avec la disponibilité de l'azote dans le sol (par exemple, l'abondance des bactéries oxydants l'ammoniac et des bactéries dénitrifiantes). Mes modèles microbiens ont également surpassé des modèles similaires basés sur des paramètres du sol couramment mesurés (pH, C total, N total, rapport C/N et teneur en eau). Cependant, dans ce chapitre, j'ai échantillonné les champs au début de la saison de croissance, et il n'était pas certain que ce soit le meilleur moment pour créer mes modèles prédictifs. Dans le chapitre 3, j'ai ensuite échantillonné notre champ expérimental toutes les deux semaines pendant une seule saison de croissance afin de trouver le moment où le pouvoir prédictif microbien était le plus élevé. Mes modèles ont mis en évidence un ensemble de paramètres microbiens très cohérents avec le chapitre 2. J'ai également constaté que le pouvoir prédictif le plus élevé pour la qualité des grains de blé et les rendements se situait au début de la saison (mai-juin), pendant les stades de croissance du blé qui sont critiques pour la nutrition azotée. Les résultats de ma thèse montrent qu'il est possible de prédire avec précision la qualité des grains de blé et les rendements à partir de quelques paramètres microbiens mesurés en début de saison et que ces modèles sont robustes à l'échelle provinciale. L'étude a également mis en évidence le rôle clé des micro-

organismes impliqués dans le cycle de l'azote pour la qualité et le rendement du blé. À plus long terme, ces modèles pourraient aider les agriculteurs à prendre des décisions pour optimiser la qualité et la quantité des grains, vers une agriculture microbiocentrique.

Mots-clés : séquençage d'amplicons, microbiome du sol, cycle de l'azote, régression pas à pas, régression Lasso, qualité des céréales.

SYNOPSIS

L'azote est un élément crucial pour la production végétale, en particulier pour les cultures qui nécessitent des niveaux d'azote plus élevés à leurs différents stades de croissance. Certaines cultures, notamment le blé, ont besoin de grandes quantités d'azote pour soutenir leur croissance, la synthèse des protéines et améliorer la qualité des grains. Toutefois, la fertilisation azotée excessive a été signalée comme un problème majeur dans la production de blé. Le défi actuel est de réussir à corréler précisément les taux de fertilisation azotée avec le rendement du blé et la qualité des grains. L'application inadéquate d'engrais azotés conduit à un gaspillage catastrophique d'azote dans l'environnement. Les conséquences les plus importantes de N non utilisé sont l'eutrophisation des eaux de surface et les émissions de gaz à effet de serre N_2O. La plupart des processus biogéochimiques dans le sol qui produisent les nutriments et les gaz atmosphériques nécessaires à tous les organismes vivants au-dessus et au-dessous du sol sont principalement réalisés par les micro-organismes du sol. On pense en particulier que le traitement et la disponibilité de l'azote dans le sol sont fortement liés au recyclage de l'azote par les microorganismes. Les micro-organismes du sol transforment l'azote en différentes formes par des étapes séquentielles, au cours desquelles diverses formes d'azote ionisées sont produites par des réactions oxydatives ou réductrices. En général, les microbes utilisent différentes formes d'azote comme accepteurs d'électrons (par exemple, les dénitrifiants) ou sources d'énergie (par exemple, les nitrifiants), créant souvent un environnement compétitif pour l'absorption nette d'azote par les plantes en affectant la disponibilité totale de l'azote du sol. Même la respiration microbienne est corrélée au potentiel redox de l'environnement du sol, et souvent certaines communautés entretiennent une relation syntrophique avec différents groupes microbiens qui donnent des électrons pour accélérer une certaine réaction géochimique microbienne. Par exemple, les communautés microbiennes oxydant l'hydrogène exercent des effets différentiels sur les bactéries respirant le CH_4 et le CO_2 en augmentant le métabolisme microbien du carbone, régulant ainsi la dénitrification du sol et les émissions de NO_2. Par conséquent, le microbiome du sol associé au cycle de l'azote et des autres nutriments joue un rôle décisif dans l'accessibilité des nutriments pour les plantes.

Les microbiomes sont considérés comme de puissants intégrateurs des processus passés et présents de l'écosystème, fournissant de multiples niveaux d'information sur la fonction du sol. Ils détiennent une fonction d'indicateur dans le sol et expliquent les états contemporains de l'écosystème. La plupart des processus de transformation des nutriments du sol sont déterminées par les interactions multitrophiques du microbiome du sol et des propriétés physicochimiques du sol. Ainsi, le microbiome du sol joue directement ou indirectement un rôle important en influençant la croissance et la productivité des plantes par le cycle

des nutriments, le maintien de la fertilité du sol et la séquestration du carbone. Par conséquent, il est assez difficile et souvent imprécis de définir les fonctions biochimiques du sol ou les taux de processus avec seulement quelques taxons ou certaines espèces microbiennes ou encore quelques propriétés physicochimiques. Néanmoins, en général, les changements dans la diversité microbienne du sol peuvent modifier le statut nutritif du sol qui, à son tour, peut influencer la colonisation active du microbiome de la rhizosphère dans les racines des plantes en modifiant l'environnement de l'interface racine-sol. Ainsi, les processus d'assemblage de la communauté des microbiomes indigènes des plantes et leur fonction sont fortement corrélés (par exemple, la diversité alpha et bêta) avec la diversité microbienne centrale et la structure de la communauté dans le sol éloigné (bulk soil). La connaissance précise de la façon dont les facteurs biotiques et abiotiques, ainsi que la gestion agricole, influencent les processus d'assemblage des communautés de microbiomes pour façonner les modèles de cooccurrence microbienne et les interactions culture-microbiome le long de l'axe sol-plante-racine est encore limitée. En réponse aux facteurs environnementaux exogènes et endogènes du sol, le microbiome du sol fournit un niveau différent de signaux de l'état actuel de processus écosystémiques donnés par le biais de leurs changements structurels. Par conséquent, ces signaux de l'état d'un agroécosystème donné dérivés du microbiome du sol central sont très utiles pour comprendre le modèle de relations entre le sol et les différents traits et la productivité des plantes.

Les indicateurs microbiens liés à la diversité et à la composition microbienne ont un fort pouvoir prédictif qui peut dépeindre l'état actuel des processus des agroécosystèmes. La capacité du microbiome du sol à prédire les propriétés biologiques et physicochimiques des sols agricoles, ainsi que le rendement et la qualité des cultures dans différents environnements, a déjà été démontrée dans plusieurs études. L'augmentation de la teneur en azote des feuilles et des grains est fortement corrélée à une absorption efficace de l'azote par les cultures. Ainsi, la séquestration du carbone et la décomposition de la matière organique du sol peuvent contribuer à une meilleure minéralisation de l'azote. Une décomposition plus importante de la matière organique peut libérer de l'azote facilement utilisable pour l'absorption par les cultures, ce qui peut favoriser la croissance des cultures et la synthèse des protéines des grains. Par conséquent, le métabolisme microbienne de différentes sources de carbone organique pourrait être une variable intéressante. L'intensité de l'utilisation microbienne de sources de carbone spécifiques indique l'activité de guildes fonctionnelles microbiennes spécifiques qui peuvent être directement liées au métabolisme de l'azote des plantes. Par conséquent, les indicateurs liés à la décomposition du carbone organique du sol pourraient être des prédicteurs utiles pour modéliser le grain et la qualité du blé. Le sol sert de réservoir microbien aux plantes pour recruter le microbiome par le biais de processus d'assemblage de communautés. La plante fournit des nutriments aux microbes par l'exsudation des racines, ce qui influence la diversité microbienne globale du sol peut influer sur la dispersion et la succession des microbes

dans les différents compartiments de la plante qui suivent l'axe sol-racine. Cette caractéristique de l'assemblage des communautés dans un agroécosystème pourrait avoir un impact sur l'ensemble des processus physiologiques des plantes en modulant les interactions plante-microbiome. Les signaux dérivés de la diversité microbienne décrivant la structure ou la composition globale des communautés du microbiome du sol peuvent être des prédicteurs potentiels du rendement et de la qualité des cultures. L'abondance de certaines communautés microbiennes associées peut être liée directement ou indirectement au cycle des nutriments du sol. Les indices de richesse de la diversité microbienne spécifique peuvent être directement ou indirectement associés à l'utilisation de l'azote dans les grains et à la synthèse des protéines dans les grains. Il existe des preuves substantielles que la diversité microbienne du sol joue un rôle important dans la régulation du cycle de l'azote, comme on peut l'observer pendant l'utilisation et le traitement microbien de l'azote dans un système fermé. La composante abiotique associée au climat ou à la situation géographique influence également la structure et l'activité microbienne, ce qui pourrait expliquer la variation de la qualité des grains de blé. Le changement climatique persistant peut favoriser une pression sélective pour la survie microbienne par le biais de changements dans les propriétés du sol et la dynamique de l'azote, établissant ainsi une nouvelle structure de la communauté microbienne suivant un processus déterministe. D'autres facteurs environnementaux et biotiques peuvent modérer les propriétés physicochimiques du sol, ce qui les rend moins efficaces pour expliquer directement les processus de l'écosystème. En revanche, la diversité, l'abondance et la fonction microbiennes ont un plus grand pouvoir prédictif. Lorsqu'elles disposent d'une niche optimale, les communautés microbiennes peuvent s'intégrer aux propriétés et aux nutriments actuels du sol, ce qui se traduit par des réponses diverses aux processus écosystémiques. L'analyse de l'abondance, de la diversité et des réponses fonctionnelles des communautés microbiennes permet de mieux comprendre l'impact des changements environnementaux sur ces communautés. Ces informations sont précieuses pour comprendre les effets de divers facteurs sur les communautés microbiennes du sol. Les profils ou les abondances des communautés microbiennes sont donc mieux corrélés avec l'activité fonctionnelle au niveau de la communauté et peuvent fournir un meilleur potentiel prédictif global pour expliquer le rendement du blé et la qualité du grain.

Les développements récents de la technologie omique peuvent être exploités pour explorer les profils taxonomiques microbiens du sol à haute résolution, permettant la capture d'une large gamme de diversité microbienne à la fois à l'échelle spatiale et temporelle. Il a été proposé que les approches multiomiques intégrées permettent d'observer l'interaction complexe entre les métabolites du sol, les minéraux et les propriétés microbiennes. Cette approche intégrée de la technologie multi-omique nous a permis de démêler la relation entre la productivité des plantes, l'activité microbienne (par exemple, la décomposition de la matière organique du sol) et la diversité microbienne. Ainsi, la technologie omique est potentiellement utile pour prédire les interactions à plusieurs niveaux entre les plantes, les microbes et le

sol. Par exemple, l'utilisation de techniques métataxonomiques ou métagénomiques basées sur l'ADN peut considérablement améliorer les efforts de recherche et approfondir notre compréhension des relations complexes entre les facteurs abiotiques du sol et la diversité et la fonction microbiennes. Grâce à ces méthodes, il peut être possible de limiter le temps et les ressources tout en obtenant des informations importantes sur l'interaction complexe entre les micro-organismes et la santé des sols. . Les données métataxonomiques peuvent déchiffrer le modèle et les principaux acteurs des groupes microbiens impliqués dans la dénitrification dans des conditions de sol acide, nous permettant de comprendre leur régulation sur les intermédiaires dénitrifiés dans un écosystème individuel.

Comme mentionné plus haut, le microbiome du sol joue un rôle crucial dans la transformation et la disponibilité des nutriments du sol et influence ainsi l'abondance des nutriments des plantes dans le sol. Le cycle de l'azote, piloté par de nombreuses communautés microbiennes aérobies et anaérobies, est la principale usine de traitement de l'azote du sol. L'incorporation de différents microbes du sol dans le cycle des nutriments (par exemple, le cycle de l'azote) dépend principalement de facteurs écologiques, notamment la diversité microbienne du sol, les propriétés physico-chimiques du sol, le climat et la séquestration du carbone. Il est clair que les taux de transformation de l'azote ne peuvent pas être déterminés uniquement par l'activité des nitrifiants et des dénitrifiants, ou d'une communauté microbienne particulière. Il existe sans aucun doute d'autres variables liées aux interactions biotiques qui peuvent influencer ce processus. . Par conséquent, les interactions biotiques entre tous les micro-organismes du sol dans un environnement restreint jouent un rôle décisif dans l'apport ultérieur d'azote aux plantes. Ici, nous nous sommes principalement concentrés sur l'étude du rôle des communautés bactériennes, archéennes et fongiques du sol, car elles sont considérées comme les meilleurs micro-habitants parmi les autres microorganismes du sol. Nous avons évalué l'incorporation d'indicateurs microbiens du sol dans les analyses de sol basées sur les nutriments du sol dans l’optique de mieux guider la prise de décision concernant la gestion des engrais dans la production végétale. Par conséquent, mon hypothèse centrale est que les microbes du sol, en raison de leur rôle central dans le cycle des nutriments et la santé des plantes, contiennent un signal qui peut être utilisé pour prédire les rendements de blé et la qualité des grains. Mon objectif principal était de mesurer les propriétés physicochimiques de base du sol, le potentiel fonctionnel microbien, la diversité, l'abondance et la composition de la communauté dans le temps et l'espace et de trouver les paramètres les plus significatifs expliquant les rendements de blé et la qualité boulangère des grains.

L'analyse de la littérature sur les processus agroécosystémiques a permis de mettre en évidence l'existence de plusieurs facteurs importants qui influencent de manière significative les résultats finaux. Ces facteurs comprennent la gestion de l'exploitation, l'environnement du sol, le type de sol et la diversité des cultures. J'ai donc émis l'hypothèse que des signaux microbiens spécifiques peuvent être détectés au début

de la saison de croissance du blé et que ces signaux ont une relation directe avec le rendement et la qualité des grains.. Pour tester ma première hypothèse, j'ai analysé les indicateurs microbiens pour déterminer le potentiel fonctionnel microbien, la diversité, l'abondance et la composition des communautés, ainsi que les propriétés physiques de base du sol dans plus de 80 champs de blé au Québec. J'ai planifié un échantillonnage du sol à grande échelle dans les fermes de blé du Québec, avec l'idée que nous pourrions obtenir des indicateurs microbiens robustes à une échelle spatiale qui pourraient être utilisés pour prédire le rendement du blé et la qualité du grain. J'ai également émis l'hypothèse que la collecte d'échantillons de sol tôt dans la saison de croissance du blé pourrait être un bon moment pour extrapoler le pouvoir prédictif microbien pour le rendement et la qualité du blé à la fin de la récolte. Par conséquent, l'objectif principal était de collecter des échantillons de sol dans des fermes du Québec afin de surveiller la variation spatiale de la diversité, la composition et les fonctions microbiennes du sol sous différentes formes de gestion agricole. Dans des études précédentes, il a été rapporté qu'en utilisant des indicateurs microbiens, il est possible de prédire le statut nutritif et biologique du sol à l'échelle continentale. Dans mon cas, l'objectif principal de la modélisation prédictive du rendement et de la qualité du blé était d'étudier dans quelle mesure cette approche de modélisation pouvait démontrer le pouvoir prédictif des communautés microbiennes du sol à l'échelle du champ, en considérant l'hypothèse clé de ce travail de recherche. Nous savons que les environnements hétérogènes du sol, y compris les facteurs édaphiques tels que la disponibilité du carbone dans le sol, le pH du sol, la température, l'humidité du sol, la perméabilité à l'oxygène du sol et le potentiel redox, modulent la diversité du microbiome et abritent une écologie systématique. Les changements de tous ces facteurs édaphiques sont principalement soumis au climat et à la gestion agricole, notamment la sécheresse, les émissions de gaz à effet de serre, l'intensification de l'utilisation des terres, les pesticides, la résistance aux antimicrobiens, etc. Le métabolisme microbien du carbone, y compris l'utilisation des polysaccharides, des acides aminés, des acides carboxyliques et des acides gras, est un indicateur essentiel du cycle du carbone et de l'azote organique du sol et de l'activité physiologique des communautés microbiennes hétérotrophes. Les différentes communautés microbiennes utilisant différentes sources de carbone servent également d'indicateurs potentiels des émissions de gaz à l'état de traces (par exemple, CO_2, N_2O, H_2 etc.) produites par la respiration microbienne. Le modèle d'utilisation du carbone microbien est fortement lié à la décomposition de la matière organique du sol. Le taux de décomposition de la matière organique du sol dépend également du ratio total de la biomasse bactérienne et fongique dans le sol. La biomasse fongique contribue largement au rapport total carbone/azote du sol (rapport C: N) en recyclant (par exemple, la nécromasse) plus de carbone organique que d'azote, tandis que les bactéries recyclent plus d'azote dans la biomasse bactérienne totale. Une plus grande utilisation du carbone organique à partir des polymères d'acides aminés peut profiter à certaines communautés microbiennes qui peuvent contribuer à un plus grand stockage de carbone organique dans le sol. Certaines études ont montré que les protistes

bactérivores construisent une biomasse dont le rapport C/N est plus élevé que celui de leurs proies bactériennes et qu'ils libèrent de l'ammoniac dans le sol sous forme de déchets bactériens, ce qui peut contribuer à la disponibilité de l'azote dans le sol. Par conséquent, le cycle des nutriments induit par les microorganismes dans un écosystème agricole présentant une dynamique C: N spécifique est très sensible à l'exposition continue à l'azote organique ou inorganique fourni. Différentes pratiques agricoles et différentes conditions climatiques à travers les champs de blé nous ont donné l'occasion d'estimer l'effet des paramètres microbiens liés au rendement et à la qualité du blé. Ainsi, le plan expérimental visant à modéliser le rendement du blé et la qualité des grains dans les exploitations de blé situées sur un transect de 500 km nous permet d'identifier les indicateurs microbiens potentiels des processus agroécosystémiques pertinents qui prédisent avec précision la qualité des grains de blé.

Des échantillons de sol ont été récoltés dans 80 fermes à travers le Québec. Dans chaque champ, des échantillons composites ont été produit en creusant à 5 points d'échantillonnage, à une profondeur de 10 cm. Si des plants de blé étaient présents dans le champ, les échantillons ont été récoltés de 10 à 25 cm d'intervalle.

Les échantillons de sol ont été divisés et utilisés pour différentes analyses biochimiques et biologiques. La teneur en eau et le pH du sol a été mesurés. Le carbone total et l'azote total ont été mesurés par un analyseur d'élément, suivant la méthode de combustion. L'ADN génomique a été extrait des sous-échantillons de sol à l'aide d'un kit commercial en suivant le protocole du fabricant. Les librairies de séquençage d'amplicons de l'ARNr 16S et ITS1 ont été préparées en amplifiant des ensembles d'amorces spécifiques à la cible en suivant le protocole de préparation des librairies de séquençage NGS d'Illumina. Les amplicons d'ARNr 16S et d'ITS1 ont été envoyés pour un séquençage NGS Illumina-MiSeq en paires en générant des librairies d'amplicons méta-barcodés et en les regroupant dans un seul tube. . Les données de séquençage ont été analysées en suivant les pipelines internes d'amplicon Tagger. Le tri initial des séquences a été effectué sur la base du code-barres de séquençage et la qualité de la diversité a été assurée en scannant PhiX-Spike dans les séquences. Les lectures uniques et de faible qualité donnant lieu à des séquences inférieures au seuil du score Phred ont été supprimées. Les variantes de séquences d'amplicons ont été générées en suivant le pipeline DADA2. Les lignées taxonomiques ont été attribuées avec un classificateur RDP en utilisant la base de données Silva et un seuil minimum de scores de correspondance des taxons a été établi. Plusieurs gènes fonctionnels associés au cycle de l'azote ont été quantifiés par PCR quantitative. On a examiné l'abondance fonctionnelle des gènes liés au processus de nitrification et de dénitrification du sol. Plus précisément, l'abondance du gène de la monooxygénase bactérienne (AOB) et archéennes (AOA), de la nitrite réductase (*nirK*) et de la réductase de l'oxyde nitreux (*nosZ*) a été mesurée parce que ces données peuvent fournir un indicateur potentiel du statut de l'azote du sol dans

l'agroécosystème actuel. En outre, l'abondance absolue des copies de gènes fonctionnels liés au cycle de l'azote peut conférer un trait microbien significatif qui peut contribuer de manière optimale à des processus donnés de l'agroécosystème, ce qui à son tour peut avoir une rétroaction négative ou positive sur le processus résultant en des différences dans le rendement des cultures. Le ratio champignon/bactérie basé sur l'abondance des gènes 16S et ITS région a également été mesuré par PCR quantitative Le ratio champignon/bactérie a été analysé comme un indicateur potentiel des processus de l'écosystème, afin de vérifier si ces ratios affectent le rendement du blé et la qualité du grain. La qualité du grain de blé et la qualité boulangère ont été mesurées en collaboration avec notre partenaire, les Moulins de Soulange, afin d'obtenir les indices de qualités des grains couramment utilisés pour évaluer la qualité boulangère. Toutes les analyses de données exploratoires, y compris la distribution des données, les tests d'hypothèse et l'analyse factorielle multiple, ont été effectuées pour examiner les modèles de données et les facteurs qui influencent le plus le rendement du blé et la qualité boulangère des grains. L'imputation et la transformation des données ont été effectuées à l'aide de divers progiciels R pour les données qui n'étaient pas normalement distribuées ou qui présentaient des valeurs aberrantes extrêmes. Pour analyser les traits microbiens potentiels impliqués dans les processus de l'écosystème expliquant le rendement et la qualité du blé, nous avons effectué un test de corrélation de rang de Spearman. Pour sélectionner les prédicteurs microbiens potentiels qui ont un effet significatif sur le rendement et la qualité du blé, nous avons effectué des approches de sélection pas à pas couplées à une fonction linéaire (lm). Pour l'analyse comparative des modèles, la méthode de sélection pas à pas simplifie l'effort de sélection des 5 prédicteurs les plus importants avec une erreur standard résiduelle faible. Les compromis biais-variance et la performance ont été évalués entre les modèles avec différents paramètres statistiques.

Il a été déterminé que le rendement et la qualité du blé varie significativement entre les différentes fermes échantillonnées. De plus, la variété du blé cultivé a eu un effet significatif sur le rendement et la qualité du blé. Il a été observé qu'il existe une relation significative entre les propriétés physicochimiques du sol et le rendement et la qualité des grains de blé. Ceci était attendu, puisque les propriétés physicochimiques du sol sont des facteurs importants dans les processus indiquant la fertilité du sol. En effet, ces propriétés sont fortement impliquées dans les cycles des nutriments du sol et dans la création de microenvironnement abritant les microorganismes.

L'utilisation du carbone microbien à partir des sources contenant des acides aminés a montré une corrélation potentielle avec la qualité du blé. Les ASVs appartenant aux taxons bactériens et archéens incluant *Acidobacteria*, *Actinobacteria* et *Proteobacteria* étaient dominants parmi les échantillons prélevés dans différentes régions. Les ASV liés aux *Planctomycètes* et aux Actinobactéries variaient selon les régions, parmi les phyla avec une abondance relative moyenne supérieure à 1%. La communauté fongique

des échantillons était dominée par des communautés saprophytes telles que les *Agaricomycetes*, les *Mortierellomycotina* et les *Sordariomycetes*. Il y avait également une corrélation significative entre les phyla fongiques relativement abondants (par exemple, *Ascomycota*, *Basidiomycota* et *Zygomycota*) et la qualité de la farine. La richesse et l'abondance des espèces associées à l'indice de diversité de Shannon pour les bactéries et les archées étaient négativement corrélées à la qualité du grain de blé, tandis que la diversité de Shannon pour les communautés fongiques était positivement corrélée au rendement et à la teneur en amidon et en protéines du grain. Enfin, les ASV microbiens les plus corrélés ont révélé une influence différente sur le rendement des plantes et la qualité des grains, ce qui montre un potentiel fonctionnel mécaniste des communautés microbiennes à des niveaux taxonomiques inférieurs. Par exemple, certaines ASV liées aux genres *Paenibacillus* et *Sphingomonas*, connues pour leur activité de promotion de la croissance, ont été fortement corrélées à la qualité de la cuisson de la farine (par exemple, le temps maximal de la farine). Même certains substrats de carbone provenant de sources d'acides aminés utilisés par les communautés microbiennes étaient significativement corrélés aux différents indicateurs de qualité de cuisson des céréales et des farines. Pour réduire le nombre de prédicteurs, les traits microbiens issus des données métataxonomiques (16S et ITS) ont été sélectionnés sur la base des corrélations les plus élevées entre les ASV et les paramètres de qualité du blé, les dix ASV microbiens les plus corrélés expliquant la relation monotone la plus élevée ayant été conservés comme caractéristiques clés. De même, les traits métaboliques du carbone les plus corrélés avec la qualité du grain et de la farine ont été sélectionnés sur la base des substrats de carbone utilisés par les communautés microbiennes. Dans un premier temps, un modèle prédictif basé sur les microbes du sol pour le rendement et la qualité du blé a été développé en incorporant tous les traits du microbiome à l'aide des variables explicatives suivantes : substrat de carbone utilisé par les microbes, ASV bactériens et fongiques associés au rendement du blé et à la qualité du grain, abondance des gènes liés au cycle de l'azote, rapports d'abondance des gènes ARNr 16S et ITS, descripteurs de la communauté microbienne (par exemple, diversité alpha et bêta). Le pouvoir explicatif du modèle linéaire augmente avec le nombre de prédicteurs inclus dans le modèle. Par conséquent, pour l'analyse comparative du pouvoir prédictif entre les indicateurs pédologiques et microbiens, une méthode statistique basée sur la sélection prospective a été appliquée pour sélectionner et réduire le nombre de prédicteurs significatifs. Pour remettre en question le modèle basé sur le sol et s'aligner sur la taille des prédicteurs du sol (pH, N total, C total, teneur en eau et rapport C: N) fortement liés aux processus de l'écosystème, seuls 5 prédicteurs microbiens pour le rendement du blé et la qualité du grain ont été sélectionnés. Il est intéressant de noter que le modèle microbien a toujours été plus performant que le modèle basé sur le sol pour prédire le rendement du blé et la qualité des grains, et qu'il a montré l'erreur résiduelle la plus faible et une grande précision. En revanche, même certains paramètres basés sur le sol n'ont pas réussi à prédire les qualités du grain en raison d'un mauvais ajustement dans la régression linéaire. Le modèle basé sur le sol

comprenait des paramètres fortement corrélés aux processus d'intérêt (tels que le rendement et la qualité du blé), ce qui a entraîné un biais élevé dans l'analyse VIF (facteur d'inflation de la variance). Un deuxième défi de modélisation était de savoir si, si les prédicteurs microbiens étaient inclus avec les prédicteurs du sol, le modèle permettrait toujours de prédire le rendement du blé et la qualité du grain. Pour répondre à ces questions en suspens sur la modélisation microbienne, un modèle combiné incorporant à la fois les données du sol et les données microbiennes dans une méthode de régression basée sur la sélection avant par étapes a été réalisé. Il a été constaté que les paramètres microbiens du modèle combiné composé de 10-11 variables explicatives (données pédologiques et microbiennes) permettaient encore de prédire avec précision (64-90% en précision) le rendement et la qualité du blé. Ces résultats suggèrent qu'en plus de l'analyse des nutriments du sol, la prise en compte des paramètres microbiens dans la gestion agricole pourrait contribuer à une évaluation plus précise du rendement et de la qualité du blé. Enfin, les modèles prédictifs avec les données du microbiome obtenues au début de la saison de croissance ont toujours été plus performants que les modèles basés sur le sol et les modèles combinés (sol et microbe), ce qui a pleinement répondu à notre premier objectif.

Il a été largement démontré et rapporté que le pH du sol, la teneur en eau et le rapport C : N jouent un rôle important dans l'établissement des communautés microbiennes du sol. C'est pourquoi il peut y avoir un lien potentiel entre les processus microbiens du sol et les indicateurs physicochimiques du sol. Mais les activités des micro-organismes du sol impliqués dans des processus particuliers ne sont pas nécessairement simples. Par exemple, la teneur en eau du sol peut déterminer la disponibilité des nutriments, mais l'équilibre des nutriments du sol et le rapport C: N sont régulés par l'activité microbienne. En outre, certains indicateurs du sol ont une forte corrélation avec le rendement du blé et la qualité du grain, mais ne sont pas directement impliqués dans les facteurs qui médient le traitement de l'azote organique ou inorganique, ce qui a un impact énorme sur l'efficacité de l'utilisation de l'azote par les plantes. Par conséquent, les paramètres du sol n'ont pas souvent prédit avec précision la qualité des grains de blé et de la farine. Même certains paramètres de qualité sont négativement associés à l'azote mesuré au début de la saison de croissance du blé, ce qui suggère un effet négatif de la fertilisation azotée sans discernement avec des apports intensifs. Après avoir analysé les paramètres microbiens à l'aide de diverses approches de modélisation, il a été déterminé que la teneur en azote n'avait pas d'effet direct sur la qualité des grains. Cela souligne l'importance de prendre en compte les microbiomes lors de la prise de décisions concernant les pratiques de fertilisation. En mettant en œuvre des pratiques agricoles qui intègrent les microbiomes du sol, il est possible de réduire l'apport excessif d'azote et d'augmenter ainsi la productivité des cultures. Cette approche peut fournir des résultats plus précis en matière de production végétale que les pratiques traditionnelles basées sur les nutriments du sol. Comme les paramètres du sol ont montré des signaux faibles dans le modèle prédictif, indiquant un manque de lien précis avec les processus du sol, l'inclusion de nombreux autres indicateurs du sol dans le modèle de

régression peut être plus performante que le modèle basé sur le microbiome, mais au prix d'une précision correcte du modèle. De nombreux prédicteurs sélectionnés dans les modèles basés sur le microbiome, tels que les descripteurs de la communauté, l'abondance des gènes fonctionnels et la capacité microbienne des substrats organiques, peuvent avoir des relations directes avec les processus du sol. Comme l'attendent les agriculteurs et les meuniers, les paramètres microbiens mis en évidence dans le modèle prédictif peuvent être rapidement testés et surveillés à l'aide d'outils plus spécifiques aux taxons ou en établissant un dosage biochimique pour une activité enzymatique microbienne spécifique. En outre, les données brutes des séquences d'amplicons de gènes spécifiques à une cible peuvent être utilisées pour évaluer la diversité bêta et alpha du microbiome du sol. L'abondance des ASV liés aux oxydants d'ammoniac *Nitrosospira* ou aux oxydants d'ammoniac complets (par exemple, *Commamox*) a été corrélée négativement avec la qualité boulangère t des grains. Cela signifie donc que les processus de l'écosystème du sol dirigés par le microbiome le plus abondant ont des effets néfastes sur la disponibilité de l'azote dans le sol et l'efficacité de l'absorption de l'azote par les plantes. En effet, l'ammoniac peut être absorbé passivement par les plantes en suivant les voies enzymatiques glutamate-glutamine synthétase-glutamine oxoglutarate aminotransférase. En revanche, les formes organiques d'azote telles que les acides aminés peuvent être directement absorbées par les plantes, car ces sources d'azote ont un meilleur rendement énergétique pour les plantes. Le nitrate (NO_3^-) peut être absorbé activement par les plantes mais il doit d'abord être réduit et converti en ammoniac, ce qui est un processus plus énergivore pour les plantes. De plus, le microbiome du sol associé à l'oxydation de l'ammoniac par les processus de nitrification joue un rôle important dans le maintien de l'équilibre entre l'ammoniac et le nitrate. En accord avec nos résultats, des études de terrain similaires ont montré que les versions archéales ou bactériennes du microbiote oxydant l'ammoniac ont des effets positifs ou négatifs sur la teneur en gluten et la teneur en protéines des grains. Il est difficile d'évaluer cette rétroaction positive ou négative de l'oxydation de l'ammoniac par le microbiote sur la synthèse des grains des plantes, car leur mécanisme peut dépendre plus indirectement du contraste du rapport AOA: AOB ou du rapport AOB: bactéries totales. Certains des paramètres spécifiques sélectionnés dans le modèle prédictif concernant l'utilisation des substrats carbonés étaient les acides aminés. Ça indique que les groupes microbiens associés à la dégradation du carbone organique ont un meilleur accès au traitement de l'azote organique qui peut améliorer la disponibilité de l'azote. De même, certains substrats organiques tels que le glucose-1-phosphate utilisé par les microbes indiquent que la décomposition efficace de l'amidon ou du glycogène peut être liée à la diversité du microbiome du sol et à une rétroaction potentiellement différentielle sur le fonctionnement de l'écosystème. Un autre facteur important associé aux processus de l'écosystème du sol est l'abondance du ratio champignons/bactéries. Une réponse positive d'un ratio champignons/bactéries plus élevé, indiquant des taux de décomposition plus élevés entraînés par la communauté fongique, peut ajouter plus d'azote au sol pour l'absorption par les plantes. Ceci est dû au fait

que les champignons ont besoin de moins d'azote par unité de biomasse que les bactéries. Par conséquent, les paramètres liés à l'axe PCoA fongique indiquant les descripteurs de la communauté fongique tels que les différences de ASV fongique entre les échantillons et le rapport fongique-bactérien ont été sélectionnés dans de nombreux modèles. Enfin, il est intéressant de noter que certaines des ASV sélectionnées dans les modèles appartiennent à des genres tels que *Paenibacillus* et *Sphingomonas*, connus pour leur activité de promotion de la croissance des plantes et présentant une corrélation négative avec la qualité des grains de blé et de la farine. Cette relation négative peut être due à la dilution de l'azote dans les grandes plantes qui réduit la qualité du grain. D'autres paramètres sélectionnés dans le modèle peuvent ne pas être directement liés au processus de synthèse du grain de blé. Les efforts de modélisation n'étaient pas destinés à se concentrer spécifiquement sur l'élucidation de la nutrition azotée du blé, mais plutôt à mettre en avant des prédicteurs significatifs qui ont des impacts potentiels sur la qualité du grain et de la farine. Par conséquent, certains paramètres peuvent ne pas être directement associés aux processus de l'écosystème, mais sont néanmoins utiles car ils peuvent covarier avec certains paramètres non mesurés, qui peuvent être associés aux processus qui créent un environnement propice à la production optimale de nutriments pour le blé. En outre, des efforts de modélisation similaires sous-tendant différentes conditions de terrain ont permis de sélectionner des paramètres complètement différents qui prédisent le rendement du blé et la qualité du grain avec des degrés de précision différents, ce qui indique un état différent des processus de l'agroécosystème. La plupart des paramètres microbiens obtenus au début de la saison de croissance ont montré un lien étroit avec le rendement du blé et la qualité du grain récolté à la fin, ce qui était cohérent avec d'autres études menées à la même période de culture du blé. Ainsi, le pouvoir prédictif microbien démontré au début de la saison de croissance du blé fournit un aperçu potentiel pour trouver une date d'échantillonnage optimale pour construire le meilleur modèle. En outre, ces résultats ouvrent de nouvelles voies pour la mise en place d'expériences de suivi pour une intervention précoce de l'activité microbienne du sol qui pourrait aider à gérer les processus microbiens d'intérêt dans un état souhaité. Mais les indicateurs microbiens robustes des interactions sol-culture identifiés dans cette étude peuvent guider les agriculteurs vers une meilleure gestion agricole pour produire des céréales de haute qualité avec peu d'intrants, pour ouvrir la voie à une production agricole durable.

Le deuxième objectif aborde principalement les questions liées à la manière dont la dynamique temporelle des communautés microbiennes dans le cadre des processus de l'agroécosystème affecte le rendement et la qualité des cultures. En particulier, le pouvoir prédictif des paramètres microbiens peut varier dans le temps, car la sélection des hôtes et l'acquisition des nutriments peuvent être influencées par la fixation du carbone à différents stades de la croissance du plant de blé. Il a été démontré que la diversité microbienne et les interactions hôte-microbiome organisées à l'interface racinaire et dans différents compartiments de la plante ont un lien significatif avec le stade de croissance de la plante. Comme les

propriétés du sol peuvent être affectées par la saison ou le climat régional, ce changement dans l'habitat microbien du sol peut limiter l'accès aux nutriments pour certaines communautés microbiennes. De tels changements temporels dans la diversité microbienne, la composition ou l'abondance fonctionnelle pendant toute la saison de croissance du blé modifieront nécessairement le pouvoir de prédiction du microbiome du sol, ce qui peut soulever des questions sur le moment le plus optimal pour une prédiction. Par conséquent, les méthodes expérimentales sont conçues en tenant compte du deuxième objectif qui se concentre principalement sur la recherche du meilleur modèle pour prédire la qualité du grain de blé à différents stades de croissance tout au long de la saison. Dans des études antérieures, il a été démontré que les indicateurs microbiens peuvent prédire le rendement du blé et la qualité du grain avec plus de précision que la fertilisation azotée. Mais pour trouver le meilleur moment pour prédire la qualité du grain de blé, un champ expérimental historique pluriannuel situé à l'INRS a été choisi pour l'échantillonnage du sol qui a été conçu avec 6 blocs aléatoires, y compris 4 traitements de manipulation des précipitations et deux génotypes de blé. Les génotypes de blé ont été sélectionnés sur la base de deux traits spécifiques, y compris les traits de tolérance à la sécheresse et de sensibilité à la sécheresse. Il a été enregistré que la teneur en eau du sol dans les parcelles traitées avec la manipulation des précipitations différait entre les échantillons prélevés à différentes dates. Le schéma d'échantillonnage a été mis en œuvre approximativement en fonction du stade de croissance du blé, du semis à la maturité de la culture. Des indices microbiens similaires utilisés pour la modélisation dans le premier objectif ont également été mesurés pour justifier leur pouvoir de prédiction à une échelle temporelle. Une autre étude réalisée avec la même expérience de terrain a montré que les épisodes d'assèchement et d'humidification du sol provoqués par des pluies soudaines à la mi-juillet ont modifié les communautés microbiennes et augmenté l'abondance des archées oxydant l'ammoniac. De tels épisodes sont assez courants dans le contexte du changement climatique récent, fait qui peut modifier le statut des nutriments du sol ou limiter la fonctionnalité du microbiome du sol. Un tel changement temporel des paramètres microbiens dû à l'influence de facteurs biotiques et abiotiques peut inverser les processus de l'écosystème et créer des environnements pédologiques plus complexes de l'échelle macro à l'échelle microscopique. Même ce changement temporaire dans la composition de la communauté microbienne peut perturber le flux des nutriments du sol et l'efficacité de l'absorption de l'azote par les plantes de blé.

Pour obtenir des données permettant de trouver les dates optimales pour la modélisation, un total de 336 échantillons de sol ont été prélevés à 7 dates d'échantillonnage différentes et traités pour des analyses biochimiques et de biologie moléculaire. De même, l'ADN génomique a été extrait et des librairies de séquençage d'amplicons ciblant le gène de l'ARNr 16 et la région ITS1 ont été préparées, comme décrit précédemment dans l'objectif 1. Des librairies qPCR ont également été préparées pour quantifier les copies d'amplicons de gènes cibles associés au cycle de l'azote et analysées comme décrit dans l'objectif 1 (Chapitre détaillé 1 & Chapitre 2). Les séquences d'amplicons ont été traitées, filtrées et contrôlées en

qualité selon le pipeline bioinformatique reproduit en interne. Le regroupement *de novo* des séquences d'amplicons a été effectué sur la base de séquences représentatives qui ont été annotées pour préparer le tableau des OTUs (unités taxonomiques opérationnelles) consensuelles. La raréfaction et l'analyse de l'arbre phylogénétique des OTUs ont été réalisées en suivant un pipeline Galaxy modifié en interne. Ensuite, diverses analyses de données en aval liées à la diversité alpha et bêta et à l'abondance relative du microbiome du sol ont été réalisées par sous-échantillonnage aléatoire des données OTUs, suivi de diverses analyses statistiques. Le profilage physiologique du microbiome du sol au niveau de la communauté, basé sur l'utilisation des sources de carbone au niveau de la communauté, a été réalisé à l'aide d'un test colorimétrique Biolog. Certains paramètres clés de la qualité du grain de blé et de la farine, notamment le gluten, les protéines, la durée maximale de la farine et le couple enregistré, ont été mesurés en collaboration avec la société de mouture du blé. Des analyses statistiques exploratoires, descriptives et multivariées ont été réalisées, notamment sur la distribution et la variation de la réponse aux différents traitements à différentes dates d'échantillonnage des données microbiennes. Des tests de corrélation non paramétriques ont été réalisés pour les sept dates d'échantillonnage afin d'identifier des indices microbiens robustes associés à la qualité du blé à une échelle temporelle. Initialement, l'objectif principal était de modéliser la qualité du blé à l'aide des données microbiennes recueillies à partir de sept dates d'échantillonnage distinctes, afin de trouver la meilleure date d'échantillonnage ayant un potentiel prédictif élevé. Comme les deux génotypes de blé ont montré des compositions microbiologiques différentes dans l'analyse multivariée, ils ont été modélisés séparément. Après le nettoyage et le traitement des données, la dimensionnalité des données microbiennes a été réduite par un processus d'orthogonalisation. Les composantes principales orthogonales expliquant la plus grande variance des OTU microbiennes et de l'utilisation du carbone ont été traitées comme des caractéristiques du modèle. Combinés aux cinq composantes principales, certains autres traits microbiens tels que l'abondance des gènes liés aux processus microbiens (par exemple, le cycle de l'azote, le rapport total champignons/bactéries) et les indices de diversité alpha ont été normalisés et utilisés comme principales entrées du modèle pour la prédiction de la qualité des grains. Une approche basée sur la régression pénalisée (opérateur de sélection et de rétrécissement le moins absolu LASSO) a été appliquée à la sélection des modèles pour résoudre les problèmes de surajustement et de multicollinéarité de l'ensemble de données. Les scores de pénalité minimum ont été estimés en établissant des seuils de validation croisée pour sélectionner les prédicteurs en fonction de la taille des coefficients de régression lorsque toutes les variables d'entrée convergent vers zéro ou presque. Les résultats prédits à partir des modèles de régression lasso ont été calculés et la linéarité des modèles a été évaluée à l'aide de différents paramètres de modèle, notamment l'erreur quadratique moyenne, le rapport de variance totale (par exemple, R^2) et les critères statistiques (par exemple, AIC, BIC).

Principalement, en accord avec les stades de croissance du blé, les dates d'échantillonnage ont affecté de manière significative tous les indices microbiens, y compris l'utilisation du carbone microbien, la diversité microbienne alpha et bêta, le rapport C: B et l'abondance des gènes liés au cycle de l'azote. La résolution de la corrélation entre la qualité du grain et l'utilisation du carbone microbien a été très fluctuante en fonction de la date d'échantillonnage. La corrélation la plus négative entre l'indice de qualité du grain et l'utilisation du carbone microbien a été observée dans les génotypes tolérants à la sécheresse. De même, le changement dans le schéma de corrélation entre les qualités de blé et les gènes liés au cycle N_2, par exemple AOA, AOB, a été observé à différentes dates d'échantillonnage pour les deux génotypes de blé. Des corrélations significatives ont été observées entre la diversité/richesse microbienne et les indices de qualité du grain dans les génotypes DS.

La meilleure précision prédictive des modèles Lasso pour les génotypes tolérants à la sécheresse a été obtenue à une date précoce d'échantillonnage du sol, qui correspond approximativement au stade de la germination ou du tallage du blé. Toutefois, les prédicteurs microbiens ont eu peu de pouvoir pour expliquer la teneur en gluten et en protéines du blé, en particulier pour les dates tardives d'échantillonnage du sol. Les modèles prédictifs basés sur le lasso pour les génotypes sensibles à la sécheresse ont montré une performance relativement faible par rapport aux génotypes tolérants à la sécheresse. De même, pour les génotypes tolérants à la sécheresse, la précision de la prédiction des paramètres de qualité du grain était plus élevée aux premières dates d'échantillonnage, principalement en mai et juin. Il était également difficile de prédire les qualités des grains avec les indicateurs microbiens obtenus à certaines dates d'échantillonnage pour les deux génotypes de blé.

Les meilleurs modèles obtenus en début de saison ont sélectionné certaines caractéristiques microbiennes, notamment les composantes principales des indicateurs microbiens (par exemple, les OTU et l'utilisation du carbone microbien). Certaines OTUs microbiennes abondantes dans l'ACP caractéristique ont été systématiquement sélectionnées dans la plupart des modèles pour toutes les dates d'échantillonnage du sol. Par exemple, les OTU bactériennes et archéennes de *Nitrosphaera* et les OTU fongiques du *Mortierella* ont été genres indicatifs les plus abondants dans les modèles les plus optimaux. En plus de ces taxons, les ACP sélectionnées par le meilleur modèle pour les génotypes tolérants et sensibles à la sécheresse ont retenues les OTU bactériennes et archéennes de *Rhodoplanes, Solirubrobacter, Gaiella, Bradyrhizobium, Terrimicrobium, Hyphomicrobium* et les OTU fongiques de *Mortierella, Ganoderma, Gliomastix, Pezizella, Tetracladium.* Les principales composantes de l'utilisation du carbone microbien ont été associées positivement ou négativement à la qualité du grain dans les meilleurs modèles Lasso pour les deux génotypes de blé. Différents indices de diversité des OTUs bactériennes, archéennes et fongiques ont été sélectionnés dans les meilleurs modèles de prédiction de la qualité des cultures. Les modèles prédictifs

en début de saison ont sélectionné les gènes de la communauté dénitrifiante qui ont été négativement associés au gluten et à la teneur en protéines des génotypes tolérants à la sécheresse. En revanche, le gène *amoA* a été négativement associé à la teneur en gluten des génotypes sensibles à la sécheresse.

Les meilleurs modèles pour la qualité du grain de blé ont été obtenus à des dates d'échantillonnage au début de la saison de croissance du blé, soutenant pleinement notre deuxième hypothèse. Effectivement la plus forte robustesse a été constatées dans les modèles basés sur les échantillonnages en mai et juin. Les dates les plus optimales générant les meilleurs modèles ont correspondu à peu près au stade de semis ou de tallage du blé: une période peut être critique pour l'absorption ultérieure des nutriments par la plante. Certains paramètres sélectionnés et impliqués dans les processus microbiens peuvent avoir des liens mécanistes avec les qualités du grain de blé. Par exemple, l'abondance relative des OTU appartenant au taxon d'archaea oxydant l'ammoniac *Nitrososphaera* était également fortement corrélée avec bon nombre des principaux composants sélectionnés dans les modèles, et l'abondance des gènes *amoA* archéaux et bactériens était souvent négativement corrélée à la qualité paramètres. Ces résultats suggèrent en outre qu'une forte abondance d'oxydants et de dénitrifiants à base d'ammoniac réduit la qualité du grain de blé en raison d'un besoin énergétique accru pour l'absorption et l'utilisation de l'azote ou par les pertes d'azote, comme indiqué aux (chapitres 2 et 3). Tous les paramètres microbiens n'ont pas été associés de manière causale à la qualité des récoltes. En fait, ils pourraient covarier avec d'autres facteurs non mesurés.

L’abondance des archées oxydant l'ammoniac (AOA) a été sélectionnée dans le modèle et a été négativement corrélée avec la qualité du grain. Même les gènes dénitrifiants *nirK* et *nosZ* sélectionnés dans le modèle étaient négativement corrélés avec la qualité du grain. Ces résultats suggèrent qu'une abondance élevée de gènes liés à la nitrification ou à la dénitrification peut réduire la qualité des cultures en raison d'une absorption ou d'une utilisation inefficace de l'azote par les plantes ou d'une perte accrue d'azote par lessivage ou eutrophisation. Tel discuté précédemment la qualité du grain est liée à sa teneur en protéines. L'absorption d'ammoniac par les plantes est efficace pour les plantes, car cette forme d'azote peut être directement transformée par les plantes en acides aminés, alors que le nitrate doit être reconverti en ammoniac. L'absorption de nitrate par les plantes nécessite également plus d'énergie que l'ammoniac. De plus, le nitrate est aussi le substrat potentiel de la dénitrification qui conduit à des émissions de gaz N_2O. L’activité de ces guildes fonctionnelles microbiennes peut être manipulée ou inhibée à l'aide d'inhibiteurs de nitrification chimiques, naturels ou synthétiques pour contrôler la disponibilité de l'azote et améliorer ainsi la qualité du grain. Nos efforts de modélisation à base microbienne ont permis de comprendre que la diversité microbienne est d'une importance capitale pour l'évaluation de la qualité des grains.

Malgré que notre approche de modélisation était principalement axée sur le blé, elle pourrait servir de base à l'exploration du pouvoir prédictif de paramètres microbiens similaires dans d'autres cultures. Les résultats obtenus à partir de la deuxième approche de modélisation ont également montré que la complexité microbienne augmente et que la précision prédictive diminue avec le temps, ce qui indique que les déterminants de la diversité et de la fonction du microbiome du sol peuvent être influencés par d'autres facteurs environnementaux internes ou externes au cours de la croissance du blé. Ces résultats, en accord avec plusieurs travaux, suggèrent que l'impact de la gestion agricole sur les indicateurs microbiens peut être plus clairement observé au début de la saison de croissance des cultures. Nos efforts de modélisation peuvent inspirer le développement d'un outil qui peut être utilisé tôt dans la saison pour guider correctement les stratégies de gestion agricole. Les indicateurs microbiens initiaux mesurés tôt dans la saison et liés aux caractéristiques des cultures, peuvent influencer ou interférer avec une activité microbienne spécifique dans un agroécosystème donné. La diversité et l'abondance microbienne peuvent être influencées par les conditions environnementales anciennes ou actuelles. Par conséquent, dans le cas d'un environnement homogène, certains indicateurs microbiens caractérisant le sol au moment du semis peuvent capturer une portion du destin de l'azote à l'échelle d'une saison complète. Les microorganismes du sol agissent comme des prédicteurs potentiels des conditions physico-chimiques passées et présentes du sol. L'utilisation de ces prédicteurs est donc appropriée pour expliquer divers processus de l'écosystème. Dans nos études, les descripteurs généraux de la communauté, comme la diversité alpha ou la composante principale, ont souvent été sélectionnés dans les modèles comme étant les meilleurs prédicteurs de la qualité des grains. Il a également été observé que la diversité alpha et les valeurs propres des composantes principales ont été des prédicteurs adéquats, par conséquent, la modélisation incluant quelques paramètres microbiens tels que la diversité alpha, la diversité bêta et les patrons d'utilisation des sources de carbone, peut être suffisante pour prédire la qualité du grain de blé. En particulier, l'approche de modélisation basée sur le lasso était la meilleure parmi les modélisations basées sur la régression linéaire. Effectivement, cette approche a produit des modèles plus parcimonieux, hautement interprétables et a fourni les résultats les plus fiables pour prédire la qualité du grain avec une précision adéquate.

En conclusion, notre premier objectif a clairement illustré que les modèles prédictifs significatifs peuvent être paramétrés en utilisant des indicateurs microbiens mesurés tôt dans la saison de croissance, sur un transect de plus de 500 km. Notre deuxième objectif s'appuyant sur l'approche de modélisation temporelle a confirmé que l'échantillonnage au début de la saison de croissance peut être plus favorable pour l'estimation future de la qualité des cultures. Prises ensemble, ces deux études suggèrent qu'en collectant des échantillons de sol uniquement pendant la saison de croissance du blé, la modélisation basée sur les microbes a une meilleure précision, faisabilité et efficacité. En outre, conformément aux travaux précédents, cette approche de modélisation basée sur les microbes a sélectionné des indicateurs microbiens

associés à d'importants processus d'azote dans le sol, tels que les oxydations d'ammoniac, fournissant un signal potentiel d'une guilde microbienne fonctionnelle qui joue un rôle décisif dans la qualité du grain de blé. Ainsi, cette recherche pose une base solide pour les efforts futurs visant à prédire et à optimiser le rendement et la qualité des cultures. Ces travaux ouvrent donc de nouvelles voies vers des solutions micro biocentriques pour résoudre les problèmes critiques auxquels l'agriculture est confrontée.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

1 INTRODUCTION

1.1 Wheat cultivation and fertilization

Wheat (*Triticum spp.*) is one of the most widely produced crops in the world, with an estimated 200 million ha. of land dedicated to wheat cultivation every year. Wheat grain is among the most consumed cereals, supplying 19% of the calories consumed by humans (Aksoy and Beghin 2004). Forty percent of global wheat production is used as a dietary supplement in the poultry and livestock industries. Wheat is also a key ingredient for producing bread, pasta, and other baked goods. Wheat grain is highly rich in carbohydrates, proteins, and minerals compared to other cereals such as rice, rye, and millet (Chung, Pomeranz and Finney 1978). Canada is the world's sixth-largest producer and one of the largest exporters of wheat, producing an average of over 25 million tons and exporting around 15 million tons, annually. There are several classes of wheat varieties that have developed through breeding to adapt to different factors such as grain hardness, grain color, seeding season, and farming location and environmental conditions (e.g., rainfall events, drought etc.). About half of all Canadian wheat is grown in Saskatchewan, followed by Alberta, Manitoba, and Quebec (Newlands *et al.* 2014). In Canada, two types of wheat varieties, known as winter and spring wheat, are widely grown in the eastern region. Winter wheat is mainly sown in autumn and is frost resistant, while spring wheat is usually sown in spring. Spring wheat is associated with lower yields and quality compared to winter wheat. Modern wheat breeding aims to produce crops that have traits such as high yield, disease resistance, and high protein content. The effects of climate change present challenges for global wheat production and the selection of high-yielding cultivars that mitigate nitrogen loss and biotic and abiotic stress. Organized agricultural practices and microbiome-assisted fertilizer management may offer future solutions to restore soil fertility and produce high yield and high-quality crops.

Nitrogen is crucial for wheat production as it plays a pivotal role in grain protein synthesis. Studies show that grain nitrogen concentration largely depends on plant nitrogen uptake efficiency and the availability of soil nitrogen (Zörb, Ludewig and Hawkesford 2018). Ensuring adequate nitrogen supply during wheat growth is important for optimizing wheat yield and quality. Nitrogen also influences both protein quantity and protein quality. The gluten protein in wheat is responsible for bread elasticity and

texture. When baking dough, yeast-mediated fermentation produces carbon dioxide and other metabolites, which causes the dough to rise and give the bread a light and fluffy texture. Continued research on wheat helps the baking industry to produce top-quality, nutritious bread while addressing the increased demand for gluten-free wheat products. This presents an exciting opportunity to explore ways to lower gluten content and increase other proteins in wheat grain, while also improving the grain overall (Goel *et al*. 2021). According to the guide of Canada grain council to wheat management, scientists from Saskatchewan have suggested that it is possible to increase the grain protein content to a maximum of 16% while maintaining or increasing yield (open.alberta.ca/publications/wheat-nutrition-and-fertilizer-requirements, Ames *et al*. 2003; Hucl *et al.* 2022). However, they suggest that when the protein content exceeds 16%, the yield is reduced (open.alberta.ca/publications/wheat-nutrition-and-fertilizer-requirements). For instance, research conducted in a farm in Lethbridge, Alberta reported that nitrogen fertilization may have increased some grain protein content such as glutamic acid, proline, methionine, cysteine, phenylalanine, and tyrosine, but that would lead to a lower yield (open.alberta.ca/publications).

Wheat has three main phases of growth that require high levels of nitrogen: sowing, tillering, and kernel formation. After sowing, wheat kernels require 6 mg of protein to maintain germination and growth (Zörb, Ludewig and Hawkesford 2018). Insufficient nitrogen (N) in the soil means less protein is produced and wheat growth is negatively affected. The second most N-intensive phase is tillering, during which N concentrations determine the tiller formation of the plant (Zörb, Ludewig and Hawkesford 2018). Some wheat genotypes have the capacity to store N and use it during the grain-filling stages when there is low N availability (Zörb, Ludewig and Hawkesford 2018). However, in some cases, N uptake efficiency of the wheat genotype remains poor in highly N-fertilized soil (Bogard *et al.* 2010). Kernel formation is the third most N-intensive stage, during which grain protein synthesis occurs (Pechanek *et al.* 1997; Barneix 2007).

Modern wheat cultivars are expected to have high grain protein concentrations, requiring optimal amino acid synthesis in the plant tissue to maintain continuous transport of the protein synthesis machinery during grain development. Due to the high N demand during wheat cultivation, plants are fertilized with high concentrations of N (up to 150 kg N ha^{1}) in the later stages of wheat growth (Zörb, Ludewig and Hawkesford 2018). More than half the commercially produced N is applied indiscriminately during wheat production without regard to soil N status, totaling to more than an estimated 180 Mt/yr (Hawkesford 2014). Unfortunately, contrary to what one might expect, there is a tendency to over-fertilize without considering the microbial factors involved in soil N cycling. Schulz *et al.* (2015) found that the yield and protein concentration obtained with a single application of N fertilizer during the tillering stage of wheat growth

was equal to the yield obtained with multiple N applications during other stages. This may be due to the gradual decomposition of a large stock of soil organic N over a long period of time, in which case, additional N may not be required in the soil. Over fertilization is one of the main causes of N loss during wheat production (Chen *et al.* 2019). This reduced form of available nitrogen in the soil can affect the genetic potential of wheat plants to process the amounts of N required during kernel development (Yu *et al*. 2017). When similar wheat genotypes are grown in the same field for decades with different level of N fertilization, the resulting soil microbiome composition significantly affects crop yield and quality (Nelson *et al.* 2011; Yergeau, Quiza and Tremblay 2020).

1.1.1 Context

It is important to understand what factors influence grain and crop quality. Crops that are of poor-quality lead to financial losses for farmers. There are several challenges in trying to predict grain quality, as it is not necessarily linked to the amount of fertilizer applied (Yergeau, Quiza and Tremblay 2020). In my thesis, I aimed to address this issue by modeling wheat yield and grain quality using microbial indicators.

Nitrogen fertilizers are produced through the Haber-Bosch reaction, a process that comes with high energetic costs. This reaction takes atmospheric N and transforms it into ammonium (NH_4^+). This man-made N results in twice as much N fixation than all of the natural biological nitrogen fixation activities in the soil (Fisher and Newton 2004). The production of synthetic N fertilizers using the Haber-Bosch process is cost-effective but is totally dependent on fossil fuels. The market price for synthetic fertilizers is quite low due to the huge subsidies provided by state governments (Zörb, Ludewig and Hawkesford 2018). The large supply and low prices of N fertilizers have encouraged farmers to overuse these fertilizers, resulting in the loss of soil N equilibrium and microbial diversity (Zörb, Ludewig and Hawkesford 2018). The majority of the applied nitrogen fertilizer is unused by the crops, with more than half of the N being lost through volatilization of NH_3, leaching of NO_3^- and emissions of NO, N_2O and N_2 (Butterbach-Bahl *et al.* 2013). Nitrate leaching (NO_3^-) or N_2O gas production (N_2, N_2O) reduces the available nitrogen stores in the soil. The N_2O gas is also highly efficient at capturing heat along with CO_2, contributing substantially to global warming. Nitrogen loss through leaching significantly reduces microbial diversity and contributes to eutrophication. The environmental issues related to continuous N fertilization have become a concern throughout the world, especially in Europe and North America.

Nitrogen cycling is a crucial part of maintaining the soil nitrogen flux, where microbes play a vital role in nitrogen transformation. For instance, microbes can be ammonia-oxidizers, nitrite-oxidizers and comammox (complete ammonium oxidation), which convert ammonia into nitrates. The protonated form of ammonia, called ammonium, is positively charged, and usually binds with the generally negatively charged soil particles, whereas the negatively charged nitrate does not bind with the soil particles and is more mobile and prone to leaching. In anaerobic conditions, nitrate is also used as an electron acceptor by denitrifiers, leading to its transformation into NO, N_2O or N_2 and the release of N into the atmosphere. As the competition for nitrogen assimilation between the plant and microbes is significant, nitrogen availability in the soil could become limited. Estimating the N reduction rate through physicochemical analyses of $NH4^+$-N or $NO3^-$-N is not enough to assess true nitrogen levels in the soil. Because microbes directly and indirectly contribute to nitrogen accumulation, processing, and distribution either through microbe–microbe or plant microbe–interactions, microbes should be directly considered during soil nitrogen flux assessment. Assessing the microbial ecosystem of the soil and its function and diversity index could give us a better understanding of the nitrogen transformation equilibrium in the soil.

Microbial-based transformations of soil nutrients are also important for plant growth and nutrition. The different forms of nitrogen that are available in the soil could increase N uptake efficiency in the plants, which could later enhance plant protein metabolism and grain protein content. A continuous supply of plant-usable nitrogen that is available due to the actions of microbes can help upregulate the physiological or mechanical functions of the plant. These microbes prime the plant to become more efficient at dealing with biotic stresses and might enhance overall plant productivity. Indeed, NH_4^+ is energetically more advantageous for plant growth as it can be passively absorbed and directly used to make amino acids. Nitrate, on the other hand, needs to be actively taken up by the plant and transformed back to ammonium for amino acid production. Microbial activity, abundance, diversity, and community structure will therefore greatly influence the fate of the applied fertilizers and the efficiency of their use by crops. However, microbes are rarely considered in the decision-making processes about fertilization. The creation of fertilization guides and regulations based on information about microbial ecosystem functions could decrease excessive N input and the associated environmental costs. These measures, along with new breeding strategies such as developing wheat varieties with high N uptake efficiency, restoring soil ecosystem function or more sustainable agricultural practices, could vastly improve overall wheat production.

My goal has been to develop a microbial-based predictive model for wheat yield and grain quality, specifically for crops grown in Quebec. Wheat is an interesting model crop, as the yields generally respond very well to N fertilization. Furthermore, I chose to study Quebec wheat crops and soil because of the protein content in bread wheat is related to the availability of nitrogen, especially NH_4^+ sources, during plant N metabolism at the grain-filling stage.

1.2 Microbial dynamics in the agroecosystem

Agroecosystems are complex and dynamic. They vary at local and regional scales and are influenced by regional, geographic, and global agricultural practices and environments. The causal relationships between microbial variables and ecosystem processes are not always linear. They rather appear as complex patterns caused by diverse and exogenous factors, positive and negative feedback loops, time periods and other non-linear dynamics (Sengupta *et al.* 2021). Furthermore, soil microbiome structures are closely tied to a hierarchical series of environmental factors that are highly susceptible to climate change (Yuan *et al.* 2021).

The community assembly process within the plant holobiont is influenced by many factors, such as soil type (Engelbrektson *et al.* 2012; Bulgarelli *et al.* 2015), plant compartments (Bodenhausen, Horton and Bergelson 2013; Edwards *et al.* 2015), host genotypes (Engelbrektson *et al.* 2012; Bulgarelli *et al.* 2015; Cardinale *et al.* 2015), plant immune systems (Glavina *et al.* 2015), plant traits, developmental stages (Donn *et al.* 2015) and the climate and season (Dombrowski *et al.* 2017). Plant root-associated soil microbiomes play an important role in limiting the growth of many pathogens (Shawy and Burns 2009; Pignataro *et al.* 2012; Qian *et al.* 2015), promoting plant growth, helping with nutrient acquisition (Berendsen, Pieterse and Bakker 2012), reducing abiotic stresses (Bodenhausen *et al.* 2014) and priming plant-induced systemic resistance (Parnell *et al.* 2016). The rhizosphere around the plant roots serves as a potential hub for plant–microbe interactions. Here, the microbial diversity, activity, and community composition vary due to rhizodeposition within the rhizosphere and the adjacent bulk soil (Kourtev, Ehrenfeld and Häggblom 2003). Rhizodeposits can fix up to 40% of the net carbon in rhizosphere soil and contain a variety of low- and high-weight molecules (amino acids, organic acids, secondary metabolites, etc.), mucilage, and sloughed cells (Parnell *et al.* 2016). The pattern of rhizodeposition plays a decisive role in the distribution of microbial communities across different plant species and cropping systems (Baudoin, Benizri and Guckert 2003). Significant changes in rhizosphere communities between the seedling and

grain-filling stages of crop growth have been shown to be host-specific and can determine the successful colonization of the microbiome in the wheat rhizosphere (Donn *et al.* 2015).

Various environmental factors have been shown to influence the composition of the plant microbiome and plant–microbiome interactions (Blanchet *et al.* 2015), thereby playing a major role in altering plant physiological processes. For example, during the periods of drought, higher fungi: bacteria ratios (Azarbad *et al.* 2018; Meisner *et al.* 2018) have been shown to correlate with the plants' ability to adapt to water-limited conditions (Evans and Wallenstein 2014). Fertilization also has a significant effect on microbial community structure, affecting many phyla either positively or negatively. *Acidobacteria, Chloroflexi, Fibrobacteres, Nitrospirae, Planctomycetes,* and *Verrucomicrobia* were all positively correlated with high pH and carbon to nitrogen (C: N) ratios, while *Actinobacteria* and *Proteobacteria* showed negative correlations with these parameters (Kavamura *et al.* 2018). It has also been reported that soils treated with inorganic N fertilizers were found to exhibit lower levels of bacterial taxonomic diversity (Kavamura *et al.* 2018).

1.2.1 Temporal variation

According to ecological theory, the succession of microbial communities in specific ecosystems is largely driven by the diversification and integration of the community structure over time. Temporal dynamics influence community assembly processes that determine changes in the stability of a microbial community and biodiversity (P. White *et al*. 2006). Temporal dynamics also influence the microbial response to ecological disturbances (Fraterrigo and Rusak 2008). The soil microbiome shows hypervariability in various ecosystem processes across time and space (Hannula *et al.* 2020). Changes in microbial diversity over various temporal scales could be an indicator used for monitoring soil microbial processes over time and microbial responses to different environmental conditions. Microbial community dynamics might stabilize over long time periods through homogenous selection processes. Researchers have found that in a microbial-driven ecosystem within a stable community structure, changes in trait-specific characteristics and plant root exudation levels can be easily observed within a very short period of approximately a few days to a week (Bais *et al.* 2006; Broeckling *et al.* 2008). Plants play an important role in shaping the composition of the soil microbiome where there are multiple interactions between different microbial communities. This can influence the growth of other plant species in the same soil (Hannula *et al.* 2019; Morriën *et al.* 2017). Thus, temporal changes in the plant microbiome at the ecosystem level may affect plant species-specific co-occurrence and diversity (Putten *et al.* 2016). The temporal dynamics of the

soil microbiome can also influence bacterial and fungal community structures. Bacteria have comparatively shorter generation times than fungi, meaning they likely respond to environmental changes faster than fungal communities (Sun *et al.* 2017; Allison and Martiny 2008; Rousk and Bååth 2007). Recently, it has been reported that the composition of soil fungal communities undergoes rapid changes due to seasonal effects that are thought to be the major source of temporal variation. However, the other ecological drivers and timescales associated with this process is not fully understood (Averill *et al.* 2019).

Seasonality can shape soil microbial community structures and diversity richness as plant-derived resources change (Wardle *et al.* 2004). The shift in microbial community structure over time is highly correlated with N turnover and plant species richness (Björk *et al.* 2007). Temporal shifts in microbial diversity during seasonal changes might affect the overall fungal–bacterial ratio and plant growth by altering soil total microbial biomass and C:N ratios (Chen and Gao 2022). Plant phenology is therefore also affected. A report on the effects of the soil microbiome on plant root systems in belowground environments describes a correlation between the changes in plant phenology and differential gene expression (Bouché *et al.* 2016). This suggests that the soil microbial assembly processes that are driven by abiotic selection increase niche differentiation and competition for nutrients among microbial assemblages, possibly reinforcing plant phenological variation(Trivedi *et al.* 2020). The soil microbiome not only contributes to plant phenological changes across environmental and geochemical gradients but also affects various biological factors that mediate plant–microbe interactions, such as the timing of leaf emergence and related ecosystem processes (Van der Putten 2012; Classen *et al.* 2015). Another interesting example of microbial temporal dynamics is referred to as legacy effects. These are the inherited feedback of the plant-soil microbiome interaction, which persists even after migration of the microbial community to a new location and can last from months to decades in agricultural soils and potentially longer, depending on agricultural practices, other environmental factors (e.g., drying-rewetting cycles), and historical dependence on the abundance of certain taxa (Evans and Wallenstein 2012; Averill, Waring and Hawkes 2016; Giauque and Hawkes 2016).

1.2.2 Spatial variation

The distribution of microbial communities is potentially linked to the large-scale biogeography of soils at continental and regional scales. Several studies on soil microbes have demonstrated that biogeographic patterns of soil microbes depend exclusively on soil properties, environmental conditions (e.g., climate, topography) and land type (Chemidlin *et al.* 2014). The factors that influence microbial distribution vary over different spatial scales and ecosystems. A report by Dequiedt *et al.* (2009) demonstrated that in North and South-East France, microbial spatial distribution is more related to soil properties and land type than to geography and climate. Another report on the biogeography of microbes have noted that rainfall patterns and soil properties play essential roles in microbial community establishment and distribution across the Mongolian Plateau in China (Chen *et al.* 2015). Land use can also have a major impact on spatial distribution. In some cases, traits that have been conserved over evolutionarily time scales are disrupted by the way the land is used. Different agricultural practices also affect the structure and composition of overall microbial communities, which vary across agricultural fields and spatial scales.

Agricultural management is the driving force behind changes in soil processes and function (Techen *et al.* 2020). Different types of agricultural field management result in variations in soil microbial community composition and function over spatial scale. Fields are usually managed by conducting various mechanical operations, such as tilling and sowing the land to prepare the soil for plant growth. These mechanical activities directly affect the soil microbiome by modifying the structure of the soil. For example, conventional tilling operations reduce soil fertility through wetting-drying and freeze-thaw cycles at the soil surface, promoting soil erosion and disrupting the soil–pore complex. This causes high levels of decomposition of soil organic matter that deplete the soil organic carbon pool (Six *et al.* 2002). Changes in soil structure through conventional management practices also affect the physical habitat that soil microbes live in. Reductions in fungal biomass can occur through the destruction of hyphal networks. Agricultural soil compactness affects community structure by increasing the activity of anaerobic and saprophytic microbial communities, resulting in reduced richness of aerobic microbial taxa and crop-associated microbiomes in the agroecosystem (Longepierre *et al.* 2021). Increased production of CH_4 (Sitaula *et al.* 2000) and N_2O (Sitaula *et al.* 2000b) can promote methanogenic and denitrifier communities while limiting the growth of methanotrophs and ammonia oxidizers. Other agricultural practices such as bio-inorganic fertilizer-based farming practices, crop rotation, pesticide application and intercropping are also important

variables for maintaining the spatial structure of microbial communities within the agroecosystem (Xu *et al.* 2020).

Another part of spatial variation of soil microbial communities through changing their composition is directly related to fertilization management, which greatly affects microbe-driven soil processes and biological activity. It has been reported that long-term intensive fertilization can greatly affect soil fungal communities (Hartman *et al.* 2018) and microbial total carbon and nitrogen biomass (Li *et al.* 2018). Decisions in using different fertilizer in agriculture, therefore, greatly influence soil microbiome function at regional and continental scales. For example, mineral fertilizers can increase soil nutrient accessibility but limit microbial respiration by reducing organic carbon sources (Janssens *et al.* 2010). Consequently, high rates of organic matter decomposition provide the necessary binding agents for the active compaction of soil aggregates and thus help increase soil fertility by also increasing fungal hyphal networks and the production of essential microbial metabolites (Řezáčová *et al.* 2021). Several studies have reported that organic fertilizers expand the potential functional guilds involved in carbon, nitrogen, and other cycles (Enebe and Babalola 2021). At the same time, repeated application of the same type of organic fertilizer can promote heterotrophic respiration and nitrogen turnover, which can increase CO_2 and N_2O emissions (Skinner *et al.* 2019). Further, several studies revealed that microbiomes associated with bulk soil are much more sensitive to fertilizer exposure than plant indigenous microbiomes (Sun *et al.* 2021; Xiong *et al.* 2021a, 2021b). Organic farm management has also been reported to impact crop microbiome and crop quality to varying degrees. These practices can also increase microbial alpha diversity and dispersal of potential bacterial communities (e.g., *Bradyrhizobium* and *Bacillus*) (Vannette and Fukami *2017;* Khoiri *et al.* 2021; Wittwer *et al.* 2021).

Microbial spatial variations can be observed in the soil through changes in microbial habitats, including niche differentiation at the macro or micro scale. Microbial colonization and succession are facilitated through the formation mechanisms of the soil matrix, such as soil particle aggregation, leading to increased soil oxygen permeability and nutrient flow. The structural stability of the soil aggregate is essential for determining microbial diversity, activity, and niche differentiation. One study confirmed that most microbial habitats are strongly co-occurred with microaggregates by demonstrating that a large proportion of soil bacteria tended to colonize inside soil macroaggregates (Ranjard *et al.* 2000). In contrast, a small percentage of microbes usually colonize the surface of macroaggregates in large pores (Mummey *et al.* 2006). Although microbial habitats in soil aggregates are spatially distanced from one other, their phylogenetic network can be re-established to some extent during wet cycling, when microbes are able to

communicate with each other through soluble nutrients, metabolites, and genetic transformation (Wilpiszeski *et al.* 2019). The spatial separation that occurs within soil microaggregates restricts microbial dispersion and promotes the establishment of microbial communities, through independent functional and evolutionary processes (Rillig, Muller and Lehmann 2017). The strong correlation between soil aggregates and soil organic matter plays a decisive role in the selection process by abiotic stress for a microbial habitat. For example, poor soil aggregation that is mediated by labile carbon compounds could facilitate fast-growing microbial copiotrophs in the carbon-rich environment (Trivedi *et al.* 2017). In contrast, consistent and stable soil aggregation (microaggregates) promotes the growth of oligotrophs, which have the metabolic machinery necessary to degrade more complex compounds (Trivedi *et al.* 2017). Soil porosity can also shape aerobic and anaerobic microbial niches over different spatial scales. The limited oxygen that is inherent to soil pores help anaerobic microbes to improve their functional activity, as observed among dentrifier communities (Kong *et al.* 2010). This suggests that tracking the porosity of soil macroaggregates, which varies between agricultural fields, can be an important way to monitor microbial-driven carbon sequestration or N_2O emission levels over multiple spatial scales.

Abiotic and biotic factors influence the structure and diversity of microbial communities, resulting in spatial variations in microbial functions within agroecosystems. These variations may explain differences in crop yield and quality. Soil nutrient abundance is related to the spatial patterns of microbial diversity, which are highly affected by agricultural practices. In heterogenous environmental conditions (different biotic and abiotic factors), taxa-specific selection and abundance based on multitrophic interactions causes microbial community to maintain their structure. Heterogeneity in microbial diversity is especially pronounced when the hierarchical series of interactions that occur between variables fluctuate at different spatial scales. Microbe-related ecosystem functions are composed of microbial inter- and intraspecific interactions and soil physicochemical properties such as pH, C:N ratio, and N availability. These variables help determine the microbial communities that shape soil microbiome structures at spatial scales (Asad *et al.* 2021b). It has been reported that variations in soil physical, chemical, and biological properties over time and space are strongly correlated with microbial diversity (Franklin and Mills 2003). For example, autocorrelated spatial patterns among microbial communities at the landscape scale vary with pairwise distances > 700km (Bru *et al.* 2011). However, microbial communities in other soil ecosystems such as arctic soils are influenced by spatial factors on a larger scale (Shi *et al.* 2015). So far, there are very few studies on the robustness of microbial spatial parameters on N processes that directly or indirectly affect crop species and yield. Soil properties such as pH, soil organic C, and C:N ratios are important factors affecting microbial enzyme-driven geochemical processes in agricultural fields with different soil types and

regional climates. Wheat cultivated by different agricultural systems across a 500-km transect in Quebec showed a wide range of diversity in microbial community composition, resulting in variations in yield and quality metrics (Asad *et al.* 2021a). Spatial variables including biotic and abiotic factors that correlate with different microbial activities can be useful tools for the classification and estimation of the specific taxa that make up the soil microbiome. However, there is still little information on the extent to which diversity and function influence crop quality and quantity at large spatial scales (Figure 1-1).

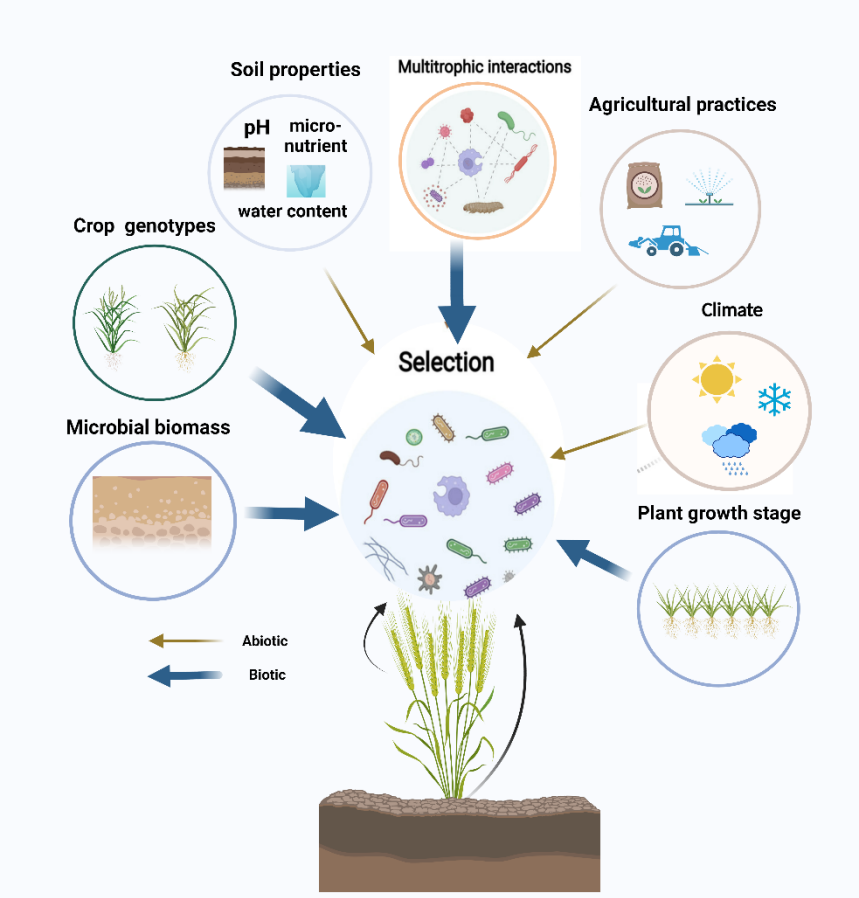


Figure 1-1: Factors influencing the microbial community selection processes in soil and crop microbiomes.

There are four main mechanisms in community assembly: 1) dispersal (community movement to a new location), 2) selection (influenced by biotic and abiotic factors, 3) diversification (genotypic variations), and 4) drift (random birth and death events). The microbiome can be transmitted from seed to crop through vertical and horizontal transfer from the soil, atmosphere, neighboring plants, and interacting insects and animals. The crop microbiome assembly process is largely influenced by factors, such as plant growth stage, genotypes, microbial biomass, climate, soil type and nutrients, and agricultural practices. In addition to deterministic selection, stochastic selection process (e.g., drift) has an essential role in community

assembly. But the contrasting results between stochastic and deterministic processes in agroecosystems vary across time and space in terms of crop growth stage, microbial biomass, and microbial composition. (Xiong and Lu 2022).

1.2.3 Abiotic drivers

Currently, intensive agricultural management is one of the driving forces transforming agroecosystem services. Ecological processes on the soil–plant continuum is largely influenced by agricultural management and fertilization practices (Schmidt *et al.* 2019; French *et al.* 2021). Recent studies have suggested that intensive fertilization could reduce the microbial capacity for processing soil nutrients associated with diazotrophic activity (Fan *et al.* 2019). Furthermore, the structure and function of plant root-associated microbiomes are affected by the regional environment and the different agricultural practices they are subjected to (i.e., conventional, and organic management) (Hartman *et al.* 2018; Asad *et al.* 2021a). As agricultural management is an important abiotic driver that shapes soil microbial community structure, a better understanding of plant–soil microbiome interactions and their effects on crop quality and quantity within the agroecosystem may be key for future microbiome engineering and agricultural production enhancement (Figure 1-1).

Temperature plays an important role in determining the growth of microorganisms in different environmental conditions. Physiological responses of microorganisms to high temperatures trigger various cellular processes that enable microbiomes to adapt to extreme environmental conditions. Advanced sequencing technologies and functional gene quantification experiments have revealed significant changes in microbial community structure and function in response to high temperatures in agricultural fields (Schindlbacher *et al.* 2011; Melillo *et al.* 2017; Romero-Olivares, Allison and Treseder 2017). The level of microbial response varies depending on the environmental factors in different ecosystems (e.g., forest compared with grassland). For example, warming temperatures have been shown to have significant impacts on soil fungal communities in different boreal forest ecosystems, resulting in either stimulation (Clemmensen *et al.* 2013) or suppression (Allison & Treseder, 2008) of fungal biomass and activity. This pattern was observed through differences in soil water content and vegetation history at different sampling sites (Allison & Treseder, 2008). A long term-study from the Harvard Forest Ecological Research Station revealed that microbial respiration and mechanisms when acclimating to rising temperatures follows four main steps: rapid carbon loss, high tendency to disperse through community reorganization, diversity

richness of oligotrophic communities with high levels of respiration, and high decomposition of organic carbon pools with significant changes in microbial community structure (Melillo *et al.* 2017).

Drought is one of the major indicators of climate change and its consequences have been thoroughly observed in grassland ecosystems. Drought is thought to become a major cause of desertification in future semi-arid and arid regions (Wang *et al.* 2022; Melillo *et al.* 2017; Azarbad *et al.* 2018; Agoussar *et al.* 2021). The increasing tendency for droughts suggests a future decline in microbial functions, which are important for ecosystem sustainability (McHugh *et al.* 2017). When the soil is dry, less water passes through soil pores, resulting in fewer resources, slower decomposition of soil organic carbon (SOC), and higher rates of CO_2 respiration (Schimel 2018). A multiyear field experiment revealed that bacteria are more sensitive to drought than fungi in grassland ecosystems (de Vries *et al.* 2018). This suggests that fungi might play a major role in the maintenance of carbon and nitrogen cycling during periods of drought (Treseder *et al.* 2018).

Another type of physiological stress for arid-soil microorganisms is the sudden dry–rewetting cycle of the soil (Schimel 2018). Although dry soil may decrease microbial activity, exogenous enzymatic activity may continue, leading to the accumulation of bioavailable substrates. During soil rewetting, these functions may suddenly be reversed (Unger *et al.* 2010; Barnard, Osborne and Firestone 2013). High mortality rates for bacteria and fungi during soil rewetting events may be an indicator of the decomposition of dead bacterial cells and bacteria-mediated viral predation (Blazewicz, Schwartz and Firestone 2014). One study performed simulations based on an empirical dataset to determine whether there was high microbial diversity in dry conditions and revealed differentiated niches in the dry soil (Šťovíček *et al.* 2017). It was shown that sudden rewetting may lead to high rates of dispersal and niche-based connectivity and increased activity among anerobic microbes. However, other microbial communities returned to their original state when the soil became dry again (Šťovíček *et al.* 2017). Soil moisture content increases due to flooding or heavy rainfall. Recent studies on climate change have shown that precipitation rates in the northern hemisphere in northern climates could increase while the amount of snowfall decreases, thereby reducing snowpack and increasing the frequency of the freeze-thaw cycle (Sorensen, Templer and Finzi 2016). As soil water content increases, soil pores fill with water, creating anoxic conditions. These conditions can be favorable for methanogenic and denitrifier communities that may in turn lead to higher CH_4 and N_2O emissions. Continuous variation in soil water content due to changes in precipitation level exhibit different patterns of soil ecosystem processes in contrast to crop microbiome responses.

Soil properties such as pH or water content are the most influential on and may limit the rate of CO_2 and CH_4 gas flux, greatly impacting soil–microbe gas exchange and nitrification processes (Ye *et al.* 2012; Levy-Booth, Prescott and Grayston 2014). Nitrate can inhibit or stimulate the methanotrophic or methanogenic microbial community (Liu *et al.* 2011; Kim *et al.* 2015) and NO_3^- can inhibit the process of acetogenic methanogenesis through competition with denitrifiers (Leilei *et al.* 2017). Nevertheless, methane oxidizers (methanotrophs) that use methane as their carbon and energy source and encode methane monooxygenase (*MMO*) genes function as analogues of ammonia monooxygenase (*amoA*) encoded by ammonia-oxidizing microbial communities. Therefore, it has been found that increased activity of methanotrophic microorganisms can compete with ammonium oxidizers and inhibit the oxidation of NH_4^+/NH_3 (Bodelier and Steenbergh 2014). This competition for NH_4^+/NH_3 oxidation in the soil can have a significant impact on the overall N-flux, which can either increase or decrease the efficiency of crop N uptake. Still, the information regarding the links between inorganic nitrogen and CO_2 and CH_4 production are not clearly understood.

1.2.4 Biotic drivers

Microbial co-occurrences and metabolic networks are essential drivers of biotic interactions among microbial communities and are associated with various ecological consequences and perturbations. Microbial networks formed through antagonism, or syntrophic or cross-feed interactions are important deterministic factors for soil microbiome function. Co-occurring functional guilds that process soil nutrients are triggered by the selection pressures from biotic and abiotic factors. For example, arbuscular fungi (AF) can promote plant growth through synergistic interactions with bacteria, including optimal nutrient acquisition and plant pathogen resistance (Artursson, Finlay and Jansson 2006). Bacterial–fungal interactions are important for reducing high-input-based agricultural cropping systems by enhancing soil biological processes that reduce reliance on agrochemicals and help maintain soil fertility and plant health (Rashid *et al.* 2016). As discussed earlier, bacterial attachment to AF hyphae is mutually beneficial in that it helps both organisms successfully colonize plant root surfaces through exchange of nutrients and carbon (Bonfante and Genre 2010). Generally, the interaction between AF and rhizobia occurs during the pre-colonization stage at the mycorrhizosphere site and leads to the establishment of symbiosis. AF have important interactive roles, promoting N_2 fixation by interacting with symbiotic or free living N_2-fixing bacteria (Gianinazzi and Schüepp 1994).

Consumers in the microbial food web are also an important part of the biotic factors that significantly shape microbial community structure and composition. For example, protists are the primary consumers of bacteria and fungi and some small eukaryotes, behaving as pathogens, predators, or pests. Some groups of protist communities can release essential nutrients into the soil that promote plant growth. Some are potential bioresources for improving soil fertility. Many protists have mutualistic or symbiotic interactions with plants, animals, and fungi (de Vargas *et al.* 2015). Soil protists play a key role in nutrient cycling by predating on bacteria and other communities. A large proportion of soil protists are bacteriophages (Clarholm 2002). Protists comprise a biomass with a higher C:N ratio than their bacterial prey and released ammonia into the soil as bacterial waste (Sherr, Sherr and Berman 1983). Protists therefore contribute to the agroecosystem by releasing free nitrogen for plants and other microorganisms. The predative nature of protists greatly impacts population dynamics and the assembly of bacterial communities (Hünninghaus *et al.* 2017). A significant correlation was observed between protist species richness and bacterial diversity, suggesting that protists not only act as predators but also play a role in maintaining bacterial diversity in soil (Saleem *et al.* 2012). Protists also serve other roles in the ecosystem, including as stimulants of important bacterial-mediated secondary metabolite syntheses, as parts of direct or indirect interactions, and as inhibitors to pathogen growth by competition or predation. Negative interactions with plants can disrupt plant secondary metabolite production or inhibit the growth of mutualistic microorganisms.

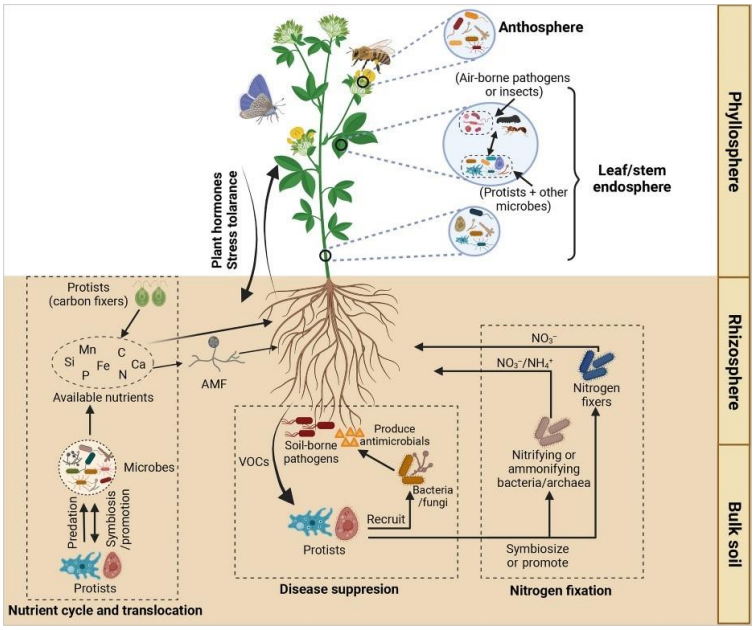


Figure 1-2: Potential roles of protist communities in soil ecosystem processes. Potential roles of protist communities in soil ecosystem processes and plant–microbiome interactions in different plant compartments (Nguyen *et al.* 2022; license free: CCBY4.0).

Nguyen *et al.* (2022) have illustrated the functional potential of protists in soil ecosystems, with a graphical representation (Figure 1-2). The authors discussed the influence of protists on host–microbiome interactions in different plant compartments (e.g., phyllosphere, anthosphere, leaf, stem, root endosphere, rhizosphere, and bulk soil). Protists predate certain groups of phylosphere microbiomes and trigger them to secrete toxic substances to protect plants from pathogens and herbivores. Certain types of protists play a selective role in modulating the function of phytohormone-producing bacteria. It has been reported that bacterivorous amoebas promote bacterial phytohormone production in the plant rhizosphere (Bonkowski and Brandt 2002; Krome *et al.* 2010). Nguyen *et al.* (2022) report that protists could interact with bacteria, archaea, and fungi (especially arbuscular mycorrhizal fungi (AMF) through predatory or symbiotic relationships, affecting nutrient cycling, plant uptake and availability of essential soil nutrients (e.g., nitrogen, phosphorous, carbon, silicon, calcium, magnesium, and iron). Nutrients translocated in the soil are important for the growth and survival of plants and microorganisms in the rhizosphere and bulk soil. Protists can contribute to soil carbon pools by releasing their digestive material after predation. It is also assumed that there are some potential relationships between the protists and the bacterial and archaeal communities that are actively involved with the nitrogen cycle. Nguyen *et al* (2022) posit that protist communities in both bulk soil and rhizosphere play an important role in nitrogen processing through various relationships (e.g., predation, symbiosis) with nitrifying or nitrogen-fixing bacteria or archaea. Predation by protists can stimulate bacterial and fungal communities to produce antimicrobial chemical agents in the rhizosphere and photosphere, thereby protecting plants from airborne or soil-borne pathogens and harmful pests. Protists induce plants to release their stress-mediating hormones by interacting directly with plants or by stimulating plant-beneficial microbiome interactions (Figure 1-2).

There is growing interest in the impact of host signals on crop microbiome assembly. A large-scale field study with 27 homozygous maize lines showed that plant growth stage is one of the main determining factors for rhizosphere community composition (Walters *et al*. 2018). The composition varies across agricultural fields according to different plant genotypes (Walters et al., 2018). Another recent multi-year field study identified some plant microbiomes belonging to the taxa of *Alphaproteobacteria*, *Betaproteobacteria*, and *Actinobacteria*, which may be transmitted vertically from seeds to the plant

rhizosphere and are influenced by plant growth stage and plant genotype (Quiza *et al.* 2022). Yet another study on maize microbiomes from pure inbred lines revealed that Operational Taxonomic Units (OTUs) belonging to *Proteobacteria* and *Actinobacteria* that were identified at the vegetative and reproductive stages can persist throughout the host life cycle (Bourceret *et al.* 2022). Similar studies with wild-type plants at the vegetative stage showed that plant growth stage plays an important role in root metabolism and the composition of the rhizosphere microbiome. This is consistent with results from rice roots (Zhang *et al.* 2018) and maize rhizospheres that provide further evidence on the importance of the plant growth stage on microbial community assembly processes and function (Moroenyane, Tremblay and Yergeau 2021; Xiong and Lu 2022). Other reports demonstrated that plant genotypes influence microbiome recruitment processes in different agroecosystem by changing their root-mediated secretory carbon profiles (Santoyo 2022). A recent study of comparative microbiome analyses between different inbred lines of maize and their F1 hybrid progeny showed a different composition of bacterial and fungal communities in the rhizospheres (Wagner *et al.* 2020).

Neighboring plants and the surrounding air seem to affect the dispersal process of the soil microbiome. Studies show that airborne microorganisms are one of the main sources of phyllosphere microbiomes in maize (52%–92%) (Xiong *et al.* 2021). Another study on the community assembly of the phyllosphere microbiome of tomato, pepper and bean plants showed that the dispersal of microbial communities from neighboring plants greatly affected the indigenous microbiomes that colonize different plant compartments (Meyer *et al.* 2022).

The fungal–bacterial ratio in soil is one of the deterministic factors for niche-based community assembly processes associated with soil nitrogen processing. The ratio of total bacteria to fungal biomass can determine the rate of nitrogen immobilization for different levels of nitrogen availability. A recent study that included an isotopic tracer experiment showed that both bacteria and fungi can immobilize soil inorganic nitrogen from the surrounding soil environment (Li *et al.* 2020). Another report showed a significant positive correlation between soil nitrate immobilization rates and fungal biomass that led to nitrate loss in forest soils (Zhu *et al.* 2013). Similarly, it was demonstrated that topsoil with high microbial biomass increased nitrogen volatilization in the Arabian Peninsula (Bargmann *et al.* 2014).

1.3 Microbial processes in the agroecosystem

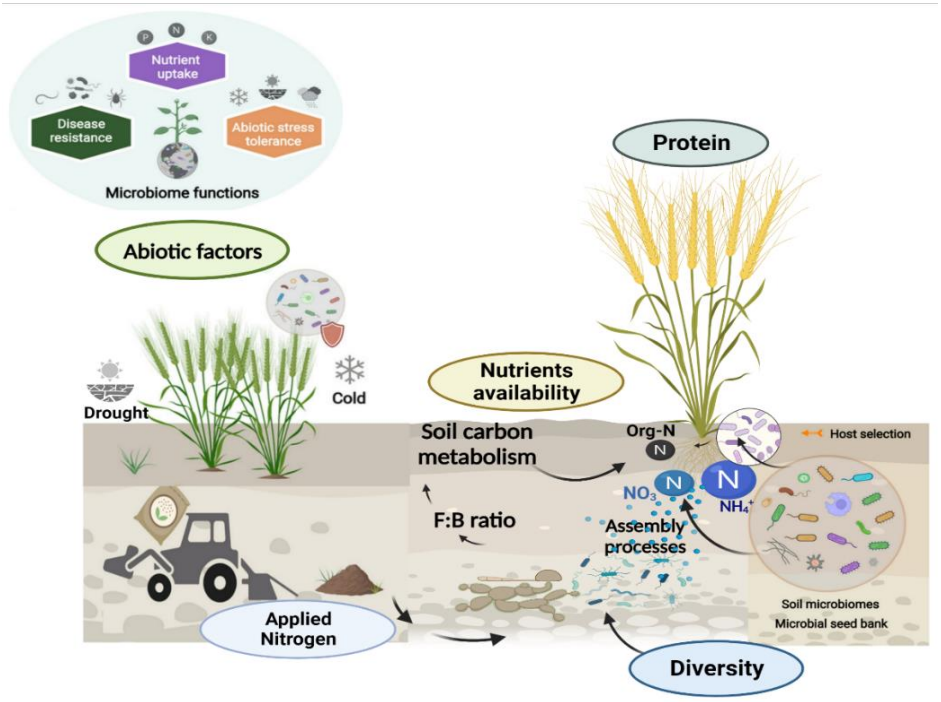


Figure 1-3: Microbial indicators associated with agroecosystem processes. The F:B ratio indicates the ratio of fungal to bacterial biomass.

The soil microbiome supports crops by providing various functions including disease resistance, nutrient uptake, and abiotic stress tolerance. Another major function of the soil microbiome is to process plant nutrients and release free nitrogen for plant absorption through nitrogen cycling. Whenever soil and plant microbiome-driven ecological processes (e.g., assembly processes) are disrupted by biotic (e.g., host selection, plant carbon sequestration) and abiotic factors (e.g., drought, cold, applied N) it limits plant nutrient availability and ultimately affects plant nutrient uptake and crop protein synthesis. There are some microbial indicators such as diversity, the ratio of fungi to bacteria (F: B ratio), and patterns of soil carbon metabolism that are directly or indirectly linked to agroecosystem processes. Therefore, soil microbial diversity provides key signals for many soil processes (e.g., Nitrogen cycle) that are strongly dependent on different levels of microbiome function (e.g., soil carbon metabolism) (Figure 1-3).

1.3.1 Plant–microbe interaction

1.3.1.1 Plant growth promotion

Microbial communities that colonize on root surfaces provide essential mechanistic support for plant growth and are commonly known as plant growth-promoting rhizobacteria (PGPR). PGPR can promote plant growth through nitrogen fixation, phosphorus solubilization, and iron sequestration through siderophores. The most reported mechanism of growth promotion by PGPR is the production of the growth-promoting hormones such as auxin. Previous studies reported that 80% of the rhizosphere microbial community could synthesize and release auxin as a secondary metabolite (Patten and Glick 1996). The root-associated microbiome (rhizosphere) induces vascular tissue differentiation, apical dominance, root initiation (lateral and adherent), cell division, and stem and root elongation through auxin synthesis (Grobelak *et al.* 2015). Optimal production of plant auxin levels required for plant growth may be disrupted when the plant hormone synthesis machinery is affected by abiotic stress. During these periods, the plant maintains its hormone levels by absorbing excess auxin synthesized by PGPR (Patten and Glick 1996). Through this mechanism, indole acetic acid (IAA) molecules synthesized by PGPR stimulate plant root development by balancing the levels of auxin (Spaepen *et al.* 2007). Another plant hormone, ethylene, helps plants tolerate various levels of stress. Plants synthesize high levels of this hormone in response to stressful conditions, including the presence of metals, extreme temperatures, and various chemicals (Ali *et al*. 2010). Many reports have demonstrated that 1-aminocylopropane-1-carboxylate (ACC) is a precursor of ethylene and that microbes degrade ACC using the ACC deaminase enzyme. Too much ethylene results in reduced plant growth, so microbes with ACC deaminase help plant growth by removing this inhibitor (Olanrewaju *et al*. 2017). Cytokinin gene expression is relatively pronounced in several PGPRs, and their induction in plant growth may lead to changes in plant phytohormone secretion levels. It has been reported that inoculating lettuce with *Bacillus subtilis* promotes plant growth by increasing cytokinin levels (Arkhipova *et al.* 2005). Alfalfa crops inoculated with genetically modified rhizobium (*Sinorhizobium meliloti*) increased the levels of cytokinin, helping plants tolerate drought (Xu, Li and Luo 2012). Gibberellic acid is another growth-related hormone that has been reported to be produced by PGPR. Among other PGPRs, *Bacillus sp*. has been identified as a significant producer of gibberellic acids (GA) (Deka *et al* 2015). A wide range of nitrogen-fixing bacteria that fix atmospheric nitrogen through a symbiotic relationship with certain plant species have also been identified. Hydrogen is a by-product of biological nitrogen fixation and is often released by the nodules of leguminous plants (La Favre and Focht 1983). Hydrogen has been shown to have a growth-promoting effect for many crops (Maimaiti *et al.* 2007), probably through the stimulation

of groups of hydrogen oxidizing bacteria (HOBs). For example, some H_2 oxidizing isolates have 1-amino-cyclopropane-1-carboxylic acid (ACC) deaminase activity (Maimaiti *et al.* 2007). It has also been observed that some bacterial isolates adjacent to legume nodules synthesize different plant-growth-promoting signal peptides during hydrogen oxidation (Maimaiti *et al.* 2007). Even these PGPR microbiomes on the root surface can interact with other microbial communities in bulk soil through an extended networking zone, especially in densely planted crops (de la Porte *et al.* 2020). Some growth-promoting rhizobacteria (PGPR) such as *Pseudomonas* sp. and *Bacillus* sp. can convert organic matter into amino acids, then convert amino acids into ammonia through the process of ammonification (Geisseler *et al.* 2010).

1.3.1.2 Pathogen suppression

Plant growth-promoting rhizobacteria (PGPR) can limit the growth of pathogens by reducing the availability of nutrients required for pathogenic growth. Barahona *et al*. (2011) reported that potential PGPRs with biocontrol properties could outcompete pathogens by blocking their target sites from attaching to part of the plant or by limiting nutrient availability.

Siderophore-producing microbes can protect plants by preventing or reducing pathogen proliferation by limiting the availability of iron for pathogens (Shen *et al.* 2013). Siderophores are low molecular-weight secondary molecules produced by microbes during iron-limited conditions. The iron-intake process involves iron binding with organic molecules such as citrate or heme. Siderophores are then synthesized by the membrane-binding apparatus, which forms the iron complex (Fe-II) in soluble form. The iron complex then binds to a specific receptor (Kramer *et al*. 2020). The structural diversity of siderophore-producing microbial communities creates variability in the mechanisms that make hosts susceptible to pathogens and repels iron, and thus limits pathogen growth (Ellermann and Arthur 2017). The closest microbial species with matching receptors can only use the same set of siderophores that have higher iron-binding affinity than microbes using a different set of siderophores. Thus, the siderophore is an important player in mediating inter- and intra-species-specific interactions (Kramer *et al*. 2020). Sometimes, plants take up the soluble form of iron (Fe-II) through a rhizosphere community that produces siderophores. The rhizosphere microbiome consists of a group of siderophores that are produced as a hydroxamate or catecholate and interact with the ferric phase (Fe-III) (Dimkpa 2016). Soil microorganisms that cannot independently produce siderophores borrow siderophores from other microorganisms that have a common binding receptor. These types of siderophores are called xenosiderophores (Winkelmann 2007). In particular, fungi produce extracellular or intracellular siderophores that function either in the transport or storage of ferric ions (Winkelmann 2007). Siderophore-mediated intracellular or extracellular iron depletion depends on the type of fungal species. Iron transport has been observed in fungal communities in

the form of intact siderophore-iron complexes, which is a phenomenon that limits the pathogen's iron acquisition (Winkelmann 2007). Iron-scavenging siderophores that are synthesized by rhizosphere communities have been proven to have a strong growth inhibitory effect on plant pathogens by creating competition for iron (Gu et al. 2020).

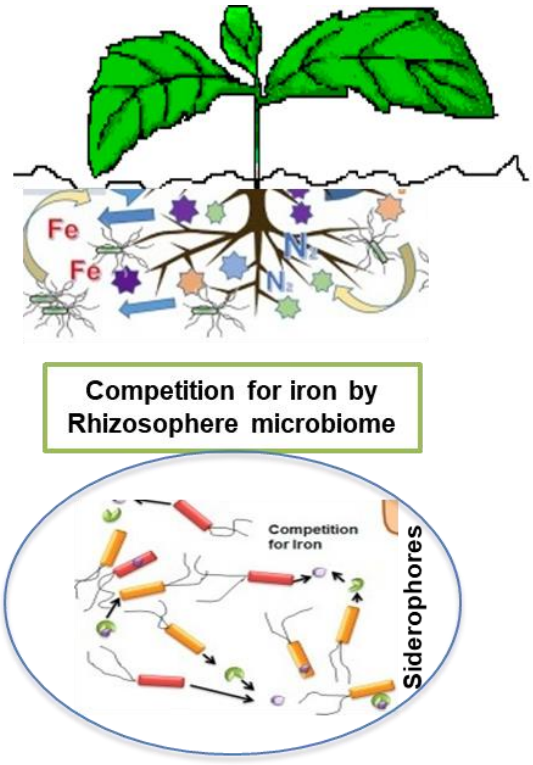


Figure 1-4: Illustration of the general processes of the rhizosphere microbiome in pathogen suppression. Rhizosphere microbiomes create a competitive environment for pathogens by limiting siderophore-mediated iron scavenging. Heterogeneous siderophores produced by microbiomes are incompatible with pathogen receptors that show the greatest inhibitory effect on pathogen density (Gu *et al.* 2020).

Another way PGPR acts against invading pathogens is to synthesize various antibiotics that completely inhibit pathogen growth and infection in plants (Raaijmakers and Mazzola 2012). In addition to plant-derived cell degrading enzymes, some biocontrol PGPRs can synthesize similar cell-wall degrading enzymes (chitinases) to inhibit fungal pathogens (Chernin *et al.* 1995). It is well established that PGPR significantly contributes to the induced systemic resistance and primary resistance to plant pathogens by the activation of signaling molecules (Halfeld-Vieira *et al.* 2006). PGPR-mediated ISR (Induced systemic resistance) and (salicylic) SA-dependent SAR (Systemic acquired resistance) are regulated by different

signaling pathways, demonstrating that PGPR-mediated immunity has potential biocontrol mechanisms with varying levels of efficacy (Ton *et al.* 2002).

1.3.1.3 Stress mitigation

Drought is an example of abiotic stress and is the main limiting factor for global food production. Global national food grain production is estimated to have declined by 9-10% over the past 43 years (Lesk *et al.* 2016). Breeding for drought-tolerant plant genotypes is not a complete solution for mitigating the negative effects of drought, as in some cases the new plant varieties do not yield high crop quality and quantity. Consequently, crop microbiome-based solutions for improving crop productivity and drought resistance have become a field of interest (Marulanda *et al.* 2009). A great deal of research on plant microbiome-assisted drought tolerance is still needed to increase plant resistance and resilience in conditions with limited water resources.

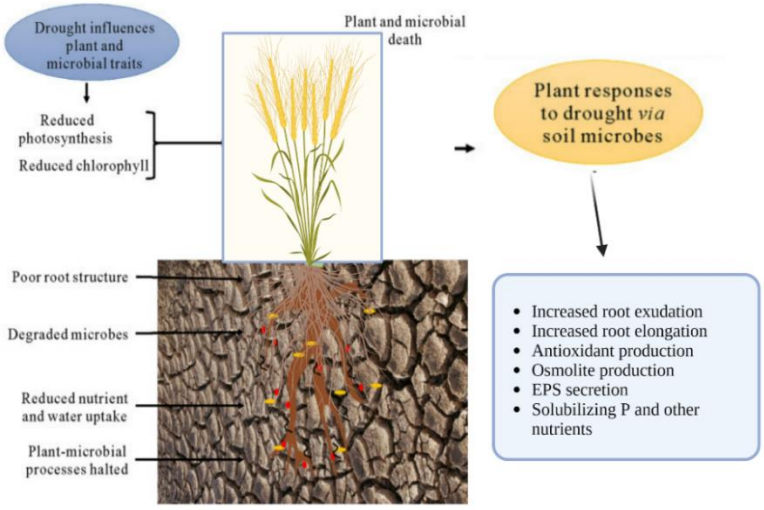


Figure 1-5: The effect of drought on microbial processes. Drought affects plant and microbial processes as well as plant-microbe interactions. Upon exposure to drought, the physiological functions of plants and microorganisms are altered, reducing photosynthesis, and affecting the rhizosphere microbiome. To mitigate stress, plants change their root structure and interact with drought-tolerant PGPR (plant growth-promoting rhizobacteria). EPS = exopolysaccharides.

The stress of drought affects multiple metabolic processes in plants (Figure 1-5), including photosynthesis, respiration, ion uptake, transportation mechanisms, carbohydrate metabolism, and nutrient metabolism or mechanisms of systemic resistance. The result is a disruption in hormonal and nutritional homeostasis in plant growth (Naseem *et al.* 2018). Drought stress also affects rhizosphere microbiome composition and abundance, and in worst-case scenarios, the total microbial biomass in the rhizosphere of selected plants has been reported to decrease by 60%–90% (Naseem *et al.* 2018).

Plant root architecture and topology are known to be very important during times of water stress, as elongated and prolific root systems allow plants to adapt to the stress and increase plant productivity (Castillo *et al.* 2013). Several studies have reported that plants treated with plant growth-promoting rhizobacteria (PGPR) can promote root growth by modifying root architecture (Ngumbi 2016). In addition, rhizospheric bacteria themselves produce hormones or exhibit induced plant hormone synthesis that promotes drought stress resistance. In cucumber plants subjected to drought stress, PGPR can synthesize a higher level of proline and provide plants an osmolyte to stabilize the osmotic pressure (Castillo *et al.* 2013). PGPR treatment typically promotes plant shoot growth. A study on PGPR-mediated shoot growth revealed that plants inoculated with *Bacillus* sp. showed relatively significant shoot growth and an increase in dry biomass compared to non-inoculant plants (Vardharajula *et al.* 2011). One common physiological phenomenon expressed by plants during drought stress is the production of excessive reactive oxygen species (ROSs) such as hydrogen peroxide (H_2O_2), singlet oxygen ($1O_2$), and superoxide radical (O_2), etc. (Cruz de Carvalho 2008). To avoid the effects of ROS, plants have different levels of antioxidative activity. It is well established that the level of antioxidant enzyme production correlates with the degree of drought tolerance (Cruz de Carvalho 2008). Several experiments have shown that PGPR induces plants to synthesize higher levels of antioxidative enzymes during drought conditions. For example, drought conditions led to the highest observed level of specific activity of the iron-scavenging enzyme CAT (Catalase peroxidase), which was 0.8 times higher in PGPR-treated plants compared to non-treated plants (Gururani *et al.* 2013). A higher rate of ethylene synthesis has been recorded as a signaling molecule induced by drought stress. Some PGPR reduce plant drought-related stress by hydrolyzing ACC (1-Aminocyclopropane-1-carboxylate) deaminase into ammonia without converting it directly into ethylene (Shaharoona *et al.* 2006). Bacteria such as *Rhizobium sp.*, *Bacillus sp.,* and *Pseudomonas sp*. and other species of rhizosphere communities can produce exopolysaccharides (EPS), an important class of polymeric compound that significantly enhances the microbiome and allows it to establish ecological niches under drought stress (Fitriani wangsa Putrie *et al.* 2013). Furthermore, it has been demonstrated that EPS-producing *Rhizobium sp.* can significantly promote soil aggregation and water retention in the rhizospheric zone (Kaushal and

Wani 2016). Kour *et al.* (2020) identified drought-adapted and P-solubilizing microbes (*Streptomyces laurentii* and *Penicillium sp*.) that could efficiently accumulate different osmolytes and increase chlorophyll content in millet during drought. Studies have shown that most drought-tolerant rhizosphere microbiomes exhibit a variety of characteristics including dense peptidoglycan cell walls, osmolyte production, dormancy, and sporulation (Xu and Coleman-Derr 2019; Schimel *et al.* 2007).

The response of PGPR-mediated drought tolerance may vary depending on the plant's development, age, level of stress and duration of drought. A better understanding of plant microbiome responses to drought stress in different agroecosystems, overall nutrient cycling, and other changes in soil microbiome diversity can help us better predict the effects of stress on plant productivity and lead to better agricultural management guidelines.

1.3.2 Biogeochemical processes

Soil provides many ecosystem services, including setting the stage for major geochemical cycles, carbon storage and turnover, water maintenance, soil structure arrangement, regulation of biodiversity and the transformation of various chemical compounds. Soil microbes play an important role in maintaining primary productivity aboveground, including within the agroecosystem. Microbes cycle the nutrients from the soil, making them available to plants. The well-studied rhizospheric effect, where microorganisms are stimulated by root exudates, results in hotspots for carbon and nutrient cycling and other ecological processes (Berg 2009). Microbial community structure, diversity and other ecological processes are also partly determined by soil pH, soil organic matter content, and climate, among others (Blanchet *et al.* 2015).

1.3.2.1 Nitrogen cycling

The most readily available forms of nitrogen such as organic nitrogen, ammonium (NH_4^+), nitrite (NO_2^-), nitrate (NO_3^-), nitrous oxide (N_2O), nitric oxide (NO), and dinitrogen (N_2) are found in the environment (Figure 1-6). Ammonia oxidation is a rate-limiting step in the nitrogen cycle that is carried out

by a major part of the microbial community, affecting nitrogen availability in autotrophic or heterotrophic conditions (Kowalchuk and Stephen 2001). Ammonia is a highly volatile compound that readily evaporates under different environmental conditions. The process during which ammonia transforms into nitrate is known as nitrification and the rate-limiting step is ammonia oxidation. Nitrification is carried out by a very narrow group of chemoautotrophic microorganisms in aerobic conditions. Only certain ammonia-oxidizing microbial communities facilitate nitrification in soil ecosystems by oxidizing ammonia into nitrate. Nitrate, being a negatively charged ion, cannot bind properly to the soil and thereby increases nitrate leaching (Kowalchuk & Stephen 2001). Organic forms of nitrogen enter agricultural ecosystems through the degradation of plants, microbial or animal material, where organic nitrogen is further mineralized into ammonium (ammonification). Inorganic nitrogen enters the soil through the application of fertilizers or through the biological fixation of atmospheric nitrogen. Ammonia can then be oxidized and transformed into hydroxylamine by ammonium-oxidizing bacteria (AOB) or by archaea (AOA) through ammonia monooxygenase. Hydroxylamine is then converted to nitrite (NO_2) through hydroxylamine oxidase. Nitrite-oxidizing bacteria then transforms the nitrite into nitrate through nitrite oxidase. Recently, some bacteria were shown to have the enzymatic machinery that enables them to oxidize ammonia directly to nitrate (*comammox*). In anaerobic conditions, nitrate can be transformed by denitrifiers through a process called denitrification. Denitrification can be carried out by a wide range of bacteria, archaea, or fungi, and occurs by reducing nitrates (NO_3^-) to nitrogen (N_2) through first transforming into their intermediate forms NO and N_2O. Incomplete denitrification therefore results in the emission of NO and N_2O instead of N_2. Nitrate reductase, nitrite reductase, nitric oxide reductase and nitrous oxide reductase catalyze the various steps of denitrification (Figure 1-6). During dissimilatory nitrate reduction to ammonium (DNRA), organic matter is oxidized, and nitrate is used as an electron acceptor by reducing nitrite (NO) into nitrate (NO_3^-) and then to ammonium (NH_4). DNRA results in the production of a soluble form of nitrogen rather than dinitrogen (Sparacino *et al* 2014; Simon *et al* 2013). Anaerobic ammonium oxidation (annamox) is an important component in the biogeochemical nitrogen cycle and has the unique metabolic activity of directly converting the ammonium, nitrate, nitrite, and nitrogen. Anammox is the most efficient process for complete ammonium oxidation and requires less energy to reduce greenhouse gases (Kuenen *et al*. 2008). The abundance of nitrogen fixers, nitrifiers, and nitrifiers in agricultural fields depends on land type and crop management. Additionally, C and N dynamics, pH, and soil texture can influence N fluxes (Kooijman, Mourik and Schilder 2009). The intensity of nitrogen mineralization can also influence the relative abundance of bacteria and fungi. A high soil C:N ratio generally favors fungi over bacteria, as fungi have a cellular C:N ratio that is ten times higher than that of bacteria (Kooijman, Mourik and Schilder 2009). Previous studies show that agricultural fields with low C:N ratios had higher levels of carbon degradation and carbon fixation genes, whereas fields with high C:N ratios exhibited elevated levels of gene expression

related to nitrogen fixation and nitrification (Kuramae *et al.* 2014). The transformation of nitrogen and its various oxidative states is the main reaction in the nitrogen cycle. It is highly dependent on the activities of a diverse group of microorganisms, such as bacteria, archaea, and fungi (Grzyb, Wolna-Maruwka and Niewiadomska 2021). Soil type, crop rotation and agricultural (e.g., fertilization, tillage) practices associated with the physicochemical properties of the soil and the environmental conditions influence the biological processes of microbial populations involved in nitrogen fixation, mineralization, and availability (Kracmarova *et al.* 2022). Studies on soil N transformation in tropical forest ecosystems reveal that the unique soil properties (e.g., low pH, rapidly fluctuating redox conditions and large amounts of Fe oxides, plant litter material, available N content) and the environmental conditions (e.g., high humidity and low annual fluctuation in temperature) determine N transformation processes (Xu, Xu and Cai 2013). For example, the rates of microbe-driven N processes showed significant variations across different tropical forest soils both spatially and through time. Soil organic matter (SOM) is one of the main precursors for maintaining soil organic nitrogen levels, as well as the presence of large microbial communities that are able to transform SOM into ammonia through ammonification (Matocha, Dhakal and Pyzola 2012). Some direct links may exist between the process of N fertilization and soil organic nitrogen, affecting the overall C:N ratio (Rasche and Cadisch 2013). Managing soil microbial communities would therefore help to optimize the efficiency of both organic and fertilizer N (Pajares and Bohannan 2016). Because there is generally more NO_3^- than NH_4^+ in agricultural solubilized soil, there is greater risk of NO_3^- leaching in soils with low water-holding capacity, such as sandy soils (Grimvall 2016).

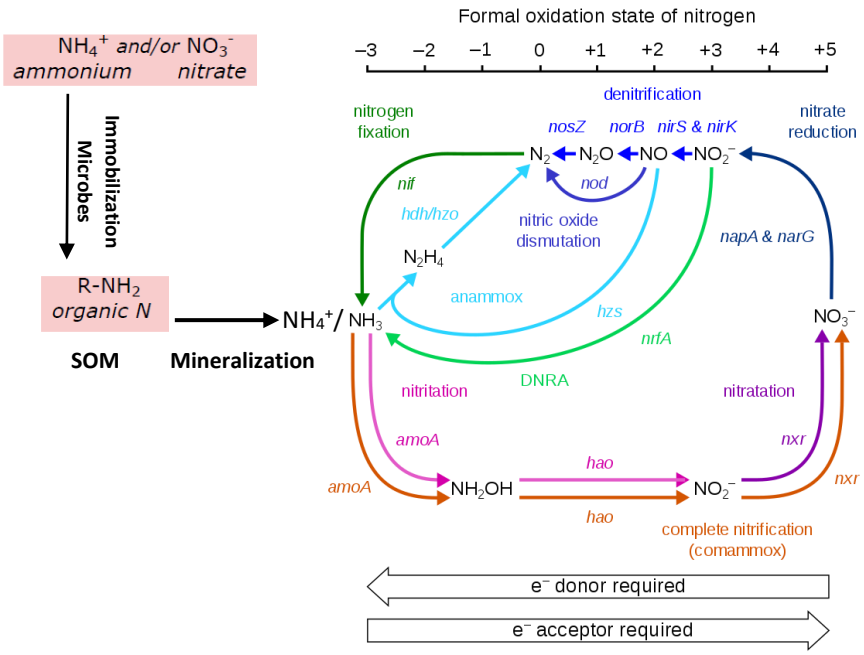


Figure 1-6: The nitrogen cycle (Nelson *et al.* 2015, license: CC BY-SA 4.0). This diagram shows the complete nitrogen cycle, which is carried out through the involvement of different microbial communities using different enzymes for nitrogen conversion, depending on the aerobic or anaerobic conditions of the soil. Nitrogen fixation is the process of converting N_2 gas into ammonia, nitrite, or nitrate by atmospheric, industrial, or biological processes. Ammonia is converted to nitrate through the process of nitrification, which is carried out by microbial communities and uses the enzyme ammonia monooxygenase (AMOA) for ammonia oxidation. Other specialists in ammonia-respiring microorganisms use hydroxyl amine oxidoreductases to convert oxidized ammonia intermediates (such as hydroxyl amine) to nitrate. Nitrite ions are converted to nitrate by other microbial groups using nitrite oxidoreductase. Some microbial communities can convert ammonia to nitrate through complete ammonia oxidation. Microbial communities use nitrite, nitrate, and nitrous oxide reductases to reduce nitrate to nitrite, then to N_2 gas through various reductive reactions involving nitrate-reducing microbial communities. Nitrite is usually reduced from nitrate and then transformed into nitrous oxide or dinitrogen through nitric oxide. Anaerobic microbial communities can directly oxidize nitrite or ammonia to N_2 gas by involving in a process known as anaerobic ammonia oxidation. Microbial communities use various nitrate or nitrite reductase enzymes to reduce nitrate or nitrite to ammonium, returning it to the nitrogen cycle through a process called dissimilatory nitrate reduction to ammonium (DNRA). Microbial communities that use the DNRA pathway primarily oxidize organic matter and use nitrate as an electron acceptor. Ammonification or mineralization is an important microbial process of converting organic nitrogen into ammonium through the decomposition of soil organic matter (SOM), which plays an important role in the nitrogen cycle. Nitrogen immobilization occurs when soil microbes take up ammonia/ammonium or nitrate, which is the opposite of mineralization. As a result, crops may not have access to the N nutrients.

1.3.2.2 Decomposition of soil organic matter and C: N dynamics

Nitrogen mineralization from soil organic matter is known to determine the intrinsic N supply of the soil for plant productivity in natural and agroecosystems (Tiessen, Cuevas, and Chacon 1994). Microbes are the main drivers of nitrogen mineralization from soil organic matter, the rate of which depends largely on the source of organic carbon, temperature, moisture, and aeration. Mineralization of organic carbon usually occurs in moist soil conditions when microbes decompose organic matter, making it easier for other microorganisms to convert organic nitrogen into mineralized forms. The carbon to nitrogen ratio in soil is one potential indicator of N mineralization (Flavel and Murphy 2006). For example, young legume crop residue and cattle manure, which are highly rich in N, result in a low C:N ratio in the soil due to the rapid mineralization process of the organic N that occurs when the organic matter is decomposed (St. Luce *et al.* 2011a).

Soil organic matter (SOM) is composed of soil organic carbon and nitrogen deposited by the degradation of plant and animal residuals. Soil carbon and nitrogen cycles are linked to each other because the elements of these processes (Lindsay *et al.* 2010) are heavily mediated by the SOM that is deposited.

SOM improves soil fertility, which has a large influence on soil structure, water-holding capacity, and plant nutrient content (St. Luce *et al.* 2011b). During carbon sequestration, carbon dioxide is photosynthesized and converted into carbohydrates and stored as plant biomass. Atmospheric N is sometimes fixed by cyanobacteria or plant symbiotic bacterial communities, but most plants depend on N in the soil. The biomass (necro-mass) is decomposed in the soil and returned into the atmosphere in the form of CO_2. A similar process occurs for the N-mineralization form of the organic substances, where the plant inorganic N uptake is triggered by nitrification and returned into the atmosphere in the form of N through de-nitrification. The rate of organic matter degradation might be linked to N mineralization (Haynes 1986). SOM decomposition in the soil could be a significant parameter for studying N dynamics and associated factors. Ammonium ions (NH_4^+) are produced during SOM decomposition and can be converted into nitrate (NO_3^-) through nitrification, where the amount of carbon used by microbial communities indicates changes in total C:N as well as the total N transformation rate in the soil (St. Luce *et al.* 2011; Robertson and Groffman 2015). Different carbon sources that are used by the microbes as a substrate provide information on the functional diversity of the soil and indicate changes in soil carbon dynamics. Spatial and temporal variabilities in soil properties could therefore provide a profile for the accumulation rate of soil organic carbon and the nitrogen linked to the processing of inorganic N. This information could in turn indicate the robustness of microbial functions across time and space for different agricultural practices and climates.

1.4 Predictive modeling

1.4.1 Modeling of microbial ecosystem processes

Soil microbes are key to ecosystem process and management. Knowledge about soil food webs and interactions between the soil, microbes and plant traits is essential to improving the predictive power of current biochemistry-based models on soil processes (Allison and Martiny 2008; Wieder, Bonan and Allison 2013; Faucon, Houben and Lambers 2017; Funk *et al.* 2017; Fry *et al.* 2019; Li *et al.* 2022). Some plant traits such as the leaf nitrogen content and plant growth rate are linked with C storage and decomposition, which can be a useful predictor for understanding temporal and spatial patterns in soil microbes (Carrillo *et al.* 2017). There is abundant literature suggesting that ecological processes can be predicted by microbial activities (Aguilar-Trigueros *et al.* 2015). For example, characterizing microbial methane oxidation and phosphate solubilization activities can effectively indicate dormancy and plant species-specific trait dominance (Caro-Quintero and Konstantinidis 2012). Parameters related to plant microbes and soil interactions have been shown to be helpful for modeling carbon and nitrogen fluxes in soils (Kanters, Anderson and Johnson 2015). Soil microbe and plant interactions as well as nitrogen and carbon use efficiency were shown to be able to predict fruit quality (Pausch *et al.* 2016). Synthetic microbial

communities were used to show that the level of phosphate accumulation and the hormone-mediated systemic immune responses of individual strains of microbes can be used to predict the phosphate uptake of the plant (Herrera Paredes *et al.* 2018). Similarly, the susceptibility of the gut microbiome to Vibrio cholerae invasion could be predicted from the relative abundance of about 100 taxa (Midani *et al.* 2018). A simple bacteria-based decision-making tool was recently developed for soil bioremediation (Horemans *et al.* 2017). Yet another example involves the degradation rate of diesel in arctic soils, which could be predicted through the initial bacterial diversity and the abundance of specific microbial groups like Betaproteobacteria that are also influenced by soil organic matter content (Bell *et al.* 2014). Another study demonstrated that the relative abundance of several taxa in contaminated soils before the experiment started was a good predictor of willow growth in contaminated soils 100 days later (Yergeau *et al.* 2015). Similarly, Zn accumulation in willows growing in a former landfill for 16 months could be predicted by the relative abundance of specific fungal taxa (Bell *et al.* 2015). Finally, soil carbon content and the relative abundance of high-affinity H_2-oxidizing bacteria were able to predict H_2 oxidation rates with more than 80% precision (Khdhiri *et al.* 2015). Although soil and microbial parameters demonstrated the greatest explanatory power in that study, 20% of the variation was unexplained. However, the authors thought that the variation may have been due to fungal communities and unmeasured physical characteristics of the soil. In a different study on the pathogenic fungi Fusarium in asparagus fields, multiple regression showed that the high abundance of Fusarium was linked to high soil organic matter, clay content, and NH_4 (Yergeau *et al.* 2010b). The model results from that study suggested parameters that could be manipulated to reduce the abundance of the pathogen Fusarium in asparagus fields. More recently, scientists in Europe have discussed the scope and potential of statistical learning-based modeling approaches for microbiome-based monitoring of agroecosystem processes (Chable *et al.* 2021). Different classical statistical predictive modeling tools such as equation modeling (SEM), redundancy analysis (RDA), variation partitioning, and multiple regression have been used to generate and test specific models based on soil microbial communities. For instance, variation partitioning, and SEM were used to identify the influences of spatial, soil and plant parameters on the structure of bacterial communities in chalk grasslands (Yergeau *et al.* 2010a).

Machine learning-based modeling has attracted recent widespread attention among microbiologists, particularly with the increased use of ML-based modeling in soil health metrics and human health and disease classification. ML-based predictive modeling has produced accurate predictions of soil health ratings (Wilhelm, Van S and Buckley 2021) and has shown promising results in predicting crop productivity, soil organic matter, and physicochemical properties (Wen *et al.* 2021). The degree of complexity of agroecosystem processes, including multitrophic interactions mediated by different abiotic

and biotic components, varies based on the spatiotemporal dynamics of microbial diversity and community-level physiological processes. It is difficult to characterize the complexity of microbe-driven soil ecosystem processes using only the physicochemical and geochemical parameters of the soil. As discussed by Correa Garcia et al. (2022), the abundance and degree of microbial function of each microorganism is triggered by the combined effects of biotic and abiotic factors, creating a multivariate niche. The presence of thousands of microbes within their own multivariate niche clearly demonstrates the integrated nature of the biotic and abiotic processes in a given ecosystem (Correa-Garcia, Constant and Yergeau 2022). However, there is a wide range of microbiome data that describe the many microbial functions driving the heterogeneous environment in agroecosystems. Most microbiome data are derived primarily from DNA, RNA, and proteins, which are analyzed using modern omics-based methods. Genomic characterization of microbiomes through metagenomics, transcriptomics, and others can provide large numbers of variables with high taxonomic resolutions, creating an abundant resource for predicting microbial functions or processes within limited microbial parameters. To understand the impact of complex agroecosystem processes on crop traits, including yield and grain quality, modeling approaches using microbiome data may be helpful in identifying important microbial parameters.

1.4.2 Statistical modeling with microbiological data

Microbiome data-based modeling approaches continue to be updated with the aim of creating the most reliable and accurate models to characterize ecological processes. Microbial data are used to predict ecosystem processes and often rely on interpretability rather than accuracy. Some non-parametric supervised learning methods are currently being applied in microbiome science to build accurate models such as neural networks. However, these methods cannot explain the underlying relationships between microbial variables because the method predicts the variable of interest based on an intrinsic layer of data, which is difficult for humans to interpret. In contrast, regression models coupled with a dimension reductionist approach is easier to interpret, but this method sacrifices model accuracy (Correa-Garcia, Constant and Yergeau 2022). Other supervised models such as random forest or support vector machines used 16SrRNA sequencing data to identify microbiome composition and accurately predict the soil health metrics in a continental scale (Wilhelm, van Es and Buckley 2021). However, an analysis of the model through the elimination of a thousand sets of taxa failed to highlight the true predictors (taxa) that contributed the most predictive power, further requiring a *posteriori* analysis (Correa-Garcia, Constant and Yergeau 2022). Meanwhile in another study, a multiple linear regression model with forward selection of

microbial indicators predicted wheat grain quality and was able to capture the most important microbial predictors by describing different degrees of relationships between predictors and response variables (Yergeau, Quiza and Tremblay 2020). Both approaches, in terms of model interpretability and non-interpretability, have tracked some common microbial predictors that greatly influence the agroecosystem. Correa-Garcia, Constant and Yergeau (2022) suggested that robust microbial predictors that exist in certain important ecosystems could be tracked using statistical modeling with different levels of interpretability. Choosing the right modeling method in soil microbiome research is highly context dependent.

1.4.3 Definition of statistical learning, model parameters, accuracy, and bias-variance

According to James *et al.* (2013), statistical learning (SL) refers to a set of tools used for modeling and understanding complex datasets. Below are some of the key terms commonly used in statistical learning:

Explanatory variable, independent variable, predictor, or input: The variables that originate from external sources of data and are called predictors or parameters in regression models, denoted by the symbol X. The predictors influence the magnitude of variation in response variables. An example of this type of variable is the indicator of microbial communities associated with agroecosystem processes.

Response variable, dependent variable, predicted variable or output: The response variable is a function of the qualitative or quantitative variables that depend on the value of the independent or explanatory variables. These are denoted as Y and is also called a dependent variable. Examples of the response variable are crop yield, grain quality, and soil health.

Prediction or classification through quantitative or qualitative data can be categorized into two different types of statistical learning: supervised and unsupervised. Supervised learning is mainly used to predict output variables (e.g., agroecosystem processes) by estimating the effects of predictors (e.g., microbial diversity index) based on input variables. In multiple linear regression, the unknown regression coefficient of each predictor or input is calculated based on the magnitude of change in the response variable.

Unsupervised learning explains only the pattern of internal relationships among the input variables or predictors without involving the response variables. Unsupervised modeling is quite commonly used in microbe-based modeling to organize data when the internal patterns of the input variables are not well understood (James *et al.* 2013).

The main limitation of statistical learning models is adjusting the true relationships of the predicted processes to the estimated explanatory variables. There are two main criteria for determining the true relationship of a model: bias and variance. A biased model is complex and fails to illustrate the true relationship between the response and explanatory variables. This is often defined as model underfitting (Correa-Garcia, Constant and Yergeau 2022). Modeling methods applied to noisy training data can result in significant variance, but they are not very effective at predicting new data. This situation is called overfitting (Correa-garcia, Constant and Yergeau 2022). For example, not all predictors are strongly associated with response variables in microbe-based predictive modeling, and sometimes some small subsets of predictors may have high predictive power. The methods used to determine robust predictors (e.g., microbial indices) that are strongly associated with predicted responses (e.g., ecosystem processes) and that fit the models that result in accurate predictions are called variable or predictor selection procedures. Regression models with poor variable selection procedures, particularly in models with trained data, may have high bias that results in low variance or low bias that may result in high variance. Therefore, reducing variables from a large dataset by selecting only the relevant variables can reduce model complexity and increase model accuracy and interpretability. Proper predictor selection enables predictors to accurately learn the structure of the data while avoiding excessive noise (James *et al.* 2013).

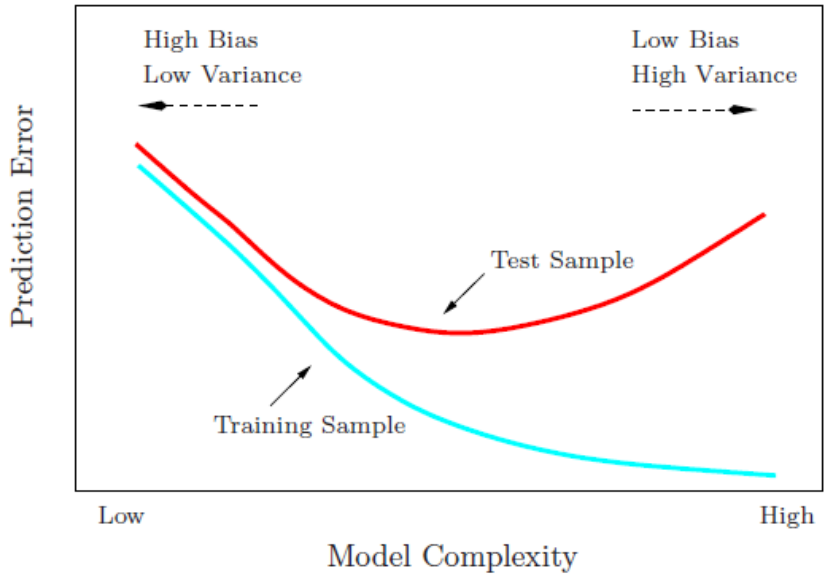


Figure 1-7: Illustration of model complexity and bias-variance trade-off. The figure also shows the inherent trend towards model complexity when building models with training and testing datasets (https://online.stat.psu.edu/ CCBY-NC 4.0 free).

To check the performance of the model with the selected microbial predictors, it is necessary to evaluate the residual error of the model. Depending on the sample size, different adjustment techniques are used to minimize model error. Model error adjustment is an effective method for selecting the best model with the optimal number of predictors. The most well-known cross-validation-based techniques used in multiple linear regression are the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and adjusted R^2. The AIC criterion evaluates a large class of different models generated during model selection, based on which one has the highest likelihood of certifying goodness of fit. The AIC values are adjusted together with the model error and adding fewer predictors decreases the AIC index. The BIC is derived from Bayesian statistics but follows a similar process to AIC. Adjusted R^2 is another model parameter used to evaluate the best model among a set of models with different predictors. Statistically, the adjusted R^2 square value is calculated as the ratio between the residual sum of squares and the total sum of squares (1-RSS/TSS). The R^2 values in linear regression models indicate the proportion of total variance and give a value between 0 and 1 that is completely independent of the behavior of the response variable (James *et al.* 2013). The residual standard error (RSE) is calculated from the standard deviation of the error term of each observation as derived from the input variables (James *et al.* 2013). The residual standard error can be used to determine the model's lack of fitness with the given data. In some statistical learning

approaches, the mean square error (MSE) is estimated by squaring the residuals from the regression model and summing them. The values typically represent the residual error that underlies the difference between the observed and predicted values. Cross-validation-based methods are also applied to obtain penalty scores for the regression models generated from regularization processes (e.g., Lasso). A consistent threshold (e.g., k-fold) for the model parameter is adjusted to the optimal λ value, contributing to model accuracy by reducing model complexity using sparse regression coefficient estimations following the least squares method (James *et al.* 2013).

1.4.4 Supervised learning

Supervised statistical learning techniques, when applied to microbiological data sets, require the right quality and quantity of data, as low-quality data may affect the overall predictive accuracy. Imprecise input data, including incorrect hypothesis testing, can result in error predictions that may require further processing. Using data types with similar characteristics has resulted in good model performance in training, indicating that the model will likely perform well when tested with unknown data from the same distribution (Goodswen *et al.* 2021). It is therefore best to choose modeling tools that have already performed well with data from the same types of microbial ecosystems. For example, two similar modeling methods, multiple stepwise regression models for the prediction of similar microbial ecosystem processes, produced models with high prediction accuracies (Yergeau, Quiza and Tremblay 2020; Asad *et al.* 2021).

1.4.4.1 Modeling approaches for predictor selection (interpretable)

Due to the high dimensionality of large datasets (microbial taxonomic data), some computational statistical methods are used for model selection by creating a subset of explanatory variables (microbial indicators). For example, the stepwise variable selection method provides a subset of microbial variables that allows for a more restricted set of models with a limited number of variables. This method involves simultaneously adding small sets of variables to the regression model to test how much they improve the linear fit. There is also a computational method that is the basis for a hybrid, forward-backward selection method. This hybrid version of the forward-backward stepwise selection incrementally adds variables and eventually eliminates less important variables that do not contribute significantly to the model fit. This approach provides the best option for subset selection, at low computational cost. This type of subset selection is called forward selection. It starts with a null model with no predictors, then variables are

included in steps until all predictors are selected. Each variable selection step identifies the best models with the lowest residual standard error (RSE) and the highest R^2. For example, using the hybrid version of the stepwise multiple regression model, researchers were able to select the best variables to predict wheat grain quality with high accuracies of 64–90% (Yergeau, Quiza and Tremblay 2020; Asad *et al.* 2021a). Stepwise multiple regression was also reported to be the best method for predicting *Fusarium* wilt in lentils (Ali *et al.* 2022).

Another popular method for model fitting with selected predictors (e.g., microbial features) is to constrain or regularize the regression coefficients or shrink the regression coefficients to zero. This compression-based method reduces problems associated with high dimensionality and model bias and is considered the best one for model selection. This method is a useful simulation tool for clustering microbiome composition from OTUs based on phylogenetic relationships and is called a phylogeny-regularized sparse regression model (Xiao *et al.* 2018). Ridge regression and lasso are two techniques commonly used for shrinking regression coefficients towards zero. Ridge regression specifically estimates the weights of the total regression coefficients following the least squares method. Unlike stepwise regression, the ridge includes all predictors with regression coefficient estimates close to zero, but the mean of the coefficients does not equal zero. Therefore, ridge regression can produce good model accuracy, but can be more challenging for model interpretation, particularly for data with large sets of predictors. Lasso (Least Squares Shrinkage Operator), a modern alternative to ridge regression, performs both regularization and predictor selection. In comparison, the statistical formula for lasso is quite similar to ridge regression. The only difference is in the penalty method, where lasso follows the L1 norm and provides a more restrictive penalty score. The L1 penalty for lasso regression forces the regression coefficient to become exactly equal to zero when the tuning parameter λ value is quite large. Therefore, lasso performs better than ridge regression in terms of variable selection and provides an easier interpretation for predictive models. Lasso regression can be computed with input and output variables that sparse the best models with large regression coefficients. Lasso is useful for selecting microbial variables in high-dimensional data where selected variables (predictors) with large coefficients can be used in predictions with linear regression models (Dong *et al.* 2020). Lasso regression can also select microbial variables in a small number of samples, where selected predictors with large coefficients can be used to test the model for prediction accuracy (Asad *et al.* 2023). While human phenotypes (e.g., age, gender, etc.) have been accurately modeled with large microbiome data sets (34,000) from two continents using a penalized regression model (e.g., ridge regularization), modeling with smaller subsamples yields models with less bias and higher variance (Rothschild *et al.* 2022). Sample size is therefore an important factor in microbial-based modeling. This

suggests that modeling microbial ecosystem processes from highly variable environments may require larger samples to produce predictive models with low variance. Another regularization method that combines L1 and L2 norms called Elastic Net is very useful for selecting subsets of variables from large datasets (James *et al.* 2013).

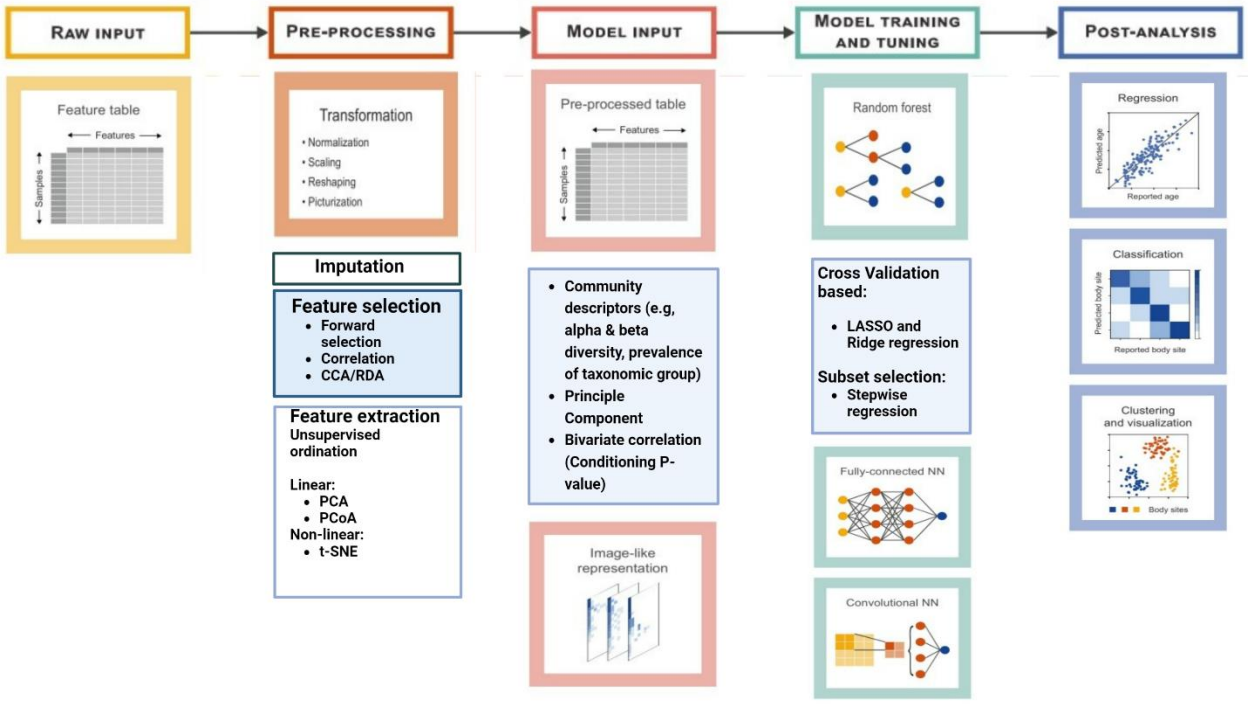


Figure 1-8: Workflow of statistical learning for microbiome analysis (Medina and Kutuzova 2022).

Most of the raw microbiome data from DNA sequencing is formulated using operational taxonomic units of the sequence variants obtained by counting sequence reads according to proximity. Before modeling, microbiome data are pre-processed following the method of transformation, imputation, relevant feature selection, or extraction. Then either processed microbiome data or feature data are used as the model input in various statistical learning algorithms in order to select the best models. The best model is then tuned and highly trained to select the one that generates the best predictive performance. Finally, using the selected model or calculated observed values, it is possible to predict or classify the specific variables (e.g., quantitative, or qualitative variable) of interest.

1.4.4.2 Modeling approaches for accurate predictions (less interpretable)

The support vector machine is a classic machine learning-based approach used for microbiome-based classification of soil health and other health metrics. A support vector machine is also called a maximal margin classifier. This approach defines the n-dimensional subspace through linear boundaries with an extension of support vectors that accommodate a nonlinear class of boundaries. These non-parametric supervised learning techniques are reported to be very effective at classifying soil metrics with 16S rRNA data based on regression or binary classification (Wilhelm, van Es and Buckley 2021). One study reported the use of support vector machines with the relative abundance of 100 taxa to predict human susceptibility to intestinal infection from *Vibrio cholerae* (Midani *et al.* 2018).

Random forest is a machine learning algorithm that combines decision trees to create flowchart-like structures. These flowcharts help determine how to split data sets by voting for trees in the same group. This approach builds multiple trees by bootstrapping the sample to select subsets and binning these subsets as a random forest to increase performance over a single tree (Tin Kam Ho 1995). Random forest algorithms have been reported to work well to predict symbiont density in sponges (Moitinho-Silva *et al.* 2017), maize and wheat yields, and can classify individual patients with various diseases (Chang *et al.* 2017; Yergeau, Quiza and Tremblay 2020). Random forests are able to increase model accuracy but there is increased complexity and training time associated with it. (Wilhelm, van Es and Buckley 2021). A random forest model was used to classify multidrug-resistant bacterial strains on tuberculosis from geographically diverse data and showed high model accuracy (Farhat *et al.* 2016). Random forest was also able to classify the origin of 8287 soil samples based on the bacterial taxonomic abundance of soil samples collected from 21 countries (Ramirez *et al.* 2017).

The gradient boosting decision (GB Boost) tree is a common method, often used to solve regression and classification tasks. This method is carried out through a sampling process with boosting, where a set of decision trees is used to predict only the labeled data. This method has high computational costs but is reported to be a potential tool to study ecosystem functions, particularly human microbiome-based gender classification, or patient countries of origin (Rothschild *et al.* 2022). Stochastic gradient-boosting ML (GBM) techniques are effective predictive tools for microbiome- (structure and dynamics) based classification of cancer cell types between cancerous and noncancerous tissues and for monitoring the progression of tumor cells (Poore *et al.* 2020). K-nearest neighbor (kNN) is an algorithm derived from

machine learning that is a non-parametric method applied for classification or prediction by grouping the nearest data points. For classification, the algorithm labels datapoints based on the k values of the nearest neighboring points. For regression, the average value of neighboring points (clusters) is calculated as the predictive value. These methods have also been reported to be useful for selecting microbial strains from enrichment cultures in different cultivation conditions (Oyetunde *et al.* 2019).

Deep learning (DL) is a class of ML algorithm that builds data architecture based on artificial neural networks. DL models develop artificial nodes (also called neurons or units) by transforming inputs and connecting other nodes along the edges to form a network. The network has multiple layers that represent different layouts or structures. For example, convulsion neural networks (CNN) can be used to picturize microbiome-driven host phenotypes (Reiman *et al.* 2020; Sharma, Paterson and Xu 2020). Using CNN, the OTU table is first converted into an image by transforming each sample into a structural form within shape of a square. The newly formed square-shaped sample images are then organized according to color intensity based on the presence or absence of microbial taxa. Figures are generated from microbial features including species abundance (e.g., OTUs abundance), presence, and absence of taxa. For a single sample, the phylogenetic tree is constructed and grouped by species abundance and then arranged in a data matrix (Nguyen *et al.* 2017). Artificial neural networks have also been reported to accurately predict the parasitic load from clinical samples (e.g., physical signs, serological test, and biochemical markers) without having any prior biological knowledge about the phenomenon (Torrecilha *et al.* 2016).

1.4.5 Unsupervised learning

Unsupervised learning for microbiological data is mainly used to cluster microbial taxonomic (e.g., OTUs) and functional genomic data (e.g., transcript). The clusters are subsequently represented as key data in predictive models. Clustering methods can be hierarchical or divisive. Divisive clustering typically identifies clustered variables that do not overlap with each other, while hierarchical clustering groups together subsets of larger clusters that are always nested. Hierarchical clustering generally does not require the user to input a value for k, whereas divisive clustering requires a value for the k mean. There are many clustering methods. Among them, k-means clustering and agglomerative nested clustering are most often used for hierarchical and divisive clustering in microbiology (Goodswen *et al.* 2021). K-means clustering was found to be effective for assessing antimicrobial resistance from different environmental samples using metagenomics data (Oh *et al.* 2018).

Unsupervised learning techniques are mainly used to look at intrinsic patterns in datasets, or when a response variable is absent. Dimensionality reduction is another method of unsupervised learning that is used for variable selection or extraction by transforming the predictor into an M-dimensional subspace (M < p). Dimensionality reduction-based methods essentially create a compressed representation of the data by computing linear and non-linear combinations of existing features. Linear methods such as principal component analysis (PCA) generate a new set (n-1) of orthogonalized variables (eigenvector) by decomposing eigenvalues. PCA captures variations in the original dataset and ranks them according to the percentage of variation explained. The first few PCA dimensional axes capture most of the variation in the original dataset that can be visualized to explore the variation pattern among the samples. For example, principal components of 16S rRNA gene amplicon data can be used as input variables to predict wheat grain quality and reduce model complexity (Asad *et al.* 2021a, 2023). High co-linearity can sometimes lead to problems of multicollinearity or heteroscedasticity in linear regression models, or model overfitting due to overlapping variables in stepwise regression. Therefore, feature selection using PCA orthogonalization (e.g., decomposition of eigenvalues) can be useful to optimize multicollinearity and bias-variance trade-offs in linear regression-based models (Asad *et al.* 2023). Alternatively, the generalized linear model (GLM)-based ordination technique was applied to distinguish the representative 16SrRNA sequences that originated from different bacterial and archaeal microbial communities with optimal niches. This technique was able to minimize the effects of negative dispersion and statistical sparsity (B. Sohn and Li 2018). T-distributed stochastic neighbor embedding (t-SNE) is a statistical technique used to visualize high dimensional data by locating the data point in sub-dimensional space (2 or 3). By embedding each high-dimensional object into two- or three-dimensional object, this nonlinear method models similar objects by nearest point and dissimilar objects by distant point based on a high probability distance matrix. This approach has been reported to reliably identify and visualize local and non-linear relationships in complex microbiome datasets (Kostic *et al.* 2015).

1.4.6 Example of statistical learning methods in agroecosystems

Microbiomes play an important informative and indicative role in trait-based ecology by providing clues about past and present ecosystem processes (Correa-Garcia, Constant and Yergeau 2022). Microbe-driven agroecosystems can be researched using genomics tools. Current trends in microbial ecological research are mostly focused on the characterization of microbial community structure and function at the genomics level. Exploratory studies on microbiomes currently focus more on compositional changes in

microbiomes as they are exposed to various environmental factors and treatments. However, statistical learning (SL) tools have recently revolutionized microbiome research, allowing observational microbiologists to build model-based frameworks on microbe-driven ecosystem functions from genomic to phenomic scales. One particular model-based microbiome study used genomic data and statistical learning tools to accurately predict and classify various disease patterns for biomarker identification (Marcos-Zambrano *et al.* 2021). The authors were able to predict the quality of water and its resources by tracking the prevalence of 30 bacterial OTUs (*Proteobacteria and Bacteroidetes*) using a random forest algorithm (Wang *et al.* 2021a). Linear regression was shown to estimate soil biodiversity in a boreal ecosystem using data on fungal richness, community composition, and relative abundance as input variables (Li *et al.* 2019). Another study by Wang *et al.* (2021b) demonstrated that soil biodiversity could be predicted through a linear regression-based model using only information on the bacterial species richness and abundance of 39 bacterial genera. In yet another study, an index of differential species abundance (log-ratio) of 140 taxa predicted potato yield with 77% accuracy (Jeanne, Parent and Hogue 2019). Logistic regression was able to accurately predict litter decomposition using data associated with microbial descriptors by grading high and low process rates (e.g., fungal, and bacterial richness) (Albright *et al.* 2020). Using different microbiome datasets from multi-scale studies, patients could successfully be categorized as healthy or disease-susceptible by applying machine learning tools (Marcos-Zambrano *et al.* 2021).

1.4.7 Major sources of microbiome data

Prior to modeling, it is important to consider the sample sources used as input variables to study the microbial traits associated with indicators of ecosystem processes. Based on the type of data used, the right statistical modeling tools must be selected to accurately model the ecosystem processes. Microbial indicators obtained from genomic data can be used to determine microbial function. Microbial genomic DNA, for example, is quite stable in soil and is not very sensitive to environmental changes over short periods of time. This is because changes in the microbial genetic material respond to the environment and usually occur over a season or year (Correa-Garcia, Constant and Yergeau 2022). Therefore, data derived from genomic DNA is a very useful for monitoring microbiome patterns that characterize soil processes, such as gas exchange, across seasons or years (Graham *et al.* 2014). DNA-based investigations of microbial communities could also be a potential indicator for studying the impact of legacy effect in agriculture. The availability of extracellular DNA in the soil can be affected by environmental perturbations, birth, and death events. In contrast, transcriptomic microbial data can capture the profile of contemporary microbial activity, but due to the short lifespan of mRNA, monitoring legacy effects or predicting future microbe-driven

agroecosystem processes might be difficult. For example, RNA-based meta-transcriptomic approaches are useful for identifying microbial communities associated with microbial enzymatic processes that mediate the phytoremediation of environmental pollutants (Yergeau *et al.* 2014, 2018). In addition to amplicon sequencing, a single sequence of shotgun meta-omics data provides vast information on microbial community structure, function, and taxa abundance. This information can be used to indicate microbial process rates and functions, including nitrogen fixation, biotransformation of toxic compounds, and decomposition of organic matter. However, data derived from omics approaches always have high dimensionality.

1.4.8 Common features of microbiome data

The microbial data derived from DNA and RNA amplicon sequencing or shot gun metagenomic sequencing are highly dimensional, meaning the number of variables is larger than the number of samples. For example, the number of ASVs (Amplicon Sequence Variants) and OTUs (Operational Taxonomic Units) calculated from 16S amplicon sequencing to generate taxa abundance or rarify tables (presence and absence) showed high dimensionality, represented as count data (Medina and Kutuzova 2022). In predictive modeling, this high-dimensional data can be modeled with a perfect fit, in which case the features are not always related to the response variable. Adding more features may increase model complexity and bias. In particular, data sparsity for microbial descriptors affects non-parametric approaches such as k-nearest neighbour. There are some solutions that can reduce highly featured data to obtain improved model accuracy or interpretability. Feature selection is one of those solutions in microbiome-based predictive modeling. An example of feature selection is the Spearman correlation-based approach, which results in most correlated ASVs or OTUs with the response variable. Selecting the top 10 correlated ASVs or OTUs improved model accuracy in the linear regression model by setting up the Spearman correlation conditioning p value with a minimum threshold (Yergeau, Quiza and Tremblay 2020; Asad *et al.* 2021a). In contrast, microbial community descriptors can be directly used to predict soil microbial processes with multiple linear regression. An example is beta diversity, which estimate microbial community differences based on the Bray Curtis dissimilarity index and then breaks them down with eigenvalues to produce principal coordinates (Asad *et al.* 2021a). Measures of microbial alpha diversity, including species richness, absolute abundance of functional genes linked to soil processes, or total ratio of the abundance of microbial marker genes (e.g., 16S, ITS), might also be used as potential microbial features in multiple linear regression or penalized regression models to accurately predict crop yield and grain quality (Yergeau, Quiza and Tremblay 2020; Asad *et al.* 2021a, 2023).

1.5 Hypotheses and objectives

General hypothesis:

As discussed above, the soil microbiome is directly associated with agroecosystem processes and crop production. Therefore, upscaling the monitoring process of soil properties from soil geochemical processes to microbial processes can reveal information for improved decision-making for sustainable agricultural practices and management. My main hypothesis is that soil microbes, because of their central role in nutrient cycling and plant health, provide a signal that can be used to forecast wheat yields and baking quality.

General objective:

My objective is to measure basic soil physicochemical properties, microbial functional potential, diversity, abundance and community composition across time and space to find the most significant parameters for explaining wheat yields and grain baking quality.

1.5.1 Specific hypotheses

1. Certain microbial indicators will be strongly linked to wheat yield and grain baking quality across different agricultural fields subjected to a wide variety of management strategies (spatial robustness).
2. Certain microbial indicators that are measured early in the wheat growing season will be strongly linked to wheat yields and grain baking quality at the end of the season (temporal robustness).

1.5.2 Specific objectives

A. To determine the microbial functional potential, diversity, abundance and community composition, and basic soil-physical properties of more than 80 wheat fields across Quebec.
B. To determine the microbial functional potential, diversity, abundance, and community composition over one growing season (sampling every 2 weeks).

1.6 Experimental approach and links between the objectives and the chapters of the thesis

1.6.1 Chapter 2: Determine the microbial functional potential, diversity, abundance and community composition, and basic soil-physical properties of more than 80 wheat fields across Quebec.

The experimental approaches consistent with the first objective described in chapter one focus on the following questions:

1) Can soil microbiomes predict wheat yield and grain quality on a spatial scale? 2) Are microbial indicators better for monitoring soil processes than soil basic physicochemical indicators? 3) How strong is the predictive power of microbial indicators for predicting wheat grain qualities? 4) Is it possible to build a model with microbial parameters for wheat yield and grain baking qualities using simple linear regression? 5) What is the magnitude of the relationship between microbial parameters and the grain quality of wheat? 6) Which microbial parameters in regression models have potential causal links with individual soil processes that can be used in future soil microbiome manipulations?

Farmers used to make fertilizer application decisions for wheat cultivation based on soil tests during the growing season. Could microbe-based information better guide farmers with agricultural management decisions? In this chapter, our study focuses on how microbial traits can be best used to evaluate future wheat yield and grain baking quality. As discussed in the literature, soil microorganisms are key players in soil nutrient cycling and availability, as they are intrinsically involved in decomposition and nitrogen fixation. Therefore, soil microorganisms can be key indicators for measuring soil fertility at the beginning of the crop season to predict the wheat yield and grain quality at the end of harvest. Soil microbes are highly complex, and their functional dynamics are driven by soil ecosystem processes. Regulatory network formation within co-occurring microbial groups is often niche optimum for functional activation. It is not possible to fully describe soil processes using a few isolated microbes and their functions because soil microbial community distribution is highly mediated by the biotic and abiotic factors associated with the agroecosystem. Thus, our modeling approaches using microbial parameters derived from community-level analysis can be used to closely monitor the heterogeneity of soil microbial characteristics that affect crop

phenology. Furthermore, microbial descriptors such as beta diversity, which describes microbial community differences at spatial scales, and alpha diversity, which describes changes in community richness or composition within samples, may be important indicators for monitoring contemporary microbial ecosystems and their compatibility with crops yield and grain quality. These descriptors may provide another avenue for investigating the way anthropogenic impacts on the environment or noise in microbial communities affect crop production.

Our research on the soil microbiome mainly focuses on indicators that are potentially associated with ecosystem functions of soil microbiota, such as the diversity of microbial taxa, *in vitro* carbon utilization patterns, and abundance of microbial N-cycle genes. To test our first hypothesis about the potential for microbes to predict the basic physicochemical properties of soil, we analyzed soil samples from 80 bread wheat fields in the province of Québec, Canada. The samples were collected early in the growing season. To identify the potential microbial parameters, we analyzed soil microorganisms at the community level using molecular biological and biochemical methods. We performed a comparative analysis to determine the predictive potential between commonly used soil and microbial parameters and analyzed basic soil properties. Both microbial and soil physicochemical data were analyzed and compared with yield and grain quality data to uncover the microbial traits that have significant predictive potential. Using these microbial traits, statistical learning methods were applied to model wheat yield and grain baking quality. We aimed to include a few select indicators in the model that only have high predictive power. We also aimed to assess microbial robustness over a 500 km transect in Quebec and determine whether soil microbiome studies can be a potential resource for future wheat quality assessment within an agroecosystem affected by agricultural practices and climate change.

In my first chapter, I also discussed some of the interesting results on microbial predictive performance in all the different types of modeling schemes. Those results highlight how microbial parameters excel at predicting soil physicochemical properties, which fulfills our first hypothesis about the predictive potential of the soil microbiome. Some of the microbial variables that were selected for the models even demonstrated causal relationships with the response variables (yield and quality data). For example, wheat plants require substantial amounts of nitrogen sources such as ammonia to synthesize higher levels of protein and gluten in the wheat kernel. Some models found that the abundance of ammonia-oxidizing bacteria in the soil had a negative correlation with the protein content in wheat grain. These microbial groups are actively involved in ammonia oxidation and convert ammonia into nitrate. Since nitrate is unstable in soil, excess ammonia oxidation can lead to nitrate leaching and increased production

of greenhouse gases. A high abundance of ammonia-oxidizing bacteria leads to more nitrate in the soil than ammonia. This is problematic for plants, which absorb ammonia more actively than nitrate. Thus, a high abundance of ammonia-oxidizing bacteria produces fewer effective sources of nitrogen for plant uptake, reducing grain quality. These results can provide a framework for customized experiments that aim to contribute to the future mechanistic understanding of plant–microbiome interactions.

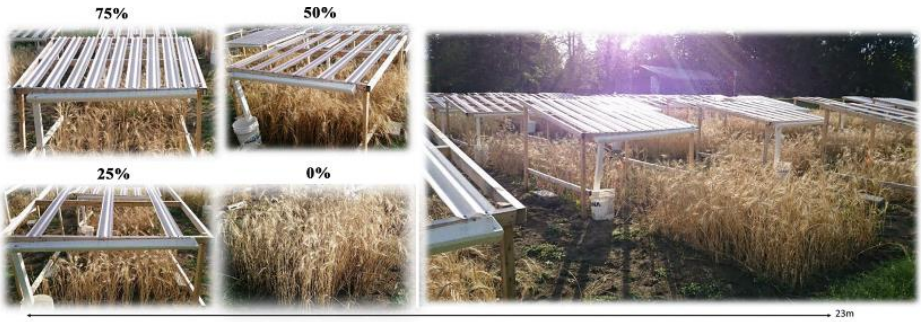
1.6.2 Chapter 3: Determine the microbial functional potential, diversity, abundance, and community composition over one growing season.

In the second chapter, we described how we sampled early in the growing season, but was that the optimal timing to determine predictive power? My second objective is to answer questions about the temporal dynamics of microbial communities, such as how microbial diversity changes over time in agroecosystem processes that affect crop yield and quality. The predictive power of microbial parameters may vary over time, as host selection and nutrient acquisition may be influenced by carbon fixation at different stages of wheat plant growth. As soil properties may be affected by season or regional climate, nutrient access may become limited at times for some microbial communities. Temporal changes throughout the wheat growing season in microbial diversity, composition, or functional abundance could alter the predictive power of the soil microbiome, raising questions about the optimal time for prediction. Our experimental method aimed to find the best model for predicting wheat grain quality at different growth stages throughout the season. Experimental approaches consistent with those for objective one and described in chapter two focus on the following questions:

1) How can we ensure that measurements of microbial indicators at the early stages of wheat cultivation can be highly predictive for wheat quality? 2) How accurate would the prediction be if soil samples were collected after seeding or at a later stage of wheat growth? 3) Do temporal variations in microbial community composition, diversity, and function affect microbial parameters that predict wheat grain quality as a result of changes in seasonal climate, soil water availability, and growth stage? 4) Is there a relationship between wheat genotypes and soil microbial parameters? Do wheat genotypes affect soil microbial diversity and function, and consequently, do genotypes have significant influence on the predictive potential of the soil microbiome? 5) If there is a difference in predictive potential among wheat genotypes, which genotype shows higher predictive accuracy? 6) On which specific dates what times of the year should soil be sampled for accurate predictive modeling? 7) Are there any alternative statistical approaches in the least squares

method that can be used to construct interpretable, simple, and accurate models with high dimensional data? 8) How can the best models be validated to demonstrate high predictive accuracy?

We previously observed that microbial indicators can accurately predict wheat yield and grain quality. However, we must determine the best moment for predicting wheat grain quality to prove our second hypothesis. We chose the experimental field located at INRS for soil sampling. This field was contained 6 random blocks, with 4 rainfall exclusion treatments and two wheat genotypes. Our sampling scheme was roughly implemented according to the wheat growth stage, starting from seedling to crop maturation. We measured the same microbial indices to investigate how they function on temporal scales. Another experiment by a researcher of our team found that soil drying and rewetting episodes triggered by sudden rainfall in mid-July shifted microbial communities and increased the abundance of ammonia-oxidizing archaea (Wang *et al.* 2022). Such episodes have become quite common in the context of climate change. Because of these events, the nutrient status of the soil may change, or the functions of the soil microbiome may be limited. We know that these temporal variations in microbial parameters caused by external factors can disrupt microbiome assembly processes and create complex microscale environments in the soil. Even a temporary change in microbial community composition can disrupt the nutrient uptake of the wheat plant. Therefore, a randomized experimental field may be a good site to more closely study the temporal dynamics of soil microbial communities and their predictive ability.



50% DT	25% DT	75% DS	50% DS	0% DS	75% DT	0% DT	25% DS	B1
0% DS	75% DS	50% DT	25% DT	50% DS	25% DS	75% DT	0% DT	B2
75% DT	0% DS	50% DT	25% DS	25% DT	50% DS	0% DT	75% DS	B3
25% DS	75% DT	50% DS	0% DT	75% DS	0% DS	50% DT	25% DT	B4
75% DS	50% DS	25% DT	25% DS	50% DT	0% DT	75% DT	0% DS	B5
25% DT	75% DS	0% DT	50% DT	0% DS	25% DS	75% DT	50% DS	B6

17m

Figure 1-9: Field testing to find robust microbial indicators. The figure depicts a multi-year historical experimental field (INRS) arranged with a randomized block design. The experimental field was separated into 6 random blocks with four precipitation exclusion (0%, 25%, 50%, 75%) treatments and two wheat genotypes: drought-sensitive (DS) wheat and drought-tolerant (DT) wheat.

We determined that the best models for wheat grain quality used data from samples collected early in the wheat growing season. This fully supports our second hypothesis and emphasizes that the most robust microbial parameters are from samples collected in May and June. The best models were from dates that occurred roughly during the seedling stage of wheat growth. This stage of growth is crucial for the nutrient uptake necessary for plant physiological processes. Another important feature of the microbial parameters in the models is discussed in the chapter 3: section 3.5 on legacy effects on the soil microbiome. Our results also illustrate that predictive accuracy decreases over time as microbial complexity increases. This indicates that the determinants of the soil microbiome and their function may be affected by intrinsic or extrinsic environmental effects during wheat growth. These patterns are clearly demonstrated in our correlation study between microbial parameters and grain quality. In this chapter, we also discuss the role of ammonia-oxidizing microbial communities and show that AOB (ammonia-oxidizing bacteria) or AOA (ammonia-oxidizing archaea) are strong microbial parameters in many models, contributing positively or negatively to wheat grain quality. Finally, we discuss host-specific (e.g., wheat genotypes) microbial predictive power and how the model showed different degrees of predictive power for the two selected wheat genotypes.

Chapter 2: Predictive microbial-based modelling of wheat yields and grain baking quality across a 500km transect in Québec

Modélisation microbienne prédictive des rendements de blé et de la qualité boulangère des grains sur un transect de 500 km au Québec

Numan Ibne Asad[1], Julien Tremblay[2], Jessica Dozois[1], Eugenie Mukula[1], Emmy L'Espérance[1], Philippe Constant[1], Etienne Yergeau[1]

Authors:

[1]Institut national de la recherche scientifique, Centre Armand-Frappier Santé Biotechnologie, 531 boul. des Prairies, Laval, QC, H7V 1B7, Canada.

[2]National Research Council Canada, Energy Mining and Environment, 6100 Royal mount Ave., Montreal, QC, H4P 2R2, Canada

Title of journal or book: FEMS Microbiology Ecology, Volume 97, Issue 12

Published online: 09 December 2021

DOI: https://doi.org/10.1093/femsec/fiab160

Contributions of the authors:

Numan Ibne Asad: Contributed to the main experimental design of the research, conceptualization, methodology, laboratory experiments, data processing and analysis, all statistical analysis and predictive modelling, original draft writing and revisions.

Julien Tremblay: Contributed to the analysis of bioinformatic data.

Jessica Dozois & Eugenie Mukula: Contributed to the soil sampling across the Quebec wheat farm, and the analysis of soil biochemistry.

Emmy L'Espérance: Contributed to the measurement of the total fungal: bacterial biomass through quantitative PCR.

Professor Philippe Constant: Conceptualization, methodology, supervision, reviewing, resources, and infrastructure.

Professor Etienne Yergeau: Conceptualization, methodology, supervision, writing and editing of the original draft, resources, and infrastructure.

2.1 Abstract

Crops yield and quality are difficult to predict using soil physico-chemical parameters. Because of their key roles in nutrient cycles, we hypothesized that there is an untapped predictive potential in the soil microbial communities. To test our hypothesis, we sampled soils across 80 wheat fields of the province of Quebec at the beginning of the growing season in May-June. We used a wide array of methods to characterize the microbial communities, their functions, and activities, including: 1) amplicon sequencing, 2) real-time PCR quantification, and 3) community-level substrate utilization. We also measured grain yield and quality at the end of the growing season, and key soil parameters at sampling. The diversity of fungi, the abundance of nitrification genes, and the use of specific organic carbon sources were often the best predictors for wheat yield and grain quality. Using 11 or less parameters, we were able to explain 64 to 90% of the variation in wheat yield and grain and flour quality across the province of Quebec. Microbial-based regression models outperformed basic soil-based models for predicting wheat quality indicators. Our results suggest that the measurement of microbial parameters early in the season could help predict accurately grain quality and quantity.

2.2 Introduction

Nitrogen fertilization is one of the most crucial factors to produce high-quality cereals and, in the province of Québec, it is recommended to indiscriminately fertilize bread wheat with 90-120 kg/ha of N (Vanasse *et al.* 2012). However, N fertilization levels are not necessarily related to wheat yields or grain baking quality (Yergeau *et al.* 2020, Ayoub *et al*.1994, Guarda *et al.* 2004, López-Bellido *et al.* 2001) suggesting that the applied N is not efficiently used. This inefficient use of N is at the core of the sustainability issues facing the agricultural sector, with the unused N either leaching to cause eutrophication of surface water or leading to the emission of the potent greenhouse gas nitrous oxide. Most of these issues are related to soil microorganisms involved in the N-cycle. Indeed, soil microorganisms are responsible for 1) nitrification, the transformation of ammonia in nitrate, 2) denitrification, the transformation of nitrate in the gaseous compounds nitric oxide, nitrous oxide, and dinitrogen, and 3) mineralization, the release of nitrogen from the soil organic matter (Tiessen *et al.* 1994). Not only some of these processes have the potential to severely impact the environment, but in the case of wheat, it changes the amount of energy needed to produce high baking quality grains with high levels of protein and gluten. For instance, nitrate needs to be taken up actively by the plant and then reduced to ammonia before being assimilated into an amino acid through glutamate or glutamine, whereas ammonia is mostly taken up passively and directly assimilated (Moreau *et al.* 2019). The uptake of amino acids or peptides would be even more energetically favorable to the plant. In view of the importance of N nutrition for cereals, most especially for bread wheat, to produce high quality grains, optimizing the usage efficiency of the applied N, or at least being able to better predict its effect on crop quality, is of outmost importance.

There is evidence that various ecosystem processes can be predicted from microbiological data. For instance, it has been reported that biodegradation of diesel in Arctic soils could be predicted with an accuracy of 60% by the relative abundance of three specific *Betaproteobacteria* taxa (Bell *et al.* 2013). Willow plants growth after 100 days in highly contaminated soil can be partially predicted by the microbial community composition (bacteria and fungi) and their relative abundance (Yergeau *et al.* 2015). Similarly, zinc assimilation by willow trees after 16 months of growth could be predicted with an accuracy of up to 63% by the relative abundance of a single fungal species at 4 months of growth (Bell *et al.* 2015). Other reports provided a bacteria-based predictive tool for soil bioremediation (Horemans *et al.* 2017) or predicted the susceptibility of the human gut to infection by *Vibrio cholerae* from the relative abundance of about 100 taxa (Midani *et al.* 2018). Finally, the phosphate content of Arabidopsis when in interaction with a synthetic community could be inferred from the results of the interaction between the plant and the individual members of the synthetic community (Herrera Paredes *et al.* 2018). More recently, with an

accuracy up to 81%, a multiple regression model was able to predict wheat yield and grain quality based on the abundance of five specific microbial taxa and N-cycle functional genes (Yergeau *et al.* 2020). In contrast, N fertilization treatments could not be related to yield and grain quality. Thus, it appears that the composition, diversity and relative abundance of microorganisms and functional genes in different environments are good predictors of future processes. However, as soil parameters such as pH, total carbon and nitrogen, and water content can also partly determine the microbial diversity and function, and thereby ecological processes (Blanchet *et al.* 2015), it is not known if microbial parameters will provide any additional predictive power. As most of these previous examples are based on samples from relatively restrained geographical areas, it is not known if these models would hold at larger scales. One recent example showed that soil health measures at the continental scale could be predicted with an accuracy of 80% from 16S rRNA gene data (Wilhelm *et al.* 2022).

Here, we hypothesized that microbial parameters would be able to better predict the yields and grain quality at a large geographical scale than selected soil parameters. Early in the growing season, we sampled soils from 80 fields across a transect of almost 500 km, without limiting our efforts to a certain type of soil, agricultural management, or variety of wheat. The regions visited and wheat varieties sowed were representative of the wheat farms of the province of Québec, Canada. We measured the following parameters from the soil samples: 1) bacterial, archaeal and fungal diversity using PCR amplicon sequencing, 2) community-level substrate utilization patterns using Biolog EcoPlates, 3) abundance of functional genes involved in the N-cycle using qPCR, 4) total bacterial and fungal abundance using qPCR, and 5) soil pH, total C, total N, soil water content and C: N ratio. We also measured yields and grain baking quality at the end of the growing season.

2.3 Material and methods

2.3.1 Soil sampling

In May-June 2018, 80 wheat fields in the province of Québec were sampled. These fields were distributed across Québec, from the Montérégie (45.1489° N, 73.3054°W) all the way to Saguenay-Lac St-Jean (48.2511° N, 71.4758° W) and were planted with the following fall or spring wheat (*Triticum aestivum*) varieties: Walton, Warthog, Harvard, Scotia, Touran, Dakosta, Helios (Supplementary Table S1). The rotation and management regime of the farms and the size of the fields varied across all farms, with some being under organic management or not. From each field, 5 soil samples of approximately 200 g each were taken from the upper 10 cm of soil from each corner and the centre of the field. If wheat plants were present (fall wheat), soil samples were taken between rows, 10-25 cm away from the plants. The five samples were then mixed creating one 1 kg composite sample per field. The samples were kept on ice packs in a cooler and brought back to the lab where a part was frozen at -20°C for molecular analyses, a part was transferred at 4°C for Biolog EcoPlates analysis, and a last part was air dried for soil physico-chemical analyses. Between each field, all soil was brushed off the hand shovel and the container used to mix the soil, which were then rinsed with 70% ethanol.

2.3.2 Soil physico-chemical properties

For soil pH measurements, dried composite soil samples were homogenized (2 mm sieve) and pulverized with mortar and pestle. One gram of the homogenized soil was mixed with 9 ml of 0.01M $CaCl_2.2H_2O$. After 30 minutes, pH was measured with a pH meter. Soil water content was measured as the difference between the weight of soil before and after overnight drying at 75^{o} C. Total soil carbon and nitrogen was measured by automated combustion techniques using an elemental analyzer (AgroEnvironnement Lab, La Pocatière, QC).

2.3.3 DNA extraction and amplicon sequencing

DNA was extracted from 500 mg sub-samples of each soil using the DNA PowerSoil Kit (Qiagen, Montreal, Canada) according to the manufacturer's protocol, with the exception that the bead-beating step was performed for 45s at speed 4 on a FastPrep homogenizer (MP Biomedicals, Southern California, USA). Amplicon libraries were generated using primers 515F and 806R (Caporaso et al. 2012) targeting the bacterial and archaeal 16S rRNA gene V4 region and using primers ITS1F and 58A2R (Martin & Rygiewicz, 2005) targeting the fungal ITS1 region following the Illumina "16S Metagenomic Sequencing Library Preparation Guideline" (Part #15 044 223 Rev. B). For each primer set, all samples were mixed equimolarly and the two resulting pools (one for the 16S rRNA gene and one for the ITS region) were sent

to the Centre d'expertise et de services Génome Québec (Montreal, Canada) for Illumina MiSeq 2 × 250 bp pair-end sequencing. A total of 19,262,942 16S rRNA gene reads and 23,107,396 ITS region reads were produced (Supplementary Table S2).

2.3.4 Bioinformatics

Sequencing data was analysed using AmpliconTagger (Tremblay & Yergeau, 2019). Briefly, raw reads were scanned for sequencing adapters and PhiX spike-in sequences. We removed single end reads that met one of the following conditions: having average quality Phred score lower than 25; having 30 bases of quality lower than Phred score 15; having 1 or more undefined bases (N). The remaining sequences were processed for generating Amplicon Sequence Variants (ASVs) in DADA2 (v1.12.1; Callahan et al., 2016). Since the quality filtering step was performed in a separate upstream step, we used more lenient parameters for the DADA2 workflow: filterAndTrim (maxEE = 2, truncQ = 0, maxN = 0, minQ = 0). Errors were learned using the learnErrors (nbases = 1e8) function for both forward and reverse filtered reads. Reads were then merged using the mergePairs (minOverlap = 12, max Mismatch = 0) function. Chimeras were removed with DADA2's internal removeBimeraDeNovo (method = 'consensus') method followed by UCHIME reference (Edgar et al., 2011). ASVs were assigned a taxonomic lineage with the RDP classifier (Wang et al., 2007) using the Silva release 128 database (Quast et al. 2013) supplemented with eukaryotic sequences from the Silva database and a customized set of mitochondria, plasmid, archaeal and bacterial 16S rRNA gene sequences (see the AmpliconTagger databases, doi:10.5281/zenodo.3560150). The RDP classifier gave a score (0 to 1) to each taxonomic depth of each ASV. For each ASV, the taxonomic lineage was reconstructed by keeping only the taxa that had a score ≥ 0.5. Taxonomic lineages were combined with the cluster abundance matrix obtained above to generate raw ASV tables. From these raw ASV tables, ASV tables only containing bacterial and archaeal ASVs or fungal ASVs were generated. In total, 34,373 bacterial and archaeal ASVs and 9,224 fungal ASVs were identified. To normalize these ASV tables, 1000-reads rarefactions were performed 500 times and the average number of reads for each ASV of each sample was then computed to obtain consensus normalized ASV tables.

2.3.5 Real-time PCR

The abundance of genes involved in key steps of the nitrogen cycle was quantified using quantitative real-time PCR (qPCR) with SyBr green. The genes targeted were: the bacterial ammonia monooxygenase subunit A gene (*amoA*), using primers amoA1-f* (5'- GGGGHTTYTACTGGTGGT-3') and amoA2-r (5'-CCCCTCKGSAAAGCCTTCTTC-3' (Levy-Booth et al. 2014), the archaeal *amoA*, using

primers crenamoA23-f (5'-ATGGTCTGGCTWAGACG-3') and crenamoA616-r (5'-GCCATCCATCTGTA-3') (Tourna et al. 2008), the nitrous oxide reductase gene (*nosZ)* using primers nosZ1-f (5'-WCSYTGTTCMTCGACAGCCAG3') and nosZ1-r (5'ATGTCGATCARCTGVKCRTTYTC-3') (Henry et al. 2006), the copper-containing nitrite reductase gene (*nirK*), using primers 876f (5' ATYGGCGGVCAYGGCGA3 3') and 1040r (5'- GCCTCGATCAGRTTRTGGTT-3 (Henry et al. 2006). The abundance of the bacterial and archaeal 16S rRNA gene and of the fungal ITS region was quantified using the same primer sets used for sequencing. Standard curves for N-cycle genes were created by linearizing (SacII restriction enzyme digestion) P-Gem T plasmid (Promega Corporation, USA) into which a target gene of interest amplified from DNA extracted from an agricultural soil was cloned. The linearized plasmids were then serially diluted (10^8-10^1 copies μl^{-1}). For the 16S rRNA gene, full-length 16S rRNA gene amplicons of *Escherichia coli* 25922 made using primers PA-27F-YM and PH-R (Edwards *et al.* 1989) , were serially diluted. For the fungal ITS region, the standard curve was prepared from serial dilutions of ITS region amplicons from the yeast *Pichia scolyti* using the primers NSA3 and NLC2 (Martin & Rygiewicz 2005). qPCR assays were performed using the iTaq universal SYBRGreen kit (Bio-Rad Laboratories Inc, Hercules, CA) on a Stratagene Mx3005P qPCR system running the MxPro Mx3005P software (v4.10; Agilent Technologies, Santa Clara, CA). Each 25-µL master mix reaction contained 1X Master Mix (HotStar iTaq DNA Polymerase, dNTPs, MgCl2, SYBR Green I dye), 300 nM of primers and 5 µL of DNA template at a concentration of 1ng/µL. Amplification conditions are given in Supplementary Table S3. Melting curve analyses were performed at the end of each run to confirm the absence of non-specific amplification. All the standard curves had R^2 of 0.98 or higher, whereas amplification efficiencies were of 94.0%, 86.9%, 108.9%, 96.8%, 54.0% and 57.3% for the bacterial *amoA*, archaeal *amoA*, *nirK*, *nosZ*, 16S rRNA gene and ITS region, respectively.

2.3.6 Community-level carbon utilization profiling

EcoPlates colorimetric assays (Biolog, Hayward, CA), comprising of 31 different carbon sources were inoculated with a 1/10 soil dilution (in water). Carbon utilization was observed using a spectrophotometer following an incubation of 168 hours in the dark at room temperature. The pink-purple color intensity that results from the reduction of tetrazolium dye following substrate utilization was used as an indicator of substrate utilization by the microbial community.

2.3.7 Yields and baking quality

Yields and a sample of grain for quality assessment were provided by participating growers on a voluntary basis. From the 80 fields sampled, we were able to retrieve yield data and a sample of grain from 33 fields. The grain and flour baking quality were analyzed in the quality control laboratory of Les Moulins

de Soulanges (St-Polycarpe, QC) for the following parameters: grain humidity, grain protein content, grain test weight, grain gluten content, grain starch content, flour ash content, flour peak maximum time (PMT, time for the dough to reach its maximum consistency following hydration), flour maximum recorded torque (BEM, maximal consistency as measured as resistance to mechanical mixing), flour coarse wheat germ (CWG), flour falling number (amount of sprout damage), flour Zeleny number (sedimentation value) (Freund and Kim 2006). A good quality grain for bread is expected to have a high-test weight, a high coarse wheat germ, and a high protein and gluten content and a low starch content. The resulting good quality flour will have a low ash content, a high Zeleny number (low sedimentation), a high falling number (low sprout damage), a high maximum torque (high consistency) and a short peak maximum time (rapid to reach maximal consistency).

2.3.8 Statistical analyses

All statistical analyses and figure generation were performed in R (v.4.0.3). The effects of wheat variety and regions on wheat yields and baking quality, soil physico-chemical characteristics and microbial taxa relative abundance was tested by ANOVA. Prior to ANOVA, the normality of the data was tested by the Shapiro-Wilk test (*shapiro.test*), and if the data was not normally distributed, log or square root transformation was performed. If the transformation failed, the Kruskal-Wallis tests were performed using the *kruskal.test* function instead of ANOVA. Before correlation and regression analyses, outlier samples were removed from the dataset using the *rstatix* package. The Spearman correlation tests were performed using the *cor.test* function with Benjamini-Hochberg p-value correction for multiple tests using the *p.adjust* function. Multiple stepwise regression analysis was performed using the *lm* function with the *step* function for stepwise forward and backward selection of variables. In some cases, to generate models that could be compared between soil and microbial parameters, the procedure was limited to select the five variables showing the highest reduction of the mean square error. The Residual Standard Error (RSE) and Akaike Information Criterion (AIC) were calculated for each model using the *stepAIC* function of the mass package. Multicollinearity in regression models was examined by calculating the Variation Inflation Factor (VIF) for each variable using the *vif* function of the car package.

2.3.9 Data availability

The raw datasets and associated metadata are available through NCBI BioProject accession PRJNA749034.

2.4 Results

2.4.1 Yields and grain quality

From the 80 fields sampled, we were able to retrieve yield data and a sample of grain from 33 fields. As expected, grain quality and yields significantly varied across the regions and the varieties (Table 2-1). The highest yields were measured in Montérégie, most specially for the Warthog variety, whereas the lowest yields were measured for the Touran variety in Lac St-Jean and for the Scotia variety in Mauricie. The Walton and Scotia varieties had grains with the highest protein and gluten content when grown in Mauricie, and similarly, the Scotia variety grown in Mauricie had the lowest PMT and the highest BEM. The regional differences were not related to geographical distances between the fields, as Mantel tests between similarity in quality parameters and geographical distance (km) did not result in significant correlations.

Table 2-1. Average yield and grain quality data across Quebec wheat farms. Average yields and grain and flour quality parameters averaged across regions and wheat varieties together with ANOVA or Kruskal Wallis tests results (N=33)

Region	Variety	Humidity %	Protein %	Starch %	Zeleny %	Gluten %	Ash %	Falling Number (Sec)	PMT (Sec)	BEM (Brabender Units)	CWG %	CN	Yield (T/ha)
Centre-du-Qc	Scotia	14.9	15.5	67.2	70.2	31.3	1.8	311.3	71.0	58.0	34.1	11.0	2.9
	Walton	14.3	14.6	67.2	64.1	29.6	2.5	422.0	64.0	58.0	33.2	10.6	3.1
Estrie	Warthog	14.6	14.0	67.1	60.2	28.1	1.5	359.0	64.5	50.5	31.3	9.9	2.9
Mauricie	Scotia	14.1	18.0	65.0	89.8	37.3	1.2	355.6	58.9	61.0	34.9	11.3	1.8
	Walton	15.2	17.5	66.1	83.2	35.2	0.8	343.0	68.0	57.0	33.6	10.8	2.7
Montérégie	Walton	13.6	13.0	67.6	52.5	25.8	1.0	370.0	101.3	43.8	31.2	9.8	4.7
	Warthog	14.0	14.0	69.2	57.2	27.1	0.8	432.3	59.7	47.3	30.6	16.0	5.8
Lac St-Jean	Touran	14.9	15.9	67.5	72.1	31.6	1.2	416.0	61.0	57.0	33.7	10.8	1.5
Lanaudière	Harvard	13.7	13.6	68.2	55.1	27.1	1.7	397.8	91.5	47.5	31.6	10.0	3.1
	Helios	13.4	16.8	66.5	77.9	34.5	3.3	412.8	69.5	61.0	35.5	11.6	2.7
	Region	n.s	**	**	**	**	n.s	*	**	**	**	n.s	***
	Variety	n.s	*	*	**	**	n.s	n.s	***	*	*	n.s	**

*P<0.05, **P<0.01, ***P<0.001, ns P>0.05

2.4.2 Soil properties

We measured five key soil parameters: pH, total N, total C, water content and C:N ratio for the 33 fields we had retrieved yield and grain quality data. Soil water content (P=0.002) and pH (P=0.00064) varied significantly across the regions sampled, with higher soil water contents in Estrie and lower soil water contents in Mauricie and higher pH in Centre-du-Québec and lower pH in Mauricie (Supplementary Table S1). Significant correlations were found between soil pH and grain protein content (r_s=-0.448, P=0.009), Zeleny (r_s=-0.466, P=0.007), gluten (r_s=-0.457, P=0.008), flour BEM (r_s=-0.453, P=0.009) and yield (r_s=0.679, P=0.00001). Similarly, protein (r_s=-0.633, P = 0.00009), starch (r_s=0.597, P = 0.0003), and gluten content (r_s=-0.634, P=0.00009), BEM (r_s=-0.513, P=0.002), CWG (r_s=-0.495, P=0.003) and yield (r_s=0.610, P =0.0002), were significantly correlated with soil water content. We also found significant correlations between yield (r_s=-0.576, P=0.0004) and flour falling number (r_s= 0.515, P=0.002) with soil C:N ratio.

2.4.3 Microbial functions

For community-level carbon utilization pattern (Biolog Eco Plates), we found significant correlations between wheat yield and the utilization N-acetyl glucosamine (r_s=-0.456, P=0.00058), L-threonine (r_s=0.418, P=0.00181), alpha keto-butyric acid (r_s=0.352, P=0.00967), D-glucosaminic acid (r_s=-0.523, P=0.00005) and putrescine (r_s=-0.490, P=0.0019). There was no significant correlation between carbon use and other grain and flour qualities, except for flour ash content which was positively correlated to the utilization of glucose-1-phosphate (r_s=0.443, P=0.00972). For the abundance of N-cycle functional genes measured by qPCR, the only significant correlations were found between the AOB (r_s=0.447, P=0.0591), *nirK* (r_s=0.479, P=0.0072), and *nosZ* (r_s=0.410, P=0.0218) genes and wheat grain moisture content.

2.4.4 Soil microbial community structure, composition, and diversity

The bacterial and archaeal communities were dominated by the *Acidobacteria,* the *Actinobacteria,* and the *Proteobacteria,* which made up about 75% of the whole bacterial and archaeal community (Fig. 2-1). Among the phyla having a mean relative abundance above 1%, the *Planctomycetes* (P= 0.0469) and *Actinobacteria* (P=0.009) varied significantly across the regions. The fungal communities were dominated by the *Agaricomycetes,* the *Mortierellomycotina,* and the *Sordariomycetes* (Fig. 1). The following fungal classes showed significant variation across the regions: *Trellomycetes* (P=0.0027), *Agaricomycete*s (P=1.34 x 10^{-5}), *Pezizomycotina* (P=0.000751), *Mortierellomycotina* (P=0.0313). We also correlated the relative

abundance of all individual archaeal, bacterial, and fungal ASVs (with a mean relative abundance above 1%) with yield and grain quality data. After Benjamini-Hochberg correction for multiple testing, several grain quality parameters such as, falling number, grain moisture content, C:N ratio, flour maximum recorded torque (BEM), flour coarse wheat germ (CWG) showed significant (adjusted P<0.01) negative correlations with A*scomycota, Basidiomycota* and *Zygomycota* ASVs, whereas flour peak maximum time (PMT) and ash content had significant (adjusted P<0.01) positive correlations to fungal ASVs (Table 2-2). On the other hand, the relative abundances of many bacterial ASVs belonging to the *Actinobacteria,* the *Proteobacteria,* and the *Verrucomicrobia* were significantly (adjusted P<0.001) correlated with protein, starch, gluten content and other quality parameter (Table 2-2). Bacterial and archaeal Shannon diversity was significantly and negatively correlated with grain moisture content (r_s= -0.411, P=0.0191). The fungal Shannon diversity was significantly and positively correlated with yield (r_s=0.379, P= 0.0388) and grain starch content (r_s= 0.397, P= 0.0268) and negatively correlated with protein content (r_s= -0.431, P= 0.0154), Zeleny (r_s= -0.431, P=0.0129), gluten content (r_s= -0.455, P=0.0101), CN (r_s= -0.429, P=0.0153), flour maximum recorded torque (BEM) (r_s= -0.388, P= 0.0309), flour coarse wheat germ (CWG) (r_s= -0.454, P= 0.0102).

Table 2-2. **Summary of correlation studies between microbial ASV and grain quality parameters. Significant (adjusted P-value <0.001 for bacteria-archaea and <0.01 for fungi) Spearman correlations between abundant bacterial and fungal ASVs (mean relative abundance above 1%) and grain baking quality (N=33).**

ASVs	Quality parameter	r_s	adj. P-value	ASV taxonomy (phylum; genus)
Bacteria				
14	Protein	0.544	0.0009	*Actinobacteria; Pseudarthrobacter*
14	Starch	-0.630	0.00006	*Actinobacteria; Pseudarthrobacter*
14	Zeleny	0.584	0.0003	*Actinobacteria; Pseudarthrobacter*
14	Gluten	0.581	0.0004	*Actinobacteria; Pseudarthrobacter*
2413	PMT	0.572	0.0003	*Gemmatimonadetes;* Uncult *Gemmatimonadaceae*
2413	BEM	-0.594	0.0002	*Gemmatimonadetes;* Uncult *Gemmatimonadaceae*
3294	PMT	0.572	0.0003	*Firmicutes; Paenibacillus*
3294	BEM	-0.595	0.0002	*Firmicutes; Paenibacillus*
3294	CWG	-0.675	0.000009	*Firmicutes; Paenibacillus*
2190	PMT	0.572	0.0003	*Nitrospirae; Nitrospira*
2190	BEM	-0.594	0.0002	*Nitrospirae; Nitrospira*
2190	CWG	-0.673	0.000009	*Nitrospirae; Nitrospira*
1397	CN	-0.610	0.000099	*Proteobacteria; Nitrosospira*

1397	Falling number	-0.695	0.000004	*Proteobacteria; Nitrosospira*
1397	PMT	0.550	0.0006	*Proteobacteria; Nitrosospira*
1397	BEM	-0.547	0.0007	*Proteobacteria; Nitrosospira*
1397	CWG	-0.614	0.000088	*Proteobacteria; Nitrosospira*
2029	CN	-0.672	0.000010	*Proteobacteria; Sphingomonas*
2029	PMT	0.572	0.0003	*Proteobacteria; Sphingomonas*
2029	BEM	-0.594	0.0002	*Proteobacteria; Sphingomonas*
2029	CWG	-0.674	0.000009	*Proteobacteria; Sphingomonas*
4604	Test Weight	-0.728	0.00000068	*Planctomycetes; Pir4 lineage*
4604	PMT	0.571	0.00033195	*Planctomycetes; Pir4 lineage*
6319	CN	-0.673	0.000010	*Verrucomicrobia; Verrucomicrobium*
6319	PMT	0.572	0.0003	*Verrucomicrobia; Verrucomicrobium*
6319	BEM	-0.595	0.0002	*Verrucomicrobia; Verrucomicrobium*
6319	CWG	-0.674	0.000009	*Verrucomicrobia; Verrucomicrobium*
Fungi				
454	CN	-0.562	0.0005	*Ascomycota; Lecythophora*
454	Falling number	-0.539	0.0010	*Ascomycota; Lecythophora*
454	PMT	0.578	0.0003	*Ascomycota; Lecythophora*
454	BEM	-0.533	0.0012	*Ascomycota; Lecythophora*
454	CWG	-0.563	0.0005	*Ascomycota; Lecythophora*
1412	PMT	0.571	0.0004	*Ascomycota; Archaeorhizomyces*
1412	BEM	-0.593	0.0002	*Ascomycota; Archaeorhizomyces*
1412	CWG	-0.673	0.00001	*Ascomycota; Archaeorhizomyces*
1412	CN	-0.671	0.00001	*Ascomycota; Archaeorhizomyces*
3671	CN	-0.667	0.00002	*Ascomycota; Peziza*
3671	PMT	0.562	0.0005	*Ascomycota; Peziza*
3671	BEM	-0.587	0.0003	*Ascomycota; Peziza*
3671	CWG	-0.669	0.00002	*Ascomycota; Peziza*
80	CN	-0.535	0.0011	*Basidiomycota; Ganoderma*
80	PMT	0.447	0.0080	*Basidiomycota; Ganoderma*
80	BEM	-0.496	0.0029	*Basidiomycota; Ganoderma*
80	CWG	-0.538	0.0010	*Basidiomycota; Ganoderma*
663	Ash	0.592	0.0003	*Basidiomycota; Ceratobasidium*
757	CN	-0.694	0.000005	*Basidiomycota; Psathyrella*
757	PMT	0.696	0.000005	*Basidiomycota; Psathyrella*
757	BEM	-0.673	0.000013	*Basidiomycota; Psathyrella*
757	CWG	-0.696	0.000005	*Basidiomycota; Psathyrella*

566	Humidity	-0.442	0.0099	*Zygomycota; Mortierella*

PMT=Flour Peak Maximum Time, BEM=Flour Maximum Torque, CWG= Coarse Wheat Germ, CN= Carbon/Nitrogen ratio.

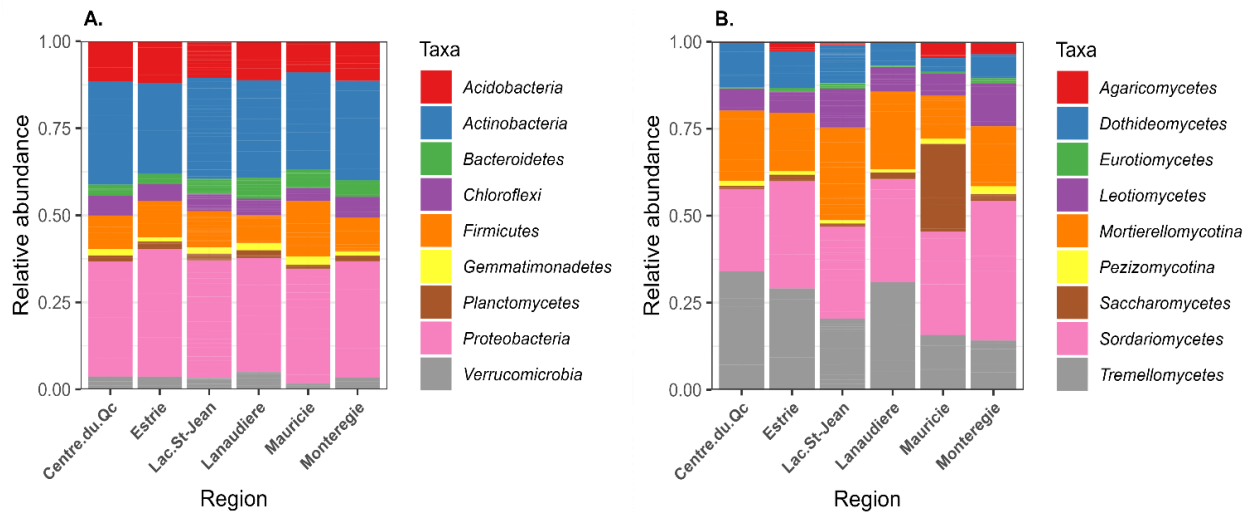


Figure 2-1: Summary of bacterial and archaeal, and fungal community composition. Bacterial and archaeal (A) and fungal (B) community composition for the phyla (A) or classes (B) having a mean relative abundance above 1%, based on Illumina amplicon sequencing of the 16S rRNA gene (A) or the ITS region (B). Values are averaged across the regions. For a clearer understanding of how fungal community composition is affected across provinces, fungal ASVs are displayed at the class level rather than the phylum level.

2.4.5 Predictive modeling of wheat grain and flour quality

We have limited our modeling efforts to yield, grain protein and gluten content and flour PMT and BEM, which are arguably the best indicators for grain and flour that produces high quality bread. Microbes-only multiple regression analysis was performed using the following explanatory variables: fungi: bacteria (F:B) ratio (calculated from the ratio of the qPCR quantifications of the ITS region and the 16S rRNA gene), fungal and bacterial-archaeal diversity indices (Shannon, Chao1, Simpson and observed ASVs), the color development for the top ten Biolog substrates (highest correlations with quality and yield data, irrespective of significance, calculated separately for each of the dependent variables), the abundance of the four N-cycle functional genes, the relative abundances of the top 10 fungal and top 10 bacterial-archaeal ASVs (highest correlations with quality and yield data, irrespective of significance, calculated separately for each of the dependent variables), and bacterial-archaeal and fungal PCoA axes 1 and 2. The soil-only multiple regression analysis used pH, total N, total C, water content and C: N ratio as explanatory variables. First, we wanted to compare the performance of soil vs. microbial variables to explain yield and grain quality. As the explanatory power of multiple regression models generally increases with the number of explanatory variables included in the model, we limited the number of variables to be selected for the microbes-only

model to 5 to match the number of variables available for the soil-only model. The following regression equations were generated for the microbes-only analyses:

Yield = 5.07+10210·**Fun757** -2.35· **γ-amino butyric acid** + 0. 013.fungal $\mathbf{S_{obs}}$-3.24· **bacterial PCoA axis1** -2.97×$10^{-5.}$ **AOB……(1)**

Protein = 13.11+7.62·**fungal PCoA axis2**– 2244·**Bact3294** +4.12·**γ-amino butyric acid** + 5.04×10^{-5} ·**AOB** – 0.031·**fungal** $\mathbf{S_{obs}}$ **…….(2)**

Gluten = 33.98+ 18.56·**fungal PCoA axis2** -2428·**Fun454** -5719·**Bact3294** -2.48·**fungal Shannon** + 8.04·**γ-amino butyric acid……(3)**

PMT = 89.2-30.19·**bacterial PCoA axis1**+ 26.99· **α-keto butyric acid** -39.20· **L-Threonine** - 0.00077·**AOB** -12470·**Bact2029……. (4)**

BEM = 47.29+9023·**Bact6319**+ 14.54·**glucose-1-phosphate** +0.00054·**AOB** -59030·**Fun757** - 0.065·**fungal** $\mathbf{S_{obs}}$**…… (5)**

For the soil-only model, we obtained the following regression equations:

Yield = -4.98+18.24· **total N** -1.27· **total C**-0.0075· **C:N ratio** +0.77· **water content** + 1.13·**pH…….(1)**

Protein= 35.74-45.31· **total N** +3.63· **total C** -0.70·**C:N ratio** -2.42· **water content** - 1.54·**pH…….. (2)**

Gluten= 82.82-119.02·**total N** +9.43·**total C** -1.88·**C:N ratio** -5.44· **water content** - 3.76·**pH…….. (3)**

PMT= 24.33-6.10·**total N** -2.28·**total C** +0.91·**C:N ratio** +5.37· **water content** +6.64· **pH…….(4)**

BEM= 132.99-203.62·**total N** +15.82·**total C** -3.39·**C:N ratio** -6.53· **water content** -4.97· **pH…….(5)**

In all cases, with the same number of explanatory variables, the microbes-only models outperformed the soil-only models, with higher R^2 values and lower Akaike Information Criterion (AIC) and Residual Standard Error (RSE) (Table 3). In the soil-only models, total N, total C and C:N ratio were highly collinear with Variation Inflation Factors (VIF) well above 5 (Table 2-4), which could partly explain their lower performance as compared to the microbes-only models. The variables selected in the microbes-only models did not show any evidence of collinearity, with VIF well below 5 in all cases (Table 2-4). Since we had much more than 5 potential microbial explanatory variables, we re-ran the analyses, including all the microbial and soil variables listed above, and let the stepwise procedure proceed until all significant variables were included in the model. Between 9 and 11 variables were included. This resulted in the following regression equations:

Yield = -0.916+1.04· **pH** +6526· **Fun757** +0.73· **water content** -0.25·**C:N ratio** + 215.6· **Bact2413** +87180·**Fun3671** -1.75· **L-threonine** -315.7· **Bact2029** -2.08· **fungal PCoA axis2**+1.45· **fungal PCoA axis1……..(1)**

Protein =22.14+6.80· **fungal PCoA axis2**-2540·**Bact3294**-2.15· **pH** -422.9· **Fun80** + 11.56·**L-threonine** -4.20· **α-keto butyric acid** +0.83· **F:B ratio** +0.000071· **AOB** +1.78· **L-asparagine** +1114· **Bact4604……..(2)**

Gluten=64.59+17.61· **fungal PCoA axis2** -6450·**Bact3294** -5.56· **pH**-89.42· **Fun80** +21.84· **L-Threonine** -10.51· **α- keto butyric acid** +2.11· **F:B ratio**-18.74· **total N**+0.0001193· **AOB……(3)**

PMT= -93. 32+26770·**Bact3294** +21.59· **pH** – 199500·**Fun757** -8377·**Bact1397** + 0.0001323·**AOA** -46.31· **bacterial PCoA axis1** -650500·**Fun1412** + 64.05· **fungal PCoA axis1** -73.31· **fungal PCoA axis2** + 9.52· **water content** + 9.41·**L-arginine……..(4)**

BEM = 86.7-11.23· **pH** -11125·**Bact3294** -8.01· **water content** -3471·**Bact14** + 2.94·**C: N ratio** -29.15· **fungal PCoA axis1** + 22.62· **L-threonine** -796·**Fun80** + 9.41· **D-glucosaminic acid** + 2.43·**F: B ratio……...(5)**

Table 2-3: Evaluation of model based on different statistical parameters. Akaike information criterion (AIC), Residual standard error (RSE), adjusted R^2 and P-value for the soil, microbial and soil + microbial models.

	Soil	Microbial	Soil+Microbial
Yield			
Nb. variables	5	5	10
AIC	-5.45	-19.24	-46.63
RSE	0.84	0.65	0.38
Adjusted R^2	0.59	0.72	0.90
P-value	2.36×10^{-5}	9.01×10^{-7}	7.98×10^{-9}
Protein			
Nb. variables	5	5	10
AIC	53.35	35.83	3.71
RSE	2.11	1.69	0.92
Adjusted R^2	0.32	0.55	0.87
P-value	0.00915	1.96×10^{-4}	1.40×10^{-7}
Gluten			
Nb. variables	5	5	9
AIC	106.48	83.95	60.51
RSE	4.85	3.88	2.48
Adjusted R^2	0.33	0.56	0.82
P-value	0.00716	1.72×10^{-4}	7.34×10^{-7}
PMT			
Nb. variables	5	5	11
AIC	193.95	168.95	140.43
RSE	20.2	16.81	9.72
Adjusted R^2	-0.063	0.23	0.74
P-value	0.677	0.0481	7.21×10^{-5}
BEM			
Nb. variables	5	5	10
AIC	140.27	123.90	107.73
RSE	8.23	7.73	5.56
Adjusted R^2	0.19	0.31	0.64
P-value	0.0617	0.0163	5.02×10^{-4}

PMT=Flour Peak Maximum Time, BEM=Flour Maximum Torque, AIC: Akaike Information Criterion, RSE: Residual Standard Error.

Table 2-4: Model evaluation for bias-variance and multicollinearity among input variables. Variation inflation factor (VIF) for the variables included in the soil, microbial and soil + microbial models.

Yield		Protein		Gluten		PMT		BEM	
Variable	VIF	Variable	VIF	Variable	VIF	Variable	VIF	Variable	VIF
Soil									
tot. N	74.05	tot. N	74.05	tot. N	74.05	tot. N	74.05	tot. N	74.05
tot. C	105.45	tot. C	105.45	tot. C	105.45	tot. C	105.45	tot. C	105.45
C/N	13.87	C/N	13.87	C/N	13.87	C/N	13.87	C/N	13.87
water	1.09	water	1.09	water	1.09	water	1.09	water	1.09
pH	1.20	pH	1.20	pH	1.20	pH	1.20	pH	1.20
Microbial									
Fun757	1.17	fun axis2	1.34	fun axis2	1.38	bact axis1	1.08	Bact6319	1.11
bacterial axis1	1.05	Bact3294	1.30	Fun454	1.15	AKBA	1.76	G1P	1.64
GABA	1.13	GABA	1.51	Bact3294	1.47	L-threonine	1.81	AOB	1.47
AOB	1.10	AOB	1.13	fun Shannon	1.70	AOB	1.17	Fun757	1.21
fun S_{obs}	1.18	fun S_{obs}	1.37	GABA	1.35	Bact2029	1.15	fun S_{obs}	1.27
Microbial + soil									
ph	2.12	fun axis2	1.69	fun axis2	1.55	Bact3294	1.58	pH	1.93
Fun757	1.61	Bact3294	1.25	Bact3294	1.27	pH	2.55	Bact3294	1.24
water	1.15	pH	1.87	pH	1.89	Fun757	2.19	water	1.37
C.N.ratio	2.27	Fun80	1.47	Fun80	1.41	Bact1397	1.15	Bact14	1.52
Bact2413	1.18	L-threonine	3.67	L-threonine	2.73	AOA	1.20	C:N.ratio	2.23
Fun3671	2.24	AKBA	2.17	AKBA	2.38	Bact axis1	1.42	fun axis1	2.12
L-Threonine	2.50	F:B.ratio	1.33	F:B.ratio	1.35	Fun1412	1.91	L-threonine	2.15
Bact2029	1.20	AOB	1.56	total.N	1.35	fun axis1	2.63	Fun80	1.35
fun axis2	1.73	L-asparagine	2.01	AOB	1.40	fun axis2	1.86	DGA	1.99
fun axis1	2.48	Bact4604	1.10			water	1.23	F:B.ratio	1.23
						L-arginine	1.34		

Values are identical for the soil models because the exact same variables were used in the five models.

GABA: Gamma-Aminobutyric Acid

AKBA: Alpha-ketobutyric acid

DGA: D-Glucosaminic Acid

G1P: Glucose-1-Phosphate

AOB: bacterial ammonia monooxygenase subunit A

AOA: archaeal ammonia monooxygenase subunit A

The predictive power of the models is depicted in Figure 2-2, where the predicted values for wheat yield and grain baking quality are plotted against the observed values. All the models were highly significant and had lower AIC and RSE and higher R^2 than the soil-only and the microbes-only models (Table 2-3). R^2 varied from 64% (BEM) to 90% (Yield), with P-values well below 0.001. The variables selected in the models did not show any evidence of collinearity, with VIF well below 5 in all cases (Table 4). The taxonomic affiliations of the ASVs selected in the models are as follows: Fun80*:* *Ganoderma* (Basidiomycota), Fun454: *Lecytophora* (Ascomycota), Fun757*:* *Psathyrella* (Basidiomycota), Fun1412: *Archaeorhizomyces* (Ascomycota), Fun3671*:* *Pezizomycetes* (Ascomycota), Bact14*:* *Pseudarthrobacter* (Actinobacteria), Bact1397: *Nitrosospira* (Proteobacteria), Bact2029*:* *Sphingomonas* (Proteobacteria), Bact2413: uncultured *Gemmatimonadaceae* (Gemmatimonadetes), Bact3294: *Paenibacillus* (Firmicutes), Bact4604*:* Pir4 lineage (Planctomycetes), Bact6319: *Verrucomicrobium* (Verrucomicrobia).

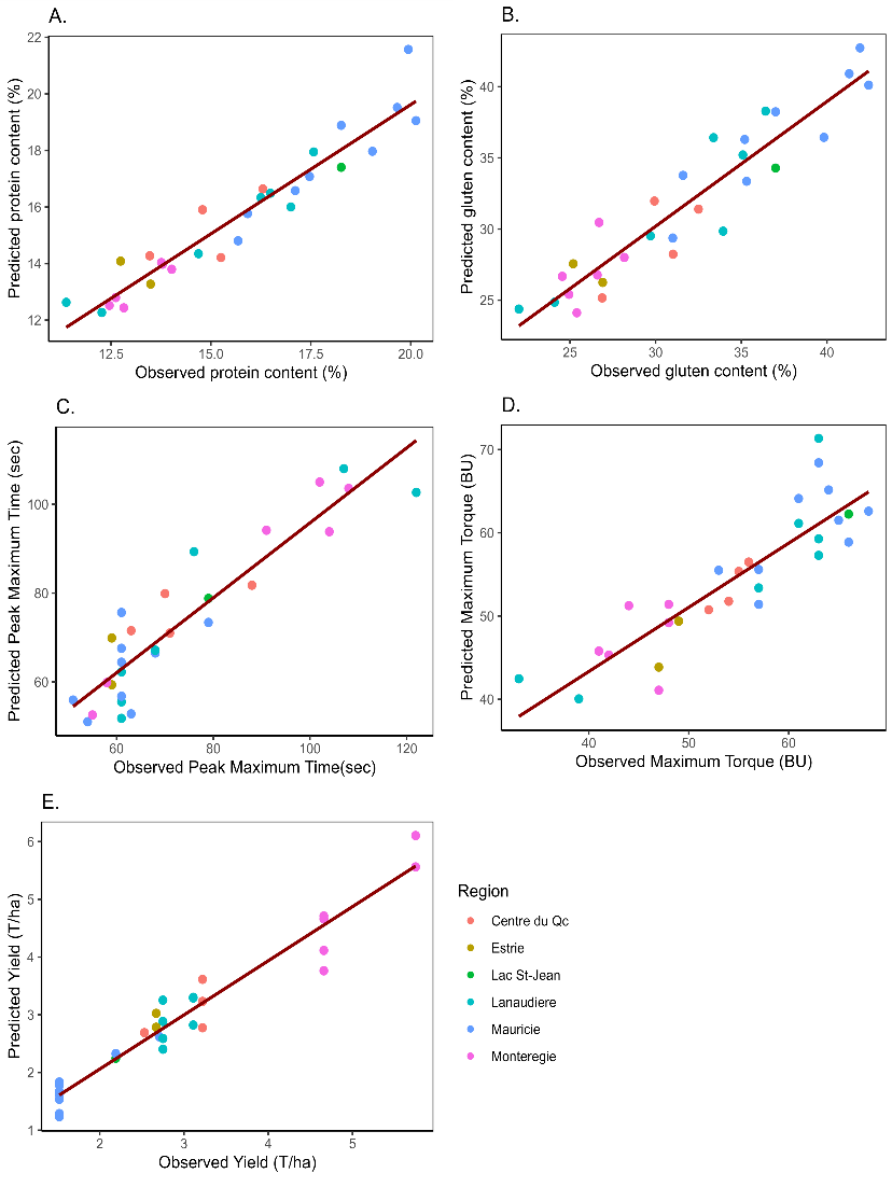


Figure 2-2: Soil and microbial-based multiple linear regression models. Observed vs. predicted yields, grain protein and gluten content, flour maximum peak time (PMT), flour recorded maximum torque (BEM) based on the soil + microbial models presented in the result section (N=33). Red lines go through the origin and have a slope of 1.

2.5 Discussion

We had hypothesized that microbial parameters measured early in the growing season would be able to better predict wheat yield and grain and flour baking quality than soil parameters measured at the same time. When using the exact same number of explanatory variables, microbes-only regression models were better able to predict wheat yield and key baking quality indicators such as grain protein and gluten content and flour peak maximum time (PMT) and maximum torque (BEM) than soil-only models. In fact, the soil parameters failed to produce a significant model to predict PMT and BEM. When combining microbial and soil parameters, we were able to generate models with R^2 of 64 to 90%. Additionally, our study confirmed the predictive power of microbial indicators for soil processes at a scale of ~500 km. Similarly, a recent continental-scale study showed that it was possible to use machine-learning to predict soil health based on 16S rRNA gene dataset with a R^2 around 80% (Wilhelm *et al.* 2022). Previous studies that had tried to predict processes based on microbial indicators had focused on a few neighbouring fields (Yergeau *et al.* 2020), lab-incubated soils (Bell *et al.* 2015), or greenhouse grown plants (Yergeau *et al.* 2015). Here, we sampled fields across the wheat growing regions of the province of Québec without limiting our efforts to a certain type of soil, agricultural management, or variety of wheat, and we were still able to produce highly significant regression models that predicted very accurately yields and grain and flour quality.

Although some of the soil parameters measured here, such as pH, water content and C:N ratio were previously reported to have a determining effect on soil microbial communities (Wan *et al.* 2015, Zhalnina *et al.* 2015), their linkages with soil processes, i.e., the activities carried out by the soil microbial communities, are not necessarily straightforward (Sánchez *et al.* 2021). For instance, nutrient availability and its depletion depend on the soil moisture content (Marschner & Rengel 2012), but the equilibrium of soil nutrients and soil C and N status is regulated by microbial activity (Paz-Ferreiro & Fu 2016). We found significant correlations between soil pH and water content and some grain quality indicators, but since these parameters cannot consider directly the microbial factors involved in the transformation of soil inorganic and organic N, which impacts plant N use efficiency, grain and flour baking quality could not be accurately predicted. Although this might vary through time, we showed here that quality parameters were negatively linked to soil total N measured at the beginning of the season. This put in question the practice of indiscriminately fertilizing at high N levels (Vanasse 2012), as, depending on the soil microbiology, good quality grain could be obtained without adding fertilizer, as previously reported (Yergeau *et al.* 2020). A recent study showed that current year cranberry yields could be predicted with an accuracy of 83% from location, cultivars, climatic indices, fertilization, and plant tissue nutrient content and soil chemical

characteristics (Parent *et al.* 2021). Similarly, by extending the number of soil parameters measured, we could probably have increased the accuracy of the soil models, but it would have become rapidly cost- and/or labor-prohibitive. Additionally, in view of the lack of strong linkages between many soil parameters and soil processes, it is doubtful if this increase in the number of parameters would have led to soil-based models that would have outperformed microbial-based models.

Microbial parameters such as the abundance of functional genes, the capacity to degrade substrates, or general community descriptors (i.e., diversity indices or PCoA axes) can be more directly related to soil processes. This is probably why we showed here, in agreement with our hypothesis, that regression models based on microbial indicators were able to predict more accurately wheat yield and grain and flour quality than models based on soil indicators. In fact, soil physico-chemical parameters were at best able to predict the grain quality with an accuracy of 33% and failed to yield a significant model for flour quality, whereas using the same number of variables, microbial models were able to predict wheat grain quality with an accuracy of up to 56%, and flour quality with an accuracy of up to 31%. Another strength of the microbial approach is that once the analyses are completed, several hundreds of parameters are available and can be added to the models to improve accuracy. In our case, using still parsimonious models of less than 11 parameters, we were able to increase accuracy up to 87% for grain quality and up to 74% for flour quality. Of course, to have practical applications for farmers and millers, the microbial parameters highlighted would have to be measured using more rapid and inexpensive methods such as qPCR for the selected ASVs and specific substrate degradation assays for the Biolog indicators. Shallow amplicon sequencing could also be useful to determine general community parameters, such as alpha and beta diversity, rapidly and inexpensively.

Many of the microbial parameters selected in our regression models can be linked to important soil processes that could have impacted wheat N nutrition, and consequently grain and flour baking quality. For instance, there were negative relationships between the abundance of the archaeal *amoA*, the relative abundances of an ammonia oxidizer ASV (*Nitrosospira*) and of a nitrite-oxidizer or comammox ASV (*Nitrospira*) with flour and grain quality. Ammonia can be taken up passively by plants and directly assimilated into amino acids, whereas nitrate needs to be taken up actively and reduced to ammonia before being assimilated into amino acids. This makes ammonia more energy-efficient for plants (Moreau 2019), especially for high N demanding crops such as bread wheat. In that case, ammonia-oxidizers that perform the first rate-limiting step of nitrification are generally thought to have negative effects on plant N nutrition. Similarly, our previous study that focused on two fields in the southern part of the province of Quebec had reported that the abundance of the archaeal *amoA* gene was negatively linked to grain and flour quality

(Yergeau *et al.* 2020). In contrast to the archaeal version of the *amoA* gene, we found that the abundance of the bacterial *amoA* was positively linked to grain protein and gluten content. These contrasting results could be due to a disconnection between AOB abundance and process rates (Hu *et al.* 2015) or to the dominance of AOA in nitrification processes of terrestrial ecosystems (Adair & Schwartz 2011, Gubry-Rangin *et al.* 2010, Leininger *et al.* 2006). The total abundance of the AOB could also not be directly relevant for soil processes, with AOA: AOB or AOB:total bacteria ratios being more important.

Some other selected parameters, such as the utilization of specific substrates, could potentially have a functional significance. For instance, some of the selected substrates such as L-Threonine and L-Asparagine are amino acids, and the capacity to degrade them efficiently could indicate a community that is better able to access to the N stored in soil organic matter, which could also improve soil N availability (Jones & Kielland 2002, Ukalska-Jaruga *et al.* 2020) and consequently plant N nutrition. Similarly, the capacity to degrade efficiently glucose-1-phosphate could be linked to the efficient degradation of glycogen or starch and potentially a heightened capacity to degrade soil organic matter. It has also been reported that the activity of microbes using glucose-1 phosphate as a carbon source impacted the diversity of the rhizosphere and bulk soil microbial communities (*Hills et al.* 2020). Another interesting parameter that was often selected and had positive relationships with quality is the fungal:bacterial ratio. A higher F:B ratio could mean that more soil organic matter is degraded by fungi, resulting in more N being released for plant uptake since fungi generally have a lower requirement for N per unit of biomass. This could also be the reason behind the selection of fungal PCoA axes, diversity indices and ASVs in many of the models. Some of the bacterial ASVs singled out by regression and correlation analyses belonged to genera, such as *Paenibacillus* and *Sphingomonas*, that contain known plant-growth promoting rhizobacteria (PGPR) (Castanhera *et al*. 2017; Liu *et al.* 2019). However, these relationships were mostly negative for grain and flour quality, which could indicate that, at constant N inputs, N is diluted in larger plants, which lowers grain quality. Other parameters could not be directly linked to processes related to wheat N nutrition. Our goal was not to produce a model that would explain wheat N nutrition, but to highlight microbial predictors for high quality grain and flour, and as such, some of the indicators selected could be co-varying with other unmeasured factors or simply have a niche optimum that is also conducive for optimal wheat nutrition. Even though they might have no functional significance, these indicators are still useful for increasing the accuracy of the predictive models and could be used as indicators to inform management practices.

Several differences were found between our previous effort to model grain quality in two wheat fields (Yergeau *et al.* 2020) and the current study. For instance, Yergeau *et al.* (2020) reported that the abundance of the copper-containing nitrite reductase (*nirK*) was a significant variable explaining wheat

grain quality, which was not the case here. The ASVs selected by the models were also different. These discrepancies could be explained by the fact that the Yergeau *et al.* (2020) study sampled two wheat varieties in two fields that were under various inorganic nitrogen fertilization regimes, whereas we sampled 80 fields across the province of Quebec growing seven varieties of wheat in various soil types, and under a range of fertilization, management, and environmental conditions, which have reduced the microbial indicators to the ones that are significant across these conditions. Alternatively, year to year differences in environmental conditions might have resulted in different microbial indicators being identified as significant in the two studies. It would therefore be crucial to test the stability in time of the indicators found here through a multi-year study. A robust indicator would be significant year after year across the whole province.

Interestingly, microbial parameters obtained early in the growing season showed strong linkages with grain quality at the end of the growing season, confirming previous results for wheat yield and grain quality (Yergeau *et al.* 2020) and for willows rhizo-remediation capacity (Yergeau *et al.* 2015, Bell *et al.* 2015). In fact, our best models were able to explain 64 to 90% of the variability in yields and quality from microbial and soil indicators derived from bulk soil sampled months earlier. It would be interesting to find the best sampling period to optimize the predictive power of our approach. However, the window of intervention to steer the soil communities will be smaller if the indicators are measured later in the season. Thus, it appears that the initial measurement of microbial diversity, functional genes and potential to use various carbon substrates could help farmers make the best management decisions. The key question is now: how can we use this information to modulate the microbial indicators identified to improve wheat baking quality? Many studies have suggested approaches to engineer or manipulate complex microbial communities (Agoussar & Yergeau 2021, Calderón *et al.* 2017, Sheth *et al.* 2016, Quiza *et al.* 2015) but the field is still in its infancy. Agricultural management practices do modify the soil and plant microbial communities (Babin *et al.* 2019) and could be used to steer the communities toward the desired state. However, more controlled studies would be needed to confirm if this is something feasible at a large scale. Identifying robust and highly accurate microbial indicators is the first step toward a better management of crop production system, to increase produce quality while reducing inputs, on the road to a more sustainable agriculture.

2.6 Acknowledgments

The entire staff of Les Moulins de Soulanges and La Meunerie La Milanaise, and more specifically Élisabeth Vachon, Stéphanie Carrière, Chafik Baghdadi and Robert Beauchemin, are gratefully acknowledged for their support of this study.

2.7 Funding

This study was supported by an FRQNT Team Grant (2019-PR-254256) and a Compute Canada Resource allocation (allocation 2020-3177) on the Graham system (University of Waterloo) to Etienne Yergeau.

2.8 Conflict of interest

The authors have no conflict of interest to declare.

2.9 References

All the corresponding references of this chapter have been included at the end of the thesis.

Chapter 3: Early season soil microbiome best predicts wheat grain quality

Le microbiome du sol en début de saison est le meilleur prédicteur de la qualité du grain de blé

Numan Ibne Asad[1], Xiao-Bo Wang[2], Jessica Dozois[1], Hamed Azarbad[3], Philippe Constant[1], Etienne Yergeau[1]

Authors:

[1]Institut national de la recherche scientifique, Centre Armand-Frappier Santé Biotechnologie, Laval, QC H7V 1B7, Canada

[2]State Key Laboratory of Grassland Agroecosystems, Center for Grassland Microbiome and College of Pastoral, Agriculture Science and Technology, Lanzhou University, Lanzhou 730020, People's Republic of China

[3]Philipps-University Marburg, Department of Biology, Evolutionary Ecology of Plants, Marburg, Germany

Title of journal or book: FEMS Microbiology Ecology, Volume 99, Issue 1, January,2023
Published online: 24 November 2022
DOI: https://doi.org/10.1093/femsec/fiac144

Contributions of the authors:

Numan Ibne Asad: Contributed to the main experimental design of the research, conceptualization, methodology, laboratory experiments, data processing and analysis, all statistical analysis and predictive modelling, original draft writing and revisions.
Xiao-Bo Wang: Contributed to the analysis of bioinformatic data.
Jessica Dozois: Contributed to the soil sampling, and the analysis of soil biochemistry.
Hamed Azarbad: Contributed to the experimental field set up.
Professor Philippe Constant: Conceptualization, methodology, supervision, reviewing, resources, and infrastructure.

Professor Etienne Yergeau: Conceptualization, methodology, supervision, writing and editing of the original draft, resources, and infrastructure.

3.1 Abstract

Previous studies have shown that it is possible to accurately predict wheat grain quality and yields using microbial indicators. However, it is uncertain what the best timing for sampling is. For optimal usefulness of this modeling approach, microbial indicators from samples taken early in the season should have the best predictive power. Here, we sampled a field every two weeks across a single growing season and measured a wide array of microbial parameters (amplicon sequencing, abundance of N-cycle related functional genes, and microbial carbon usage) to find the moment when the microbial predictive power for wheat grain baking quality is highest. We found that the highest predictive power for wheat grain quality was for microbial data derived from samples taken early in the season (May–June) which coincides roughly with the seedling and tillering growth stages, that are important for wheat N nutrition. Our models based on LASSO regression also highlighted a set of microbial parameters highly coherent with our previous surveys, including alpha- and beta-diversity indices and N-cycle genes. Taken together, our results suggest that measuring microbial parameters early in the wheat growing season could help farmers better predict wheat grain quality.

Keywords: wheat microbiome; LASSO regression; grain quality; amplicon sequencing; nitrogen cycle; community level physiological profiling

3.2 Introduction

Integrated microbiocentric approaches to optimize plant production are promising and have often been proposed to solve some of the many problems agricultural production faces (Figuerola *et al*. 2012; Schloter *et al*.2018). Soil microorganisms play a key role in many ecosystem processes that are central to agricultural production. For instance, soil microorganisms recycle organic matter, cycle nutrients, abate abiotic stresses, change soil structure and porosity, and promote plant growth (Ortiz & Sansinenea 2022). However, although it is theoretically known how to modify microbial communities (Agoussar & Yergeau 2021), it is in practice still a very daunting task because of the complexity of the communities and their interactions. A first step towards this goal would be to create microbial-based models predicting agricultural processes, to identify clear targets and key functions or taxa to manipulate.

However, soil microbial communities are very dynamic, which makes it difficult to predict process rates and to identify key players that would be amenable to manipulation. Soil microbial communities are strongly influenced by biotic and abiotic factors, such as temperature, precipitations, and plant growth stage, which all vary in time, often in an unpredictable manner. We recently showed that dry-rewetting cycles lead to a complete overhaul of the soil microbial communities, much more than small decreases in soil water content (Wang *et al*. 2022.). Soybean and wheat growth stages were shown to profoundly influence the microbial diversity associated with the plant, often in interaction with plant compartment, plant genotype, soil water content and soil history (Moroenyane *et al*. 2021; Azarbad *et al.* 2022; Azarbad *et al.* 2020). Similarly, the effect of the genotype on root and rhizosphere microbial communities varied over time (years) and with wheat growth stages (Quiza *et al*. 2022). These microbial shifts related to plant growth stages were previously linked to changes in the composition and concentration of plant root exudates during development (Chaparro *et al*. 2013). The timing of sampling is thus expected to influence the predictive power microbial parameters, but it is still uncertain what the best sampling time would be and whether robust time-independent indicators could be identified.

Recent microbial-based modeling from our group showed that early sampling of wheat field soil microbial communities, around seeding or emergence could accurately predict wheat yield and grain baking quality obtained at the end of the growing season (Asad *et al*. 2021; Yergeau *et al*. 2020) . For instance, with as little as 5 predictors, such as the abundance of archaeal ammonia-oxidizers, measured shortly after seeding in May, we were able to predict wheat grain quality with an accuracy of up to 81% (Yergeau *et al*. 2020). In contrast, different ammonium nitrate fertilization regimes did not significantly influence yields or grain baking quality. In another study encompassing 80 fields across a transect of 500km, microbial indicators from samples taken in May-June could robustly predict the wheat grain quality and yields at the end of the growing season (Asad *et al.* 2021). In line with this, earlier work showed that the growth of

willows after 100 days in highly contaminated soil could be predicted by the initial soil microbial diversity (Yergeau *et al*., 2015), whereas willows Zn accumulation after 16 months of growth could be predicted by the relative abundance of specific fungal taxa present at 4 months (Bell *et al.* 2015). Therefore, it seems that the early soil microbial data can accurately predict ecosystem processes, such as plant productivity and produce quality. However, these studies did not compare microbial data taken at different timepoints, so it is unclear if early sampling has the highest predictive power in microbial-based models.

Here, we sampled the same experimental field every two weeks over the course of a single growing season. We sequenced the bacterial and archaeal 16S rRNA gene and the fungal ITS 1 region, quantified the abundance of key N-cycle genes and measured the community level physiological profiles as microbial indicators and linked them to grain baking quality using LASSO regression. Our goals were to 1) identify the most appropriate sampling date for modelling, and 2) identify robust microbial indicators linked to grain baking quality.

3.3 Methods

3.3.1 Experimental design and sampling

We aimed at collecting samples from a single site for which we knew that the microbial communities varied through time and across treatments. For that purpose, we sampled an ongoing multi-year field experiment on our campus that looked at the effect of rainfall manipulation and wheat genotype on the transmission of the microbiota. We had previously determined that the microbial communities varied through time and across the treatments (Wang *et al.* 2022). The experiment comprised four rainfall manipulation treatments that were set-up in 2016 at the Armand-Frappier Sante Biotechnologie Centre (Laval, Québec, Canada) using 2m x 2m rain-out shelters that excluded passively 0%, 25%, 50%, and 75% of the natural precipitation. The rainfall exclusion treatments were performed using rain-out shelters, which were covered with various amount of transparent plastic sheeting. The rain was intercepted by the plastic sheeting and guided in a gutter and downspout and collected in 20L buckets that were manually emptied following significant rainfall events. Two wheat genotypes were seeded under these shelters (drought sensitive, *Triticum aestivum* cv. AC Nass and drought tolerant, *Triticum aestivum cv. AC Barrie*), and the experiment was replicated over 6 fully randomized blocks, resulting in 48 plots (4 treatments x 2 genotypes x 6 blocks). Seeds harvested from each of the plots were re-seeded in the exact same plot the following year. Soil was sampled every 2 weeks on May 10th (seeding time, T = 0), May 24th, June 7th, June 21st, July 5^{th,} July 19th, and August 1st 2018. A composite soil sample was prepared by collecting 10-cm deep soil cores from the 4 corners and the centre of each plot (4 treatments x 6 blocks x 2 cultivars x 7 sampling dates = total 336 samples). From 2016 to 2018, the average daily rainfall recorded on this site was 2.2 mm-3.5 mm. Soil water content within rainfall exclusion treatments showed significant differences among soil sampling dates (Wang *et al*. 2022).

3.3.2 Amplicon sequencing and data analysis

Total genomic DNA was extracted from the 336 soil samples with the DNeasy PowerLyzer Power Soil Kit (Qiagen) following the manufacturer's instructions. The concentration and the quality of the DNA was checked using a Nano Drop ND-1000 Spectrophotometer (Nano Drop Technologies Inc., Thermo Scientific, U.S.A.). The amplicon sequencing libraries for the bacteria and archaeal 16S rRNA gene and ITS regions were prepared according to the previously described protocols (Asad *et al*. 2021). The primers pairs used for the amplification were 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') (Caporaso *et al.* 2012) and ITS1F (5'-CTTGGTCATTTAGAGGAAGTAA-3') and 58A2R (5'-TACGGYTACCTTGTTACGACTTT-3') (Martin & Rygiewicz, 2005), for the bacterial and archaeal 16S rRNA gene and the fungal ITS 1 region, respectively. PCR amplifications were conducted in a T100™ Thermal Cycler (Bio-Rad, U.S.A.) as

previously described (Wang *et al*. 2022). PCR products were confirmed through visualization in 1% agarose gel and purified using AMPure XP beads (Beckman Coulter, Indianapolis, U.S.A.). PCR libraries were pooled together and sent to the Centre d'expertise et de services Genome Québec (Montréal, Canada) for Illumina MiSeq 2 x 250 bp amplicon sequencing as detailed previously (Wang *et al*. 2022). A total of 17,084,986 16S rRNA gene reads and 22,411,001 ITS 1 region reads were produced. The raw sequencing data and its meta data were deposited in the NCBI BioProject under accession PRJNA686206.

Sequence pre-processing, including filtering and quality testing, was performed using UCHIME (Edgar *et al*. 2011), following previously published bioinformatic pipelines (Wang *et al*. 2022). The classification of Operational Taxonomic Units (OTUs) was performed using the RDP 16S rRNA Reference Database (Wang *et al*. 2007) and the UNITE ITS Reference Database (Nilsson *et al.* 2019). The uniformity of the amplicon sequences belonging to the same operational taxonomic units (OTUs) was tested using UPARSE (Edgar *et al.* 2013). Sample rarefaction was performed using an in-house galaxy pipeline as previously discussed (Wang *et al*. 2022.). Alpha (e.g., Shannon, Simpson, Chao1, Abundance-based Coverage Estimators), beta (Bray-Curtis dissimilarity) and phylogenetic diversity were calculated as detailed in Wang *et al* (2022).

3.3.3 Quantitative real-time PCR (qPCR) and community level physiological profiling (CLPP)

We measured the abundance of the 16S rRNA gene, the ITS 1 region, and N-cycle related genes (bacterial and archaeal *amoA*, *nirK*,and *nosZ*) for the 336 samples using real-time PCR SYBR Green assays, as previously described (Asad *et al.* 2021). The abundance of N-cycle related gene copies was measured using primers amoA1-f* (5'-GGGGHTTYTACTGGTGGT-3') and amoA2-r (5'-CCCCTCKGSAAAGCCTTCTTC-3') (Levy-Booth, Prescott and Grayston 2014), the archaeal *amoA*, using primers crenamoA23-f (5'-ATGGTCTGGCTWAGACG-3') and crenamoA616-r (5'-GCCATCCATCTGTA-3') (Tourna *et al.* 2008), the copper-containing nitrite reductase gene (*nirK*), using primers 876f (5'- ATYGGCGGVCAYGGCGA-3') and 1040r (5'-GCCTCGATCAGRTTRTGGTT-3') (Henry *et al.* 2006), the nitrous oxide reductase gene (*nosZ*) using primers nosZ1f (5'-WCSYTGTTCMTCGACAGCCAG-3') and nosZ1r (5'-ATGTCGATCARCTGVKCRTTYTC-3') (Henry *et al.* 2006). The abundance of the 16S rRNA gene and of the ITS 1 region was measured using the same primers as for the amplicon sequencing (described above). The Fungal: Bacterial (F:B) ratio was then calculated by dividing the ITS 1 region abundance by the 16S rRNA gene abundance. Community level physiological profiling (CLPP) was performed using Eco Plates colorimetric assays (Biolog, Hayward, CA) with diluted soil (1/10 in water) and a 168-hour incubation, as previously described (Asad *et al.* 2021).

3.3.4 Wheat grain and flour quality

Wheat grain was harvested from the 48 plots at the end of the growing season (August 8th, 2018) and the grain and flour baking quality were analyzed in the quality control laboratory of Les Moulins de Soulanges (St-Polycarpe, QC). Four main quality indicators were used in our modeling efforts: grain protein content, grain gluten content, flour peak maximum time (PMT), and flour maximum recorded torque (BEM) (Freund and Kim 2006). PMT and BEM were measured with a GlutoPeak instrument (Brabender, Duisburg, Germany). To do so, the flour sample is mixed with water and stirred at constant speed while the instrument records the torque used to move the mixing paddle. As the gluten network forms, the torque increases until a maximum value, after which it decreases as the gluten network is destroyed by excessive mixing. The time it takes to reach the peak is the PMT (in seconds) and the height of the peak is the BEM (in Brabender Units, an arbitrary unit of viscosity). A good quality grain for bread is expected to have a high protein and gluten content. A good quality flour with strong gluten will have a high peak (high consistency) and a short peak time (rapid to reach maximal consistency) when hydrated.

3.3.5 Statistical analysis

All the statistical analyses were performed in R (v.4.1.2). To visualise the differences in the microbial community (amplicon dataset and CLPP derived from EcoPlates assays) across sampling dates, treatments, and cultivars, we used the function *cmdscale* of the vegan package (v.2.6-2) (Oksanen *et al.* 2013) to produce principal coordinate analysis (PCoA) based on the Bray-Curtis dissimilarity index. The effect of sampling date, treatments, block, genotypes on the microbial community structure and carbon utilisation patterns was tested using permutational multivariate analysis of variance (PERMANOVA) based on the Bray-Curtis dissimilarity index (*adonis2* function of the vegan package, v.2.6-2). Three-way repeated measures analysis of variance (rmANOVA) using the *aov* function was used to test for significant differences in alpha diversity, N-cycle related genes and ITS 1 region and 16S rRNA gene abundance. The normality of the residuals was examined graphically using *ggqplot* (ggpubr package v.0.4.0) (Kassambara and Kassambara 2020) and was tested by the Shapiro-Wilk test using the *shapiro.test* function. If the data did not meet the requirements of the tests, it was log or square root transformed. The homoscedasticity of the data was evaluated using the Mauchly's sphericity test of the rstatix package (v.0.7.0) (Kassambara 2020). Correlation analyses between microbial parameters and wheat grain quality were performed with the *cor.test* (stats package v.4.2.1) (Worldwide 2020) function together with the *p.adjust* function to adjust the p-value with the Benjamin-Hochberg correction for multiple tests.

3.3.6 Predictive modeling

Our goal was to model grain quality (protein, gluten, BEM and PMT) using the microbial indicators measured (bacterial and fungal alpha diversity, bacterial and fungal beta-diversity, carbon utilization patterns, F:B ratio, and N-cycle gene abundance), for each sampling date separately to find the optimal sampling date for modeling. Since our PERMANOVAs revealed that the two wheat genotypes harbored significantly different microbial communities, we modeled them separately. This resulted in 14 different microbial datasets containing each 24 samples. We excluded outlier data points using the rstatix package (v.0.7.0).

To reduce the dimensionality of the 16S rRNA gene and ITS 1 region amplicon OTU tables and of the microbial carbon usage, we performed a procedure called orthogonalization. In brief, we performed a principal component analysis (*PCA* function of the FactomineR package v.2.6) (Husson *et al.* 2016) on Hellinger-transformed (*decostand* function of vegan package v. 2.6-2) OTU tables or carbon usage patterns and used the 5 first principal components in the models. Individual OTUs and carbon substrates were then correlated to these 5 components to have an idea of the taxonomic composition of the OTUs or carbon substrates influencing each of the components. We kept OTUs and carbon substrates with correlation having a $P<0.05$. For the OTUs correlated with the principal components, a taxonomic summary at the genus level was generated using the Phyloseq package (v.1.40.0) (McMurdie and Holmes 2013).

We chose least absolute shrinkage and selection operator (LASSO) regression as a modeling method to predict wheat quality for the following reasons: (i) to avoid overfitting, which may be problematic with other regression methods (least square regression or general linear model), especially when there are many explanatory variables and a few samples, (ii) to be able to select only the most important predictive variables (i.e., feature), to reduce the mean square error of the model, and (iii) to have an interpretable model. Indeed, LASSO regression shrinks the coefficient of the non-significant predictors to zero, keeping only the predictors with the highest explanatory power.

The microbial features included: principal components 1-5 derived from the microbial OTU and carbon usage tables, the abundance of N-cycle related gene, the F:B ratio, and the bacterial and fungal alpha-diversity. First, we standardized the data (other than the PCs) using the *scale* (scales package v.1.2.1) function and then selected the optimal lambda values with 10-fold cross validation using *c.v glmnet* function of the glmnet package (v.4.1-4) (Friedman *et al.* 2017). We generated the models with penalty scores based on the lowest lambda value, which indicates non-collinear effects and low levels of inflated variance in the selected variables. The predicted outputs values from these LASSO models were calculated using the *predict* function of the stats package (v. 4.2.1). The predictive accuracy of the models was then evaluated by calculating R^2 and mean squared error values (MSE) between the observed and the predicted values.

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were also calculated to evaluate the quality of the models, while taking into account the trade-off between goodness-of-fit and simplicity. Due to the lack of sufficient samples (n=24) in each model, we could not split the datasets in training and test datasets to further test the models' performance. Finally, we compared the accuracy and performance across the different sampling dates. The R code used for data manipulation, statistical analyses, and predictive modeling is available on GitHub (https://github.com/numanibneasad/Soil_Microbiome) whereas the data used for the analyses is available on Zenodo (https://doi.org/10.5281/zenodo.7293949)

3.4 Results

3.4.1 Effect of experimental treatments on microbial parameters

The sampling date significantly affected all microbial parameters, including microbial carbon utilization, microbial alpha and beta diversity, the F:B ratio, and the abundance of N-cycle-related genes (Tables 3-1 and 3-2). Furthermore, the structure of the fungal community was influenced by wheat genotypes (Table 3-1). There was a significant interactive effect ($P < 0.05$) of the precipitation treatment and wheat genotype on the abundance of the archaeal *amoA*, *nirK* and *nosZ* genes (Table 3-2).

Table 3-1: Multivariate statistical analysis to test treatment effects on microbial indices. Permanova based on Bray Curtis dissimilarities for microbial carbon utilization profiling (Biolog EcoPlate) and community structure based on 16S rRNA gene and ITS 1 region amplicon for the effect of precipitation exclusion treatments, sampling dates and genotype.

	Biolog			**16S**			**ITS**		
	R^2	F	Pr(>F)	R^2	F	Pr(>F)	R^2	F	Pr(>F)
treatment	0.013	4.95	0.002**	0.003	0.95	0.419	0.004	1.45	0.086
date	0.105	39.71	0.001***	0.01	5.06	0.001***	0.01	2.90	0.001***
genotype	0.002	0.58	0.754	0.00	1.04	0.29	0.01	2.61	0.003**
block	0.005	1.83	0.108	0.01	4.36	0.001***	0.03	11.72	0.001***
genotype× treatment	0.002	0.81	0.506	0.00	1.51	0.061	0.00	1.71	0.039*

Treatment: precipitation exclusion (0%, 25%, 50%, 75%). Date: sampling dates. Genotypes: drought-sensitive wheat and drought-tolerant wheat. "." $0.1 < P < 0.05$; "*" $P < 0.05$; "**" $P < 0.01$; "***" $P < 0.001$

Table 3-2. **Parametric statistical analysis to test treatment effects on N-cycle related gene abundance. Three-way repeated measure ANOVA for bacterial and archaeal ammonia monooxygenase, nitrite reductase, nitrous oxide reductase gene abundance and the Fungi: Bacteria ratio for the effect of precipitation exclusion treatments, sampling dates and genotype.**

	AOA	***AOB***	***nirK***	***nosZ***	***F:B ratio***
treatment	1.449	0.241	0.940	1.027	0.467
date	46.382***	40.379***	40.176***	79.707***	86.755***
genotype	0.205	0.006	0.388	0.689	0.043
block	2.180*	3.175**	2.682*	0.995	0.918
treatment × genotype	4.782**	0.993	4.356**	3.188**	0.854

F-values are shown in the table.

Treatment: treatments with precipitation exclusion (0%, 25%, 50%, 75%). Date: sampling dates. Genotype: drought-sensitive wheat and drought-tolerant wheat. ANOVA significance, "." $0.1 < P < 0.05$; "*" $P < 0.05$; "**" $P < 0.01$; "***" $P < 0.001$

3.4.2 Correlation between microbial and grain quality parameters

We performed Spearman correlations to test if some microbial parameters covaried with wheat quality data (grain gluten and protein content and flour peak maximum time (PMT) and maximum recorded torque (BEM)). We did not find a significant effect of rainfall exclusion treatment on grain qualities but found a significant effect of wheat genotype on protein content ($P<0.001$) and PMT ($P<0.001$), so we decided to treat the two genotypes separately and all the precipitation treatments together. Correlations between grain quality and microbial carbon use fluctuated over time (Table 3-3). The correlations between carbon sources and grain quality indicators were all negative for the DT genotype whereas both positive and negative correlations were found for the DS genotype (Table 3-3). The abundance of microbial N-cycling genes was found to be correlated to grain quality measurements mostly for soil collected on the early (May and June) sampling dates (Table 3-4). The *amoA* (archaeal and bacterial), *nirK* and *nosZ* genes quantified in the DT genotype samples on May 10 and May 24 were negatively correlated to protein and gluten content (Table 3-4). Only the F:B ratio was positively correlated to protein content (Table 3-4). For the DS genotype, the *amoA* (archaeal and bacterial) and the *nosZ* genes were negatively correlated to the grain quality parameters and the F:B ratio was positively correlated to PMT for soil samples collected on May 24 (Table 3-4). The F:B ratio was positively correlated with BEM for both genotypes on July 5 and

June 21. For the DT genotype, we found some positive and negative correlations between *nosZ* and PMT, AOA and protein (July 19 and August 1), while for the DS genotype *nirK* was positively correlated with gluten on July 5. Many significant correlations between microbial richness/diversity indices and grain baking quality were found, mostly for the DS genotype (Table 3-5). Significant correlations between microbial community descriptors (PCA axes for OTUs and microbial carbon use) and grain quality indicators for sampling dates in May and June were also identified.

Table 3-3: Spearman correlations between microbial carbon utilization and grain quality. Significant (P<0.05) Spearman correlations between microbial carbon utilization and grain baking quality for each sampling date (N=24).

Drought tolerant				**Drought Sensitive**			
Carbon source	Quality	R_s	P-value	Carbon source	Quality	R_s	P-value
10-May				**10-May**			
Beta methyl D-glucoside	Protein	-0.609	0.002	N-acetyl D-glucosamine	Gluten	0.537	0.008
Phenylethylamine	BEM	-0.587	0.003	**07-Jun**			
				4-hydroxy benzoic acid	Gluten	0.522	0.009
24-May							
α-keto butyric acid	Gluten	-0.628	0.001	**21-Jun**			
21-Jun				Tween.40	Protein	-0.601	0.002
N-acetyl D-glucosamine	PMT	-0.562	0.005	**05-Jul**			
05-Jul				L-Serine	Protein	-0.547	0.007
Glycogen	PMT	-0.552	0.006	D-L alpha glycerol phosphate	Protein	-0.550	0.007
01-Aug				**19-Jul**			
Pyruvic acid methyl ester	Gluten	-0.599	0.002	L-phenylalanine	PMT	0.576	0.004
				01-Aug			
				L-asparagine	PMT	-0.575	0.006

Table 3-4: Spearman correlations between functional gene abundance and grain quality. Significant (P<0.05) Spearman correlations between functional gene abundance and grain baking quality for each sampling dates (N=24).

	Drought tolerant				Drought sensitive		
Gene	Quality	R_s	P-value	Gene	Quality	R_s	P-value
10-May				24-May			
nosZ	Gluten	-0.406	0.054	***AOB***	Gluten	-0.504	0.012
24-May				***AOA***	Protein	-0.406	0.055
AOB	Gluten	-0.450	0.031	***nosZ***	BEM	-0.400	0.059
nirK	Protein	-0.441	0.035	***F:B ratio***	PMT	0.425	0.043
AOA	Protein	-0.578	0.004	07-Jun			
F: B Ratio	Protein	0.547	0.007	**n*irK***	Gluten	-0.441	0.035
07-Jun				21-Jun			
F: B Ratio	Protein	0.426	0.048	***F: B Ratio***	Protein	0.406	0.054
21-Jun				***F: B Ratio***	PMT	-0.406	0.055
AOA	Protein	-0.563	0.005	***F: B Ratio***	BEM	0.492	0.017
AOA	PMT	0.404	0.056	19-Jul			
05-Jul				**n*irK***	Gluten	0.558	0.009
nirK	Gluten	-0.443	0.034				
nosZ	PMT	0.401	0.058				
F: B Ratio	BEM	0.479	0.021				
19-Jul							
AOA	Protein	-0.426	0.042				
01-Aug							
nosZ	PMT	0.392	0.058				

Table 3-5: Spearman correlations between microbial diversity indices and grain quality. Significant (P<0.05) Spearman correlations between bacterial and archaeal and fungal richness and diversity and grain baking quality for each sampling dates (N=24).

	Drought tolerant				**Drought sensitive**		
Diversity	Quality	R_s	P-value	Diversity	Quality	R_s	P-value
16S				**16S**			
07-Jun				**10-May**			
ACE	Protein	-0.409	0.058	Chao1	Protein	-0.454	0.029
05-Jul				**24-May**			
Chao1	BEM	0.472	0.023	Shannon	BEM	0.468	0.024
ACE	PMT	-0.467	0.025	Chao1	Protein	-0.414	0.050
ITS				**21-Jun**			
10-May				PD	Protein	-0.472	0.023
Shannon	Gluten	-0.444	0.034	Chao1	Protein	-0.482	0.020
Simpson	Gluten	-0.416	0.048	Chao1	Gluten	-0.520	0.011
				ACE	Protein	-0.418	0.047
				ACE	Gluten	-0.446	0.033
				19-Jul			
				Chao1	Gluten	0.549	0.007
				ACE	Gluten	0.549	0.007
				01-Aug			
				Simpson	PMT	0.434	0.044
				ITS			
				21-Jun			
				ACE	BEM	-0.465	0.025
				PD	Gluten	0.439	0.036
				01-Aug			
				Chao1	PMT	0.512	0.015
				Chao1	BEM	-0.483	0.023
				ACE	PMT	0.493	0.020
				PD	PMT	0.491	0.020

3.4.3 Model performance in predicting grain quality at different dates

We applied least absolute shrinkage and selection operator (LASSO) regressions for each sampling date separately, to identify the date where model accuracy would be maximal to predict grain quality. In the case of the DT genotype, the best models for grain quality indicators had mean square errors ranging from 0.08 to 0.51 and AIC ranging from -17.00 to -8.35 (Table 3-6 and Fig. 3-1). The best models identified were based on microbial indicators from May 10, May 24, and June 7. For gluten and protein content, the LASSO regression had the highest accuracy for microbial indicators measured from samples collected on May 10. These models selected 11 and 8 variables, resulting in R^2 of 0.95 and 0.76, for gluten and protein

respectively (Table 6 and Figure 1). The model's accuracy for gluten and protein content prediction decreased over time, (Table 6). For BEM and PMT, the best sampling dates for model generation were June 7 (R^2=0.92) and May 24 (R^2=0.57), respectively (Table 6 and Fig. 1). The most parsimonious model across all quality indicators was the one predicting PMT which only included 2 predictors (Table 3-6). For some sampling dates, no microbial predictor was selected by the LASSO procedure, resulting in null models (Table 3-6).

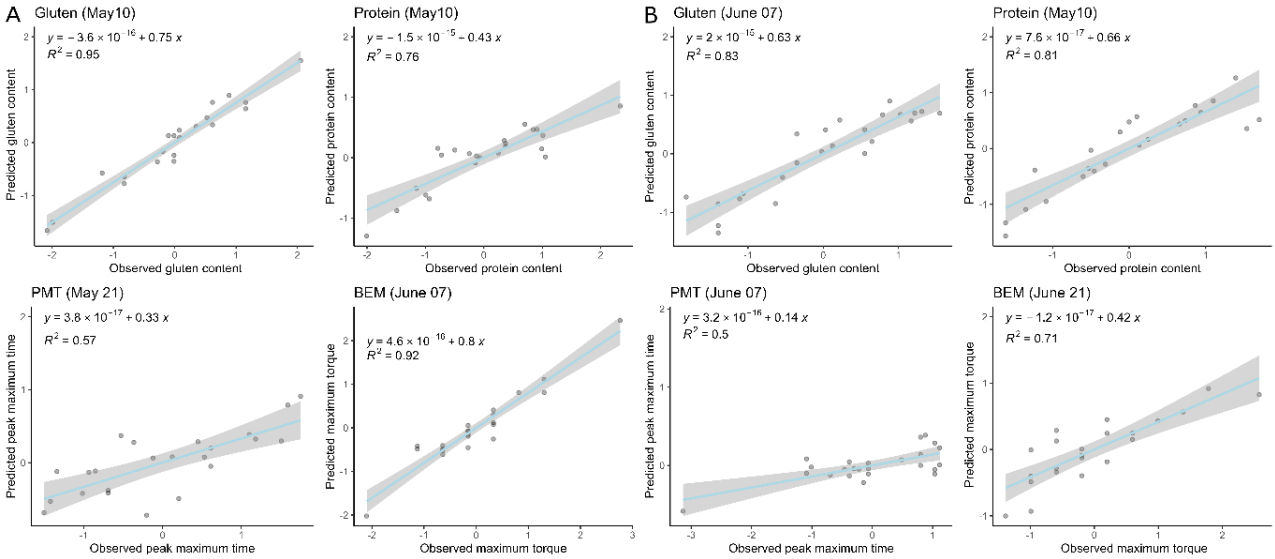


Figure 3-1: Microbial-based optimal models on optimal soil sampling dates. Observed values vs. predicted values from LASSO regression models for wheat grain gluten and protein content and flour maximum torque (BEM) and peak maximum time (PMT) for the drought-tolerant (A) and drought-sensitive genotypes (B).

Table 3-6: Comparative model analysis for drought-tolerant genotype. Comparative analysis of the LASSO model performance for the wheat grain quality of the drought-tolerant genotype (DT).

	T1	T2	T3	T4	T5	T6	T7
Date	10-May	24-May	07-Jun	21-Jun	05-Jul	19-Jul	01-Aug
Gluten (DT)							
C.V (best Lambda)	0.04	0.38		0.56	0.72		
AIC	-16.14	2.00		1.33	2.00		
BIC	-15.09	3.04		2.42	3.09		
Nb of variables:	11	1		1	1		
MSE (Mean Square Error)	0.08	0.95		0.92	0.95		
R^2	**0.95**	0.15		0.54	0.54		
Protein (DT)							
C.V (best Lambda)	0.15	0.24	0.19	0.28	0.46	0.36	0.18
AIC	-11.64	-9.21	-9.56	-3.40	2.00	2.00	-8.24
BIC	-10.51	-8.07	-8.47	-2.26	3.14	3.14	-7.15
Nb of variables:	8	5	7	2	1	1	2
MSE (Mean Square Error)	0.36	0.47	0.43	0.73	0.96	0.96	0.53
R^2	**0.76**	0.69	0.72	0.33	0.22	0.14	0.57
PMT (DT)							
C.V (best Lambda)		0.21					0.42
AIC		-8.35					2.00
BIC		-7.21					3.18
Nb of variables:		2					1
MSE (Mean Square Error)		0.51					0.96
R^2		**0.57**					0.19
BEM (DT)							
C.V (best Lambda)	0.38	0.25	0.03			0.14	0.20
AIC	-2.34	-7.31	-17.00			-8.92	-5.32
BIC	-1.21	-6.17	-15.91			-7.78	-4.14
Nb of variables:	1	2	10			7	4
MSE (Mean Square Error)	0.77	0.55	0.09			0.48	0.65
R^2	0.35	0.50	**0.92**			0.58	0.47

Missing values indicate failure to build models on specific sampling dates using LASSO regression. A total of 40 variables were used as inputs.

PMT=Peak Maximum Time, BEM= flour maximum recorded torque, Nb=Number, AIC= Akaike Information Criterion, BIC=Bayesian Information Criterion, C. V= Cross validation

The overall model performance (based on R^2 values) in predicting grain quality for the DS genotype was lower than the DT genotype (Table 3-7). Maximum accuracy of LASSO regression model was observed on June 7 for gluten and PMT, on May 10 for protein, and June 21 for BEM (Table 3-7). The best PMT and BEM predictive models used about half the number of the total predictors used in the best gluten and protein predictive models (PMT: 4, BEM: 6, gluten: 14 and protein: 11) (Table 3-7). Predictive modeling of protein content between May 24 and July 5, and on August 1 was unsuccessful and the level of accuracy of the model weas low on July 19 (Table 3-7). A similar trend was observed for PMT: sampling dates after June 7 resulted in less accurate or no model at all (Table 3-7). BEM prediction was also unsuccessful for samples collected on June 7. Overall, like for the DT genotype, the predictive models for the DS genotype dataset showed the best accuracy for quality prediction with microbial data from the May and June samplings.

Table 3-7: **Comparative model analysis for drought-sensitive genotype. Comparative analysis of the model performance of LASSO for the wheat grain quality of drought-sensitive genotype (DS).**

	T1	T2	T3	T4	T5	T6	T7
Date	10-May	24-May	07-Jun	21-Jun	05-Jul	19-Jul	01-Aug
Gluten (DS)							
AIC	-8.28	2.00	-16.01	-14.60	-1.97	-14.66	0.67
BIC	-7.14	3.14	-14.84	-13.46	-0.83	-13.52	1.76
C.V (best Lambda)	0.18	0.45	0.08	0.09	0.34	0.10	0.32
Nb of variables:	6	1	14	11	1	10	1
MSE (Mean Square Error)	0.51	0.96	0.21	0.23	0.78	0.23	0.89
R^2	0.61	0.22	0.83	0.81	0.30	0.81	0.17
Protein (DS)							
C.V (best Lambda)	0.06					0.17	
AIC	-15.15					-7.69	
BIC	-14.02					-6.55	
Nb of variables:	11					6	
MSE (Mean Square Error)	0.21					0.54	
R^2	0.81					0.53	
PMT (DS)							
C.V (best Lambda)	0.19	0.41	0.33		0.38		0.24
AIC	-5.76	-3.94	-3.56		2.00		-1.55
BIC	-4.63	-2.80	-2.38		3.14		-0.46
Nb of variables:	4	1	4		1		1
MSE (Mean Square Error)	0.62	0.70	0.73		0.96		0.79
R^2	0.35	0.45	0.50		0.15		0.24
BEM (DS)							
C.V (best Lambda)	0.13	0.36		0.19	0.18	0.13	0.17
AIC	-10.11	-0.70		-10.37	-8.21	-8.67	2.00
BIC	-9.02	0.39		-9.28	-7.11	-7.58	3.04
Nb of variables:	11	2		6	5	4	1
MSE (Mean Square Error)	0.40	0.83		0.39	0.49	0.47	0.95
R^2	0.65	0.32		0.71	0.61	0.56	0.03

Missing values indicate failure to build models on specific sampling dates using LASSO regression. A total of 40 variables were used as inputs.

PMT=Peak Maximum Time, BEM= flour maximum recorded torque, AIC= Akaike Information Criterion, BIC=Bayesian Information Criterion, C. V= Cross validation, Nb=Number

3.4.4 Microbial features selected in the optimal models

The best LASSO models for the DT genotype contained microbial features that varied but were often the principal components derived from OTU tables or carbon utilization patterns, or the alpha diversity indices. Bacterial and archaeal OTUs from the *Nitrosphaera* (an ammonia oxidizing archaeal genus), *Rhodoplanes, Solirubrobacter,* and *Terrimicrobium* were the main contributors to the principal component 2 (explained variance: 5.1%) calculated from the May 10 dataset that was selected in the models for gluten and protein content (Fig. 2 and Table 8). In contrast, the main contributors to the bacterial and archaeal principal component 1 (explained variance: 6.0%), 2 (5.2%) and 3 (5.1%) selected for the model predicting BEM on June 7 were from the *Conexibacter, Gaiella, Nitrososphaera, Hyphomicrobium* and Gp16 (an uncultured genus of Acidobacteria) genera (Fig. 3-2A). The fungal OTUs that contributed to the principal components selected in the May and June models belonged to the *Mortierella, Ganoderma,* and *Gliomastix* genera (Fig. 3-2B). We found a negative relationship between the bacterial phylogenetic diversity index and gluten content and a positive relationship between bacterial Simpson diversity and gluten content and BEM in the May 10 and June 7 models (Table 3-8).

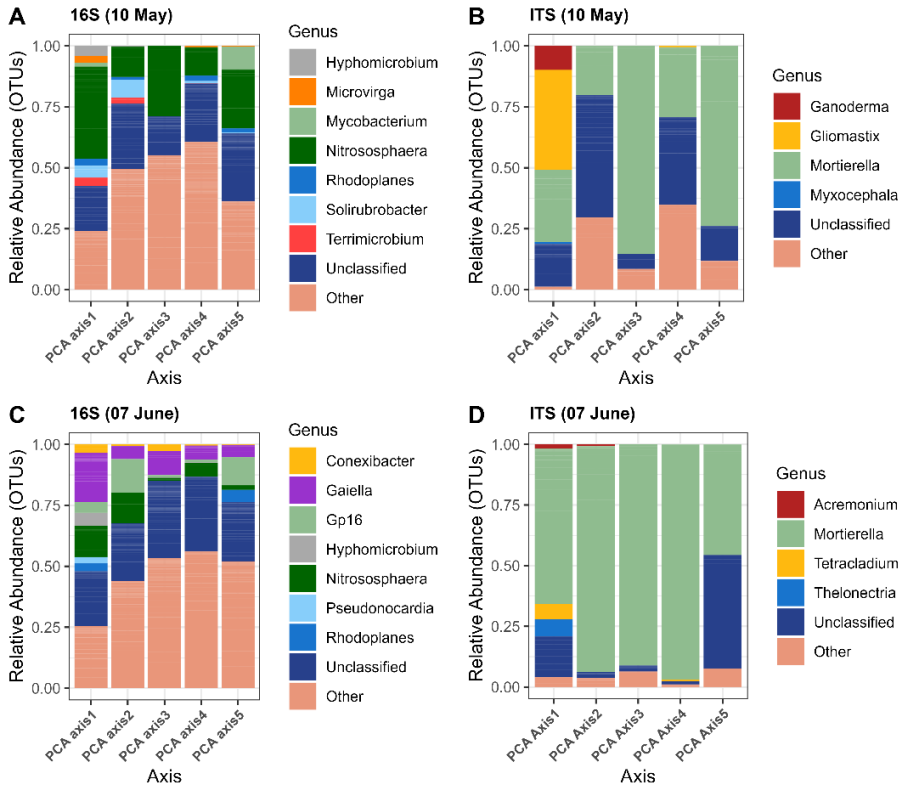


Figure 3-2: The relative abundance of the bacterial and archaeal, and fungal genera for drought tolerant genotype. The relative abundance of the bacterial and archaeal (A, C) and fungal (B, D) genera significantly correlated with the first five principal components for the drought tolerant genotype for the May 10 (A, B) and June 7 (C, D) sampling dates. Others: various genera with relative abundances below 0.1%.

Principal components derived from carbon utilization patterns were also included in all our most accurate models for the DT genotype (Table 3-8). The models predicting protein and gluten content (May 10) selected 3 to 4 of the top 5 principal components included, for which the most important contributing carbon substrates were Putrescine (r_s=-0.91; P<0.001), L-Arginine (r_s=0.74; P <0.001), Pyruvic Acid methyl ester (r_s=-0.62; P<0.001), Glycogen (r_s=0.59; P<0.001) and L-Threonine (r_s= -0.56; P<0.001). The model predicting BEM (June 7) selected principal component 2 (explained variance: 9.3%), 3 (7.4%), and 4 (4.7%) and the most important contributing carbon substrates of the principle components were alpha-cyclodextrin (r_s=0.69; P=0.002), alpha-keto butyric Acid (r_s=0.68; P=0.003), γ-amino butyric acid (r_s= -0.66; P =0.006),

Glucose 1-phosphate (r_s=-0.71; P=0.001). Finally, the principal component 2 (explained variance: 7.5%) selected in the model predicting PMT (May 24) was correlated to glycogen (r_s=0.59; P=0.002), alpha-cyclodextrin (r_s=0.68; P<0.001) and γ-amino butyric acid (r_s=-0.65; P=0.004). We also observed a negative relationship between protein content and *nirK* (regression coef. = -0.183) and gluten content and *nosZ* (regression coef. = -0.235) in the models obtained on May 10 (Table 3-7).

Table 3-8: Selected microbial feature in models of drought-tolerant genotypes. Microbial parameters included in the LASSO models for wheat grain quality of the drought-tolerant genotype (DT).

Gluten-May10		**Protein-May10**		**PMT-May21**		**BEM-June07**	
Variables	Coefficients	Variables	Coefficients	Variables	Coefficient	Variables	Coefficient
Intercept	-1.60×10^{-14}	Intercept	-2.50×10^{-15}	Intercept	-2.00×10^{-16}	Intercept	3.63×10^{-14}
Bacteria.PC2	0.492	Bacteria.PC2	-0.141	Biolog.PC2	-0.433	Bacteria.PC1	-0.184
Fungi.PC3	-0.011	Fungi.PC1	-0.184	ACE fungi	0.225	Bacteria.PC2	0.254
Biolog.PC2	0.354	Fungi.PC3	-0.188			Bacteria.PC3	-0.051
Biolog.PC3	0.016	Fungi.PC5	-0.072			Fungi.PC2	-0.102
Biolog.PC4	-0.153	Biolog.PC1	0.111			Fungi.PC4	-0.099
Biolog.PC5	-0.086	Biolog.PC4	-0.185			Fungi.PC5	0.643
Simpson bacteria	0.628	Biolog.PC5	-0.122			Biolog.PC2	0.365
PD bacteria	-0.997	*nirK*	-0.183			Biolog.PC3	-0.249
ACE bacteria	0.270					Biolog.PC4	0.381
Chao1 fungi	-0.202					Simpson bacteria	0.498
nosZ	-0.235					Chao1 bacteria	-0.080

PD: Phylogenetic diversity, ACE: Abundance-based Coverage Estimators, PC: principal component.

As for the DT genotype models, the models for the DS genotype were mainly composed of principal components calculated from the OTU tables and from the carbon utilization patterns, and from alpha-diversity indices (Table 9). The LASSO model predicting protein content selected the bacterial principal component 4 (explained variance: 4.9%) for the May 10 sampling date (Table 3-9). This principal component was correlated with OTUs belonging to the *Nitrososphaera, Rhodoplanes, Solirubrobacter*, and *Terricomicrobium* (Fig. 3-3A). On the same date, the fungal OTUs contributing the most to the principal component 1 (explained variance: 7.3%), 3 (5.6%), 4 (5.5%), and 5 (5.2%) belonged to the *Acremonium, Mortierella, Pezizella*, and *Tetracladium* (Fig. 3-3B). On June 7, the models predicting gluten content and PMT selected the bacterial principal components 2, 4, and 5 (Table 3-9). These axes explained 4.7-4.5% of the variation and were correlated to OTUs related to *Gaiella*, Gp6, *Hyphomicrobium, Nitrososphaera,*

Rhodoplanes, and *Solirubrobacter* (Fig.3-3C). On June 21, the model predicting BEM selected the bacterial principal components 2 and 4 (Table 3-9), which explained 4.8% and 4.7% of the variation and were correlated to OTUs related to *Arthobacter*, *Nitrososphaera*, *Gaiella*, Gp6, *Hyphomicrobium*, *Bradyrhizobium*, *Terrimicrobium* and *Rhodoplanes* (Fig 3-4A). The fungal PC 1 (7%), 2 (6.2%), 4 (5.0%), and 5 (4.9%) selected for the June 7 were correlated to OTUs related to *Acremonium*, *Mortierella*, and *Tetracladium* (Fig. 3-3D). The fungal PC2 (5.8%) selected in the model for BEM in June 21 was linked to OTUs related to *Acromonium* and *Mortierella* (Fig. 3-4B). Fungal richness and diversity were selected in the LASSO models for Gluten, Protein and BEM, with either positive (Protein and BEM) or negative (Gluten) coefficients (Table 3-9).

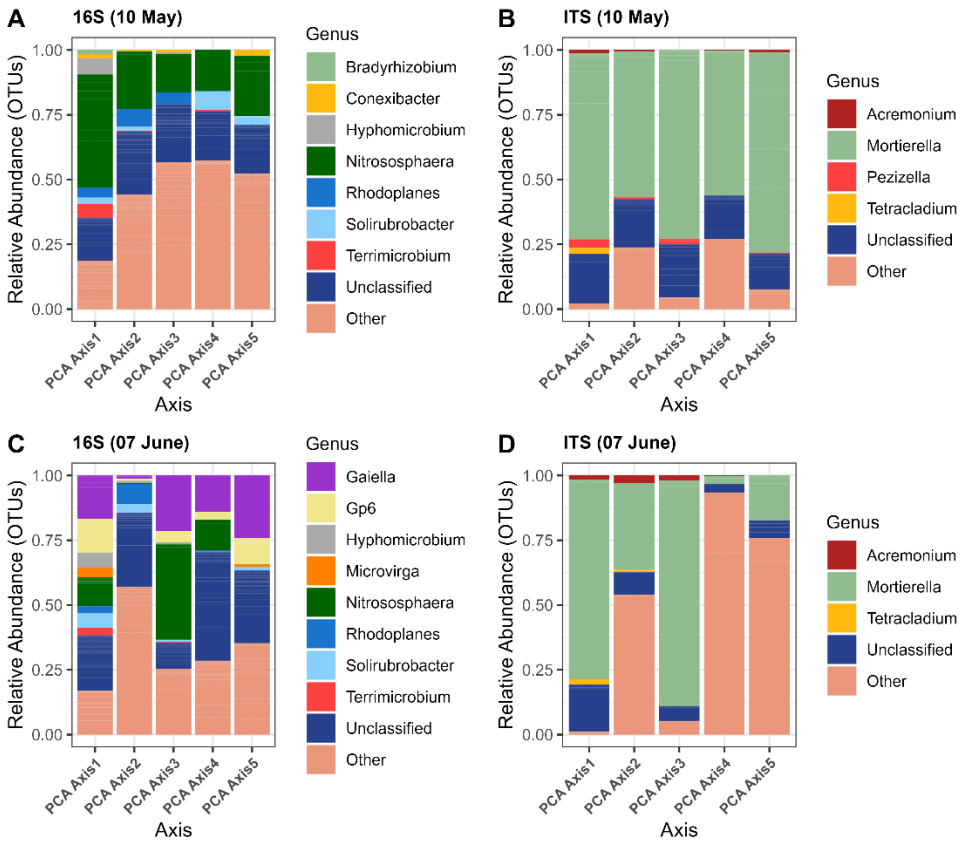


Figure 3-3: The relative abundance of the bacterial and archaeal and fungal genera for drought sensitive genotype. The relative abundance of the bacterial and archaeal (A, C) and fungal (B, D) genera significantly correlated with the first five principal components for the drought tolerant genotype for the May 10 (A, B) and June 7 (C, D) sampling dates. Others: various genera with relative abundances below 0.1%.

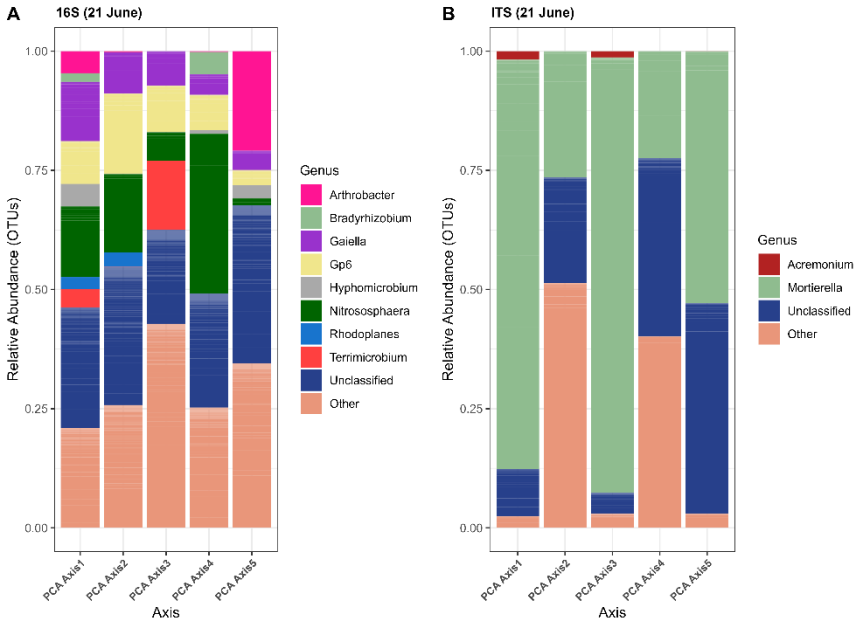


Figure 3-4: Relative abundance of bacterial and archaeal and fungal genera for drought-sensitive genotypes on June 21. The relative abundance of the bacterial and archaeal (A) and fungal (B) genera significantly correlated with the first five principal components for the drought sensitive genotype for the June 21 sampling dates. Others: various genera with relative abundances below 0.1%.

For the May 10 model (protein), the carbon substrates contributing the most to the selected principal components were beta-methyl D-glucoside (r_s=0.61; P=0.001), D-glucosamine acid (r_s= -0.58; P=0.003), D-galactonic acid y- lactone (r_s =-0.53; P=0.008). For the June 7 models (gluten and PMT), the carbon substrates contributing the most to the selected PC were Glucose 1-phosphate (r_s=0.81; P<0.001), D-galactonic acid y-lactone (r_s=0.64; P=0.0005), 4-hydroxy benzoic acid (r_s= -0.66; P=0.0005), 2-hydroxy benzoic acid (r_s=0.56, P=0.003). Finally, for the June 21 model (BEM), the carbon substrates contributing the most to the selected PC were L-phenylalanine (r_s= 0.55; P= 0.003) and alpha-cyclodextrin (r_s=-0.49;

P=0.011). We also observed that the models selected the fungal: bacterial ratio, which negatively influenced the gluten content on June 7 and positively influenced BEM on June 21. There was a negative relationship between the abundance of the bacterial *amoA* gene and gluten content, and a positive relationship between *nosZ* and gluten content on June 7 (Table 3-9).

Table 3-9: Selected microbial feature in models of drought-sensitive genotypes. Microbial parameters included in the LASSO models for the wheat grain quality of the drought-sensitive genotype (DS).

Gluten-June 07		Protein-May10		PMT-June 07		BEM-June21	
Variables	Coefficients	Variables	Coefficients	Variables	Coefficients	Variables	Coefficients
Intercept	7.96×10^{-15}	Intercept	-1.80×10^{-17}	Intercept	3.57×10^{-16}	Intercept	-8×10^{-17}
Bacteria.PC2	-0.018	Bacteria.PC4	-0.541	Bacteria.PC4	-0.009	Bacteria.PC2	-0.010
Bacteria.PC5	0.216	Fungi.PC1	-0.219	Fungi.PC4	0.146	Bacteria.PC4	-0.263
Fungi.PC1	0.012	Fungi.PC3	-0.093	Fungi.PC5	0.120	Fungi.PC2	-0.086
Fungi.PC2	-0.361	Fungi.PC4	-0.262	Biolog.PC5	0.026	Biolog.PC4	-0.151
Fungi.PC4	-0.026	Fungi.PC5	0.141			Chao1 fungi	0.182
Fungi.PC5	-0.072	Biolog.PC3	0.027			F:B ratio	0.213
Biolog.PC1	0.317	Biolog.PC4	-0.099				
Biolog.PC3	-0.078	Biolog.PC5	-0.009				
Biolog.PC5	0.024	Chao1 bacteria	-0.284				
Simpson bacteria	0.089	Chao1 fungi	0.154				
PD fungi	-0.150	PD fungi	0.012				
AOB	-0.460						
nosZ	0.115						
F:B ratio	-0.031						

PD: Phylogenetic diversity, F: B: Fungal: Bacterial ratio, PC: principal component.

3.5 Discussion

Plant- and soil-associated microbial communities vary throughout the seasons/plant growth stages (Chaparro *et al*. 2013, 2014; Moroenyane *et al*. 2021; Azarbad *et al*. 2022; Azarbad *et al.* 2021; Wang *et al*. 2022) and it was unsure what was the best timing to create models to predict wheat grain quality. By sampling the same field every 2 weeks and measuring a wide range of microbial parameters, we were able to show with LASSO regression that the predictive value of microbial parameters is optimal during the earlier stages of wheat growth, at the seedling (May) or tillering stages (June). Many classes of microbial

parameters (e.g., alpha diversity indices, principal components, N-cycle genes) were consistently singled out by the regression models, which could allude to a mechanistic link between grain quality and the parameter identified, or simply to covariation between the microbial parameter and grain quality due to a third unmeasured parameter. Our work focused on wheat, and although it would be interesting to see if similar patterns apply to other crops, it is the first and necessary step to start building microbial-based predictive models for crop yields and quality.

All the best models were made with data collected before the end of June, which is at the early stages of wheat growth in Quebec. This is coherent with our previous results that showed that good predictive models could be made with soil samples taken in May or June (Yergeau et al. 2020; Asad et al. 2021) even though different sampling point were not compared. Similarly, in another system where microbial communities play a key role in the process of interest, phytoremediation, it was shown that early microbial community composition could predict the potential of the plants to decontaminate soil or to survive (Bell *et al*. 2014; Yergeau *et al*. 2015). Navarro-Noya *et al*. (2022) showed that the complexity of microbial structure and diversity increases with maize development, and that the effect of agricultural practices on the soil microbiome was more evident at the early stages, which could explain why we found here that early microbial indicators performed better. This is encouraging for future work, as the ultimate goal of this type of predictive modeling is to have a tool that could be used to guide management strategies for farmers. Maximum usefulness will happen if indicators of yields or quality can be measured early, when it is still possible to intervene. It could be that the sampling dates highlighted are the ones that are the most critical for wheat grain quality, but for wheat, it is generally thought that the grain filling stage (around mid July in Quebec) is the most critical stage in term of N nutrition for high quality grain (Zörb *et al*. 2018). However, unless there is an unlikely massive microbial immigration, the microorganisms that can modulate or are indicative of soil N availability are already present in the soil early at seeding, and it is likely that their abundance and diversity at this stage could predict wheat grain quality. In fact, it was recently suggested that, because of their potential to be influenced by legacy and current environmental conditions, microbial communities act as multivariate integrators of the current and past physico-chemical conditions of their immediate environment, making them highly suitable predictors for ecosystem processes (Correa-Garcia *et al*. 2022).

Microbiome data have characteristics (sparsity, high dimensionality, zero-inflated) that often make them challenging to use in models. Here, we transformed the OTU and carbon utilization patterns tables using eigenvalue decomposition, namely principal component analysis, which reduces the dimension of the datasets to (number of samples)-1 principal components that are orthogonal (not collinear) and ordered in decreasing order of variance explanation, moving from several thousands of descriptors to 23, in the case

of the OTU tables. We further reduced the dimensionality by only utilizing the first 5 principal components in our LASSO regression, with the idea that these components contained a large part of the variation in the original dataset. One downside of this approach is that it makes the models less directly interpretable, with principal components being composite variable for many OTUs or carbon sources. However, using correlation analyses of individual OTUs with the principal components we were able to identify taxonomic groups and carbon sources that were linked with the principal components. We also used LASSO regression that selects of the most significant variables and shrinks the regression coefficient of the other variable to zero, generally producing parsimonious, highly interpretable models containing a few variables. Although non-parametric methods (neural network, random forest, support vector machine, etc.) could produce more accurate models, they are often less interpretable, meaning that the predictors influencing the output cannot be easily identifiable. Still, our models had high accuracy of 50-95%. The predictive performance of LASSO regression to predict biological characteristics from microbiome data was shown to be excellent for zero-inflated data such as microbial OTU count tables (Xiao *et al.* 2018; Dong *et al.* 2020). We also had good results using linear regression coupled with forward/backward selection with a preselection of individual OTUs that showed the strongest correlations with the predictors (Yergeau *et al*. 2020; Asad *et al*. 2021).

General community descriptors, like alpha-diversity indices or principal components, were often selected as the best explanatory variables in the models and/or significantly correlated to quality parameters. Alpha diversity indices and eigenvectors (such as principal components) derived from microbial community structures are integrators of many parameters. Interestingly, it suggests that shallow sequencing to recover alpha and beta diversity patterns together with community level carbon utilization profiling would be sufficient to model wheat grain quality. Additionally, some specific microbial parameters, such as the abundance of N-cycle functional genes or the utilization of specific carbon substrates, were consistently singled out by the LASSO regression and the Spearman correlation analyses. For example, the negative relationships between wheat quality and the abundance of the *nirK*, *nosZ* and bacterial *amoA* genes were well aligned with previous work (Yergeau *et al*. 2020; Asad *et al*.2021). The relative abundance of OTUs belonging to the ammonia-oxidizing archaea taxon *Nitrososphaera* were also highly correlated with many of the principal components selected in the models, and the abundance of both the archaeal and the bacterial *amoA* genes was often negatively correlated to quality parameters. These results further suggests that a high abundance of ammonia oxidizers and denitrifiers reduces wheat grain quality through an increased energy requirement for nitrogen uptake and utilization or through nitrogen losses, as discussed before (Yergeau *et al*. 2020; Asad *et al*. 2021; Wang *et al*.2022). Indeed, since grain quality is linked to its protein content, it is energetically more efficient for the plant to uptake ammonia, which can directly be incorporated into amino acids, whereas nitrate will need to be transformed back to ammonia (Beeckman et

al., 2018). Nitrate uptake also requires more energy than ammonia uptake (Beeckman et al., 2018). Finally, nitrate is prone to leach and is a substrate for denitrification, which will lead to loss of nitrogen to the atmosphere. Manipulating or inhibiting the activity of these microbial guilds using, for instance, natural or artificial nitrification inhibitors may increase wheat grain quality. However, this strategy will need to be further studied to understand potential unwanted effects, as a common nitrification inhibitor, nitrapyrin, was shown to have off-target effects on the soil microbial community (Schmidt *et al.* 2022) and that nitrate stimulates lateral root elongation and affects various signaling pathways in the plant (Beeckman et al. 2018). It was recently shown that biological nitrification inhibition (BNI) genes could be introduced into wheat cultivars from a wild grass species (*Leymus recimosus)* (Subbarao *et al.* 2021). The root exudates of some Australian wheat genotypes were also shown to be able to inhibit a strain of bacterial ammonia-oxidizer (O'Sullivan et al. 2019). Microbiome manipulation is still in its infancy and, because of ecological processes underlying community assembly, it will be a challenge (Agoussar & Yergeau, 2021). It is also unclear if microorganisms involved in nitrification and denitrification are sufficient indicators for accurate modeling of the grain quality, and, consequently, if solely targeting these groups will result in the expected increase in grain quality. As our model showed, general community structure and diversity seem to also have a prime importance in determining wheat grain quality.

Our previous work showed that significant predictive models could be parametrized using microbial data measured early in the growing season, across a transect of more than 500 km (Asad et al. 2021). Here, we sought to confirm that early microbial measurements were optimal for such predictive models by focussing on a single field and sampling it every two weeks for a complete growing season. Taken together, the two studies confirm that our microbial-based models are effective at a large spatial scale and that they are optimally build using samples taken early in the season. Although we used a different modeling approach than previously, the selection of ammonia-oxidizers by the models was shared with our previous studies (Yergeau *et al*. 2020; Asad *et al*. 2021), suggesting a potential key role of this functional guild for wheat grain quality. Our manuscript lay the foundation for future attempts to predict and optimize crop yields and quality, on our way toward microbiocentric solutions to the pressing issues facing agriculture.

3.6 Acknowledgments

The entire staff of Les Moulins de Soulanges and La Meunerie La Milanaise, and more specifically Élisabeth Vachon, Stéphanie Carrière, Chafik Baghdadi and Robert Beauchemin, are gratefully acknowledged for their support of this study. All members of the Labo Yergeau are thanked for their help in maintaining and setting up the field experiment. Emmy L'Espérance provided the R code used for repeated-measures ANOVAs.

3.7 Funding

This study was supported by an FRQNT Team Grant (2019-PR-254256) and a Compute Canada Resource allocation (allocation 2020-3177) on the Graham system (University of Waterloo) to Etienne Yergeau.

3.8 Conflict of interest

The authors have no conflicts of interest to declare.

3.9 References

All the corresponding references of this chapter have been included at the end of the thesis.

4. GENERAL DISCUSSION AND CONCLUSION

4.1 Discussion

A sustainable and modified agrifood system is required to address the existing and future challenges in food production due to climate change. Microbiome-based research is a newly established field that is creating new opportunities to improve human nutrition and health management while addressing environmental issues. Applied microbiome research in agriculture is revolutionizing the comparative study of soil, plant and farm animal health monitoring systems and the ability to predict aspects of agricultural production and productivity (Callens *et al.* 2022). After conducting a thorough literature review, I discovered that microbial indicators related to soil structure and function are expected to have great potential for future applications in soil and plant phenotype assessment. One of the main findings in my thesis highlights the ability of the soil microbiome to explain agroecosystem conditions and the soil nutrient status affecting plant physiology and overall crop yield and quality. My thesis introduces a conceptual framework for a soil–microbiome–crop quality axis that underlies the various mechanisms of plant–microbiome interactions at both spatial and temporal scales. Banerjee *et al.* (2022) showed that the particular soil microbiome shared between different health cohorts interconnects soil, plant, and human microbes more than previously imagined. Other reports have described soil as a microbial seed bank that creates microhabitats for pathogens and beneficial microorganisms, providing a diverse niche for various important species (Xiong and Lu 2022). A sustainable soil ecosystem provides the nutrients needed to grow plants with high levels of productivity, resistance, and resilience (Carrillo *et al.* 2017). This is further supported by the indicative and predictive characteristics of soil microbial communities for wheat production, described in Chapters 2 and 3. In my thesis, I applied integrated statistical learning tools using genomics data for predictive modeling. The model used extensive functional and genomics data of the soil microbiome. I chose genomic approaches because they provide large amounts of data associated with microbial community structure and composition, which are needed for the predictive modeling of wheat grain quality. Other classical microbiological methods do not allow for the estimation of large spatial and temporal variations in microbial parameters.

Basically, chapter 2 describes how I measured the basic properties of soil that are commonly tested prior to crop-growing season and the microbial parameters that capture the largest variation in soil microbial processes at multiple wheat farms across Quebec. First, I explored the indicative features of soil microbes

that explain wheat yield and grain quality at the community level, derived from high throughput16S and ITS1 amplicon sequencing data. Then, I focused on the comparative model analysis for predicting wheat yield and grain quality, with two separate and combined parameters for soil physicochemical and microbial indices. The aim was to test whether significant microbial or biochemical indicators in the soil accurately predict wheat yield and quality when these indices are modeled individually or interactively with soil physicochemical indicators in a multiple linear regression, as this relationship has not been well-studied. My second objective (Obj-B), described in Chapter 3, was to identify the ideal dates for soil sampling during the wheat growing season. The distinct microbial traits measured from these samples would lead to the most accurate predictions of wheat-grain quality.

Soils sampled early in wheat growth showed the greatest accuracy for making predictions (Chapter 3), providing evidence of the high predictive performance of early-season microbial-based modeling. I tested my hypothesis (H-2) through the second specific objective (Obj-B). I followed a soil sampling scheme that was conducted during the wheat growing season and divided the scheme into 7 time points, according to different wheat growth stages. This allowed me to test the predictive performance of the same microbial indicators that were used in previous models of wheat quality at the regional scale. Based on the two main results from Chapters 2 and 3, it is clear that microbial communities are the best indicators for crop quality assessment and are useful parameters for future wheat production.

As discussed, microbiome assembly processes are driven by endogenous factors (e.g., species, genotype, developmental stage), biotic stress (e.g., herbivore, pathogen), anthropogenic factors (e.g., agricultural practices, urbanization, agrochemicals, nanomaterials), and environmental perturbations (related to geographic location, temperature, season, moisture, rainfall, local dispersal, etc.) (Zhan *et al.* 2022). In any given situation, the specific factor associated with the microbial assembly process determines the potential outcome for the whole microbe-driven agroecosystem. Moreover, the soil microbiome directly influences the nutrient uptake and cycling through nitrate fixation, nitrification and denitrification, phosphate solubilization and siderophore formation (Fierer 2017). The soil microbiome also provides an environment that is resistant to some soil-borne pathogens, promoting disease-suppressing soil. Furthermore, the root-associated soil microbiome contributes to a wide range of services including plant productivity and nutrition. Based on these recent data, agricultural and soil management decision approaches that do not incorporate these findings about soil microbial status will likely be ineffective.

The main focus of my thesis is on current sustainability issues in agriculture and the potential for microbes to inform agricultural management decisions for wheat production. I also discuss the importance of spatial and local climatic variations in soil microbiome performance. This is evidenced by the different qualities of grains from the same genotype grown in different regions and implies that soil microbiome

distribution and composition vary at the regional scale because of different agricultural management practices (e.g., fertilizer, pesticides, varieties). It is interesting to note that crop rotation is a regular part of agricultural practice in some wheat fields in Quebec. This practice can affect the composition of the soil microbiome. Among these, in Canadian canola production, the effects of crop rotation on soil and plant root-associated microbiomes are particularly significant (Town *et al.* 2023). Another study showed how different levels of nitrogen fertilization can greatly affect the recruitment process of the crop microbiome, particularly during the development of Canadian canola cultivars (Li *et al.* 2023b). It's important to recognize the significant impact of soil microbiome signatures on crop health and productivity, as they integrate with abiotic conditions that are influenced by agricultural practices. Regardless of the role of the soil microbiome in nutrient processing, excessive fertilizer applied to the soil results in high nitrogen loss through leaching, volatilization, eutrophication, and denitrification. Unfortunately, this is the case throughout Quebec, as many wheat farmers apply fertilizer indiscriminately (Vanasse 2012). As a result, overall bread wheat quality across the province may decline, especially if guidelines for proper fertilizer management are not established. Furthermore, there is no direct link between intensive fertilization and higher wheat yield and grain quality (Yergeau, Quiza and Tremblay 2020). Low-quality grain may be rejected by millers and can only be used for fodder, which is a loss for the farmer. Fortunately, my work shows that with the inclusion of soil microbial indicators associated with microbiome composition, structure, and function along with soil physicochemical parameters in soil testing, wheat yield and grain quality can be more accurately assessed.

My study was the first to reveal the potential role of the soil microbiome in agroecosystem processes in multiple wheat farms across a 500-km transect in the province of Quebec. I observed that model parameters aligned with some of our empirical findings, which was consistent with what we would expect given what we know about soil biochemical processes. For example, several model parameters related to fungal richness and higher microbial uptake of organic amino acids were positively associated with grain quality, suggesting that microbial processes associated with higher decomposition may make more nitrogen available to wheat by storing organic sources of nitrogen in the soil. (Chapter 2; Asad *et al.* 2021). Another indirect effect of the root-associated microbiome was observed in the regression model, through a negative relationship with grain quality. This effect suggested that bacterial activity might promote the growth of some wheat plants that continuously uptake high levels of subsidized inorganic nitrogen. Other smaller wheat plants might then be restricted to lower levels of nitrogen uptake, thereby reducing overall wheat quality (Chapter 2; Asad *et al.* 2021). We found that ammonia oxidizing bacteria (AOB) played an important role in predicting wheat yield through a univariate relationship between AOB and wheat grain quality, illustrating the importance of nitrifiers in soil nitrogen processing and plant nitrogen use. By tracking a potential indicator involved in nitrification at specific spatial scales, our model

may open a new method for co-occurrence mapping (Bru *et al.* 2011) of AOB or AOA across Canadian wheat farms. This can be useful for understanding the functional roles of AOB and AOA, developing landscape-level regulations, and estimating abundance. While this technique may not be useful for site-specific estimations of ammonia oxidation at the microbial community level due to differing microhabitats, there are several other contexts in which it can be used. Predictive modeling at specific spatial scales can provide a context-dependent, exploratory framework for the soil–microbiome–crop quality axis and contribute to future work on easily-accessible tools for farmers.

The specific time points associated with optimal soil microbiome composition and function, influencing the regulation of plant nitrogen, carbon, and protein metabolism. These changes in the soil microbiome may result in different predicted values for wheat grain quality. The soils microbiome is an important indicator of plant–microbe interaction, especially when plants rely on soil microbes to process nitrogen during times of starvation. This is usually expressed by the plant phenotype (Sessitsch, Pfaffenbichler and Mitter 2019). In fact, the composition of the soil microbiome can often be predicted based on the aboveground plant species (Mazza Rodrigues and Melotto 2023). Different plant species in specific soils (rhizosphere and bulk) harbor unique microbial communities. These microbial communities may have a distinct, taxa-specific relationship associated with the distinct traits of the plants. Examples of these microbes include mycorrhizal fungi, some fungal pathogens and nitrogen fixing bacteria (e.g., rhizobium) (Bright and Bulgheresi 2010; Wassermann *et al.* 2021). Microbial structure and function can also influence aboveground plant communities. However, the coexistence of specific microbes and particular plant species are sometimes context dependent. It takes years to demonstrate the effect of different vegetative phenotypes of plant species on microbial community composition (Kusstatscher *et al.* 2021). In our microbial-based predictive modeling of wheat grain quality relative to time, we demonstrate that there are different predictive accuracies for two wheat genotypes. Furthermore, these two genotypes have different associated microbiome compositions. Another interesting feature of the plant microbiome is it can interact with hosts through gene regulation and influence phenotypic traits. The plant microbiome can be inherited by the next generation of plants and dispersed to new plants from seed. Thus, the host-microbiome coevolutionary patterns are conserved and circulated throughout the host life cycle. (Abdelfattah *et al.* 2022) described that the inheritance process includes three main stages: 1) microbiome transfer from plant to seed, 2) seed to dormancy, and 3) seed to seeding. This endophytic microbiome assembly process is also highly influenced by the soil microorganisms and environment, determining microbiome structure and composition. The seed microbiome makeup is not directly dependent on horizontal gene transfer but by interactions with soil microbes that expand microbiome to other plant's compartment, shaping the plant microbiome at the seedling stage. This means that seeds from specific wheat genotypes may have microbiome compositions that each interact differently with the soil microbiome at

the seedling stage. This might subsequently affect the overall nitrogen acquisition and thus the grain nutrient content of the plant. It has been found that the yield metrics of some Canadian crop genotypes belonging to the Brassica family are closely related to the root and rhizosphere fungal microbiome (Li *et al.* 2023a). It appears that the root-associated microbiome can greatly influence crop productivity due to distinct patterns of interactions between plants and microbes under particular environmental conditions. Therefore, depending on the strength of the plant-soil microbiome interaction, farmers might be able to use grain quality information from predictive models to make decisions about which wheat varieties to seed in specific soil types. My work can help plant breeders focus more on microbiome-assisted breeding strategies that address future climate change issues.

Microbial dynamics were shown to change over time throughout the plant growth stage (Chapter 3). Some of the microbial patterns observed through predictive accuracy were shown to fluctuate during plant growth for the two wheat genotypes (GENOTYPE 1: Drought tolerant and GENOTYPE 2: Drought sensitive). This indicates that plant carbon sequestration through root exudation alters the root-associated microbiome on a small scale as well as the soil microbial guilds (Zhou *et al.* 2022). Other studies have shown that there is less microbial complexity and diversity at the beginning of crop growth compared to later stages of the plant life cycle, resulting in a lower rate of dispersal for particular microbial communities or functional guilds in the soil. Understanding this relationship, this pattern of microbiome at early crop growth stage can be used as microbial predictors or biomarkers for monitoring specific plant traits in wheat genotypes (Navarro-Noya *et al.* 2022). The potential mechanistic link between patterns of soil microbiome distribution and grain synthesis may depend on the pattern of plant–microbe interactions. These patterns of plant-microbe interactions are reflected in the fluctuations in soil microbiome composition with seasonal changes in plant growth, microbial functional abundance, and soil properties. The presence of ammonia-oxidizing archaea (AOA) in soil represents an indication of such interactions. One study observed an abundance of ammonia oxidizing archaea (AOA) (*Nitrososphaera*) along with other microbial genera at the beginning of the wheat-growing season, and the activity level of AOA (e.g., ammonia monooxygenase gene copy) remained quite stable throughout wheat growth (Chapter 3; Asad *et al.* 2023). The abundance of AOA was higher during the late wheat season after intermittent rainfall (Wang *et al.* 2022), as these communities are more dependent on the soil ammonium content to nitrate ratio and the condition of soil micro-habitats. Since the ammonia: nitrate ratio is environmentally related, the active functions of ammonia-oxidizing bacteria (AOB) and ammonia-oxidizing archaea (AOA) vary with environmental changes (Prosser and Nicol 2012). However, their distributions within the soil are quite different, and it was estimated that the abundance of AOA is 700 times higher than AOB in deep, anoxic soil (Leininger *et al.* 2006). This difference could potentially be a useful indicator of soil nitrification. Another interesting feature of ammonia-oxidizing microbial communities is that their efficiency increases with soil depth and

temperature, especially at the end of the growing season (Ouyang, Norton and Stark 2017). This phenomenon is consistent with findings obtained using our model parameters that identified large groups of microbiomes with more prominent of AOA. In this work, we observed an intensified organic carbon use pattern by soil microbiome in the late growing season, especially as the soil water content increased (Wang *et al.* 2022) which was more correlated with drought sensitive genotypes. These observations suggest that there is a unique microecology associated with the rhizosphere microbiome for the drought sensitive wheat genotype. This microecology might be influenced by specific plant genotypes with different nutrient uptake capacities.

We have found some important correlation between shallow sequencing reads, microbial community descriptors, abundance, wheat yield, and grain quality in the samples obtained from different locations and cultivation time periods. It's possible that the microbial indicators associated with crop yield quality are connected to the current processes of the agroecosystem, whether directly or indirectly. In Chapter 2 of our soil microbiome study, we discovered that particular ASVs or OTUs are associated with taxa that play a significant role in determining crop yield and quality. For example, we found a significant correlation between *Sphingomonas* sp., *Paenibacillus* sp., and various grain quality parameters (Chapter 2, Table 2.2) . It is known that such species have the ability to promote plant growth (Castanheira *et al.* 2017; Luo *et al.* 2019; Khan *et al.* 2020) and interact with crops to increase productivity. We have also found several ASVs that are closely linked to the taxa that can oxidize ammonia (i.e., *Nitrosospira*) and nitrite (i.e., *Nitrospira*) which have a detrimental effect on crop quality. It has been observed that the crop quality has a negative correlation with the dominant archaeal ammonia oxidizer *AOA*, as explained in Chapter 2, Section 2.5. Furthermore, Chapter 2, Table 3-4 highlights a significant inverse relationship between the abundance of ammonia oxidizer (*AOA* and *AOB*), nitrite reducer (*nirK*), and the quality of grain in the two specific wheat varieties.

Regardless of time and space, our study revealed a significant correlation between the abundance of microbial taxonomic and functional properties and crop quality during the early growth stages of wheat crops. We observed that excessive ammonia oxidation or nitrite reduction earlier may inhibit the availability of reactive nitrogen species to plants, resulting in lower crop quality in the future. Elucidating the complex relationship between soil microbial indices and crop yield and quality can be challenging, especially when considering positive or negative feedback from soil microbes. However, understanding the links between microbial indicators is important in developing precise and interpretable models, especially when selecting important variables for predicting crop yield and quality. Our models identified several microbial parameters that may influence ecosystem processes and revealed causal links. Nevertheless, some variables cannot fully explain their relationship with the process predicted in a linear model. It is plausible that these

variables may be linked to other soil microbiological factors that are associated with optimal agroecosystem processes for plant nutrient uptake and grain synthesis. While we have several hypotheses for how soil microbial ecological functions are linked with soil processes and plant traits, based on the results of modelling, a causal relationship for crop production requires more experimental proof. Therefore, the goal of my thesis was not to validate the causal link underlying the process of agroecosystem function, but to illustrate the predictive power of the core soil microbiome.

I applied metagenomic approaches to investigate soil microbiome diversity and function. I used specifically high throughput '16S and ITS amplicon sequencing' approaches, which are now widely considered to be the most reliable tools to explore microbiome features. These approaches enable close examination of the microbial taxonomic orientation with a resolution of just a few micrometers. We used some basic soil properties to analyse the soil physicochemical indicators that explain soil geochemical processes, as these properties are well-linked to the dispersion of soil microbial communities. For example, one can measure pH from soils collected at regional scales and observe a strong correlation with microbial community composition, making pH a good predictor for explaining microbial community composition (Lauber *et al.* 2009; Griffiths *et al.* 2011). However, this relationship is not as clear when soil samples are collected over smaller areas. As such, spatial scales and associated factors can also influence soil microbial community structure. This was considered when designing the second part of my study. The focus was to characterize microbial indices based on genomic methods rather than further quantifying soil properties within field samples from a small geographic area. Another advantage of using genomic-based methods is the ability to identify specific functional taxa that may affect specific agroecosystem processes, at a higher resolution. These functional taxa provide more information on the biotic and abiotic factors that mediate these processes. For example, identifying the specific taxa involved in ammonia oxidation (nitrification) can help us predict the rates of ammonia oxidation. Because taxa specifically involved in nitrification may exhibit different enzyme kinetics subject to different environmental constraints (Webster *et al.* 2005).

Information at the taxonomic level is not sufficient to predict the rates of geochemical processes in each ecosystem or the microbial processes that may change in response to environmental disturbances (e.g., climate change or land use) (Fierer 2017). Therefore, it is often suggested that other omics-based approaches such as transcriptomics or metabolomics be applied to examine microbial function at the gene level. However, microbially-driven soil geochemical processes are not the result of a single metabolic pathway, but rather the product of an integrated metabolic network that is governed by a wide range of microbial taxa (Fierer 2017). For example, heterotrophic microbial communities may contribute to the metabolism of organic matter or nitrogen mineralization. Microbial catabolism of organic carbon may require multiple metabolic processes that are carried out by a wide range of microbial taxa (Pepe-Ranney

et al. 2016). Analyzing the abundance of individual genes using DNA- or RNA-based methods from dormant or relatively inactive microorganisms (Blagodatskaya and Kuzyakov 2013) can link functional processes to specific microbial communities. However, there are a few obstacles that must be overcome, including inaccuracies associated with gene annotation (Schnoes *et al.* 2009), low taxonomic resolution, and rapid changes in transcriptomes, proteins, and metabolites in response to environmental changes (Moran *et al.* 2013). It is also important to note that this approach provides relative abundance of microbial taxa rather than absolute abundance. This is relevant because it has been suggested that soil processes are more influenced by the absolute number of taxa, genes, or gene products rather than to the relative composition of microbial taxa (Fierer 2017). Another limitation of taxa-specific functional predictions is the microbial use of different enzymatic cascades resulting in different substrate affinities, even in closed systems. For example, when cultivated in a laboratory, isolated strains of methane oxidizers have affinities for different substrates (Knief and Dunfield 2005). With these limitations in mind, our main objective was to use metagenomic approaches to identify microbial signals based on taxonomic abundance for wheat grain quality prediction.

Research on microbial communities has rapidly evolved with the advent of DNA sequencing technologies. DNA-based sequencing allows us to perform in-depth analyses of microbial structure, function, and interaction within specific ecosystems by generating billions of data points. However, new models that integrate microbiome data with omics data obtained from genomics, transcriptomics, metabolomics, and proteomics need to be further developed.

There is growing interest in multi-omics technologies for studying microbiomes, particularly in the context of soil metataxonomics, proteomics and metabolomics. Experts suggest that multi-omics technology requires a proper context-specific experimental design such as one that uses two integrated methods for monitoring specific soil processes together. For example, microbial taxonomic abundance and its relevant functional profile can be combined to provide comprehensive information. One study showed that three different soils might have the same taxonomic profile but different microbial processes involved in soil nutrient cycling (Ferrocino *et al.* 2023). In addition, Yao and colleagues (2018) (Yao *et al.* 2018) investigated microbial adaptation to P-deficient soil through an integrated metagenomics and proteomics approach and found that the genes and proteins of soil microbial communities exhibited adaptive responses to changes in nutrient limitations. Another comprehensive study using a multi-omics approach on microbial interactions in agroecosystems showed the complex relationship between soil metabolic, mineral, and microbial components (Ichihashi *et al.* 2020). In our study, we applied omics data integrated with statistical learning tools to interpret the relationship between the soil microbiome and its associated functional abundance and the effect on that affects the yield size and quality for wheat crops. Generally,

metataxonomic data are used to study microbiome composition or abundance in the laboratory or field, with or without specific treatments (Correa-garcia, Constant and Yergeau 2022). This information can provide researchers with a better understanding of the potential and value of meta-taxonomic data in microbiome-based predictive research.

It has been suggested that model interpretability helps users create a customized experimental setup to validate model-driven hypotheses (Correa-Garcia, Constant and Yergeau 2022). It has also been suggested that microbial predictors selected in microbial-based models can lead us to other questions related to microbial processes (Widder *et al.* 2016). For example, they hypothesized that some of the microbial predictors selected may not be directly related to microbial processes, but rather environmental factors that might mask the microbial processes. Other questions to be explored are whether some microbial predictors have causal relationships with the abundance of specific functional guilds (Asad *et al.* 2021). Lastly, some microbial predictors might shed light on the state of past ecosystem processes and how they contribute to certain current ecosystems. An interpretable model can identify powerful microbial parameters, which can then be manipulated through microbiome engineering (Agoussar and Yergeau 2021). For example, two separate studies showed that ammonia oxidizer was the main predictor in multiple linear regressions and random forest models, as these exhibited a negative relationship with grain quality at the end of the season (Yergeau, Quiza and Tremblay 2020; Asad *et al.* 2021). The negative effect of these microbial predictors supported an experimental design that inhibits nitrification by applying inhibitors that target ammonia oxidizers, with the hope of improving grain quality (Schmidt *et al.* 2022). In another experiment, authors used unsupervised learning to identify certain key rhizosphere microbial taxa early in tomato growth. These were shown to predict the future susceptibility of tomato plants to *Ralstonia solanacearum* wilt (Gu *et al.* 2022). Inoculation of healthy tomato plants with five bacterial isolates of those bacterial taxa reduced infection by 30–100%. These examples highlight some of the potential real-life applications of microbiome-based predictive modeling. These modelling approaches using omics-integrated statistical learning tools could allow for experiments to be customized for mechanistic understanding and suggest new directions in microbiome-based forecasting.

Soil samples were collected from 80 wheat farms that were registered with various local wheat-producing companies and utilized different farming techniques, both conventional and organic. However, the soil sampling was not evenly distributed based on the two agricultural practices. Conducting extensive soil sampling across the province is crucial to gain a better understanding of the spatial diversity and heterogeneity of soil microorganisms in agroecosystems. Furthermore, to ensure accurate monitoring of soil microbial processes that affect crop quality, it may be necessary to create separate predictive models for farms that use different agricultural practices. The study conducted did not measure all microbial genes

associated with nitrification and denitrification. To gain a better understanding of the impact of microbial diversity on the nitrogen cycle in Quebec soils, metagenomic or metatranscriptomic studies can be conducted to observe the variation of all potential genes related to the nitrogen cycle. In our research, we focused on analyzing the abundance of ammonia monooxygenase (*AOA* and *AOB*), nitrite reductase (*nirK*) and nitrous oxide reductase (*nosZ*), as they are directly linked to nitrogen loss (e.g., Nitrate leaching, N_20 and N_2 gas emission). This is important because the potential microbial activity linked to nitrogen loss may indirectly affect grain quality. Therefore, we initially measured a select few genes to determine if they have any relationship with grain quality. Using quantitative molecular biological assay (e.g., qPCR), we estimated the abundance of gene copies and assessed whether their inclusion in the models improved overall predictive performance. In the future, monitoring agricultural soil processes could benefit from utilizing these types of marker genes. Quantitative measurement of such functional genes can provide valuable insights in this regard. Furthermore, in case of our predictive modelling using microbiome data, we prioritized using representative sequencing reads, such as ASVs or OTUs, rather than relying solely on low taxonomic abundance data. This approach enabled us to better understand the changes of specific microbial communities and their association with specific ecosystem processes or differences in crop quality. Nevertheless, despite advances in current soil microbiome research, it remains a challenge to fully understand the unique functional properties of different taxa selected in statistical models that influence specific processes in complex agroecosystems. In order to confirm the efficacy of microbial parameters like ASVs, OTUs, or functional genes for future soil microbiome engineering or agricultural forecasting, it is crucial to carry out further field and laboratory experiments or large-scale surveys in diverse agricultural settings.

4.2 Conclusions

We need to explore new technologies to address existing and future sustainability issues related to the agrifood system. Research in soil microbiology is creating new avenues for addressing ongoing agricultural challenges such as productivity, fertilizer management and agricultural economics. From the laboratory to the field, it has been demonstrated that microbiome composition and function can be improved through manipulation. Soil microbiome-based solutions offer opportunities for creating more sustainable and resilient agricultural production systems (Wassermann, Müller and Berg 2019) and can be used to monitor crop quality by identifying the key functions that support agroecosystems. As such, the work presented in this thesis could contribute by helping to monitor soil microbiological parameters in specific regions at the beginning of the wheat growing season and improving soil fertility management through optimal fertilizer use. The conceptual framework based on microbiome modeling that I have introduced here can also help to validate experiments and lead to potential applications that can solve real-life

agricultural problems. For example, if we can measure a single microbial parameter such as the abundance of ammonia oxidizers (nitrification) early in the season, we may be able to intervene and reduce nitrification rates. This would help retain more nitrogen in the soil for plants and improve crop yields. My work could also be used to inform guidelines for the early application of soil nutrients (e.g., NPK) to promote wheat growth or increase wheat nutrient uptake. Microbiome-informed data derived from these predictive modeling tools can be used to support the sustainability of agricultural practices. Native microbiomes could be modified to improve the resilience of wheat to abiotic stress and high grain quality can be achieved by creating environments in which plants process soil nutrient more efficiently. Furthermore, integrated wheat breeding systems that consider microbiome modification by selecting specific wheat genotypes can generate high-quality, high-yield wheat production (Hohmann, Schlaeppi and Sessitsch 2020).

In conclusion, my thesis demonstrates, for the first time, the potential and maximum utility of microbial indicators through spatial and temporal studies at a regional scale in an agroecosystem. My research explains the role of microbial indicators as the third unmeasured parameter in the relationship between intensive N fertilization and wheat yield through microbe dependent predictive modeling. It also shows that the inclusion of microbial parameters in soil nutrient testing leads to better fertilization management, which can improve wheat yield and quality. Microbe-related parameters measured at the beginning of the wheat-growing season can inform decision-making about agronomic practices or fertilizer application and help improve microbiome function, increasing soil fertility. Furthermore, I have shown that microbial indicators derived from the core soil microbiome structure, function and composition across time and space may provide better signals to predict future agroecosystem processes and wheat yield and grain quality. I have developed microbial predictive models based on two different wheat genotypes that clearly show that breeding or seeding decisions based on soil microbial parameters could result in better wheat qualities. I found it fascinating that the interaction between wheat genotype and soil microbes can vary over time and greatly impact the quality of wheat grain. By utilizing genotype-specific models to gather microbial predictors, we can gain a deeper understanding of the complex relationships between plant and soil microbiomes in the early stages of crop growth. This insight can be highly beneficial in predicting the crop yield at harvest and managing the future conditions of agricultural ecosystem processes. When choosing genotypes for wheat breeding, it's crucial to take into account the host-microbiome composition and the wide range of interactions between various wheat genotypes, soil, and other microorganisms within a particular environment. Making informed decisions based on soil microbiome can greatly improve wheat genotype selection and breeding. This approach can lead to higher yields and better-quality crops by selectively cultivating varieties that can fight plant pathogens and inhibit microbes that compete with plants for nutrients. Furthermore, we have discovered specific microbial parameters that can serve as catalysts to augment the assimilation of nitrogen by crops. For example, nitrogen loss for excessive nitrification caused

by the increased activity of ammonia oxidizers (e.g., AOA and AOB) can be controlled by using chemical inhibitors or planting selective crops and plants with biological nitrification-inhibiting properties. I also found in several models that the soil fungal-to-bacterial ratio appeared as a potential biological parameter because the balance of soil fungal and bacterial biomass has a conserved role in maintaining soil carbon-nitrogen dynamics. This finding indicates that the relative abundance of fungi plays a significant role in releasing more nitrogen, even when the amount of organic matter and microbial abundance remains constant. Taken together, these findings open the door to new possibilities in sustainable agriculture, particularly in fertilization management, and in harnessing microbial functions by modulating soil microorganisms to enhance soil productivity. I hope that my research on microbial-based predictive modeling contributes to soil microbiome-based solutions for improving soil health and fertilizer management for optimal wheat production.

5 BIBLIOGRAPHY

Abdelfattah A, Tack AJM, Berg G *et al.* From seed to seed : the role of microbial inheritance in the assembly of the plant microbiome. *Trends in Microbiology* 2022;xx:1–10.

Adair K, Schwartz E. Stable isotope probing with 18O-water to investigate growth and mortality of ammonia oxidizing bacteria and archaea in soil. *Methods Enzymol* 2011; 486:155–69.

Agoussar A, Azarbad H, Tremblay J *et al.* The resistance of the wheat microbial community to water stress is more influenced by plant compartment than reduced water availability. *FEMS Microbiology Ecology* 2021; 97:1–11.

Agoussar A, Yergeau E. Engineering the plant microbiota in the context of the theory of ecological communities. *Current Opinion in Biotechnology* 2021;70:220–5.

Aguilar-Trigueros CA, Hempel S, Powell JR *et al*. Branching out: Towards a trait-based understanding of fungal ecology. *Fungal Biology Reviews* 2015; 29:34–41.

Aksoy MA, Beghin JC. *Global Agricultural Trade and Developing Countries*. World Bank Publications, 2004.

Albright MBN, Johansen R, Thompson J *et al.* Soil Bacterial and Fungal Richness Forecast Patterns of Early Pine Litter Decomposition. *Frontiers in Microbiology* 2020;11, DOI: 10.3389/fmicb.2020.542220.

Ali B, Sabri AN, Hasnain S. Rhizobacterial potential to alter auxin content and growth of Vigna radiata (L.). *World Journal of Microbiology and Biotechnology* 2010; 26:1379–84.

Ali Y, Qin A, Aatif HM *et al.* A stepwise multiple regression model to predict *Fusarium* wilt in lentil. *Meteorological Applications* 2022;29, DOI: 10.1002/met.2088.

Allison SD, Treseder KK. Warming and drying suppress microbial activity and carbon cycling in boreal forest soils. *Global Change Biology* 2008; 14:2898–909.

Ames NP, Clarke JM, Dexter JE *et al.* Effects of Nitrogen Fertilizer on Protein Quantity and Gluten Strength Parameters in Durum Wheat (*Triticum turgidum* L. var. *durum*) Cultivars of Variable Gluten Strength. *Cereal Chemistry Journal* 2003;80:203–11.

Arkhipova TN, Veselov SU, Melentiev AI *et al.* Ability of bacterium Bacillus subtilis to produce cytokinins and to influence the growth and endogenous hormone content of lettuce plants. *Plant and Soil* 2005; 272:201–9.

Artursson V, Finlay RD, Jansson JK. Interactions between arbuscular mycorrhizal fungi and bacteria and their potential for stimulating plant growth. *Environmental Microbiology* 2006 ;8 :1–10.

Asad NI, Tremblay J, Dozois J *et al.* Predictive microbial-based modelling of wheat yields and grain baking quality across a 500 km transect in Quebec. *FEMS Microbiol Ecol.* 2021 ; 97(2) : fiab160.

Asad NI, Wang X, Dozois J *et al.* Early season soil microbiome best predicts wheat grain quality *FEMS Microbiol Ecol.* 2023:1–13.

Averill C, Cates LL, Dietze MC *et al.* Spatial vs. temporal controls over soil fungal community similarity at continental and global scales. *The ISME Journal* 2019; 13:2082–93.

Averill C, Waring BG, Hawkes C V. Historical precipitation predictably alters the shape and magnitude of microbial functional response to soil moisture. *Global Change Biology* 2016;22:1957–64.

Ayoub M, Guertin S, Fregeau-Reid J *et al.* Nitrogen fer- tilizer effect on breadmaking quality of hard red spring wheat in Eastern Canada. *Crop Sci* 1994 ;34. DOI: 10.2135/cropsci1994.0011183×003400050038x.

Azarbad H, Bainard L, Agoussar A *et al.* The response of wheat and its microbiome to contemporary and historical water stress in a field experiment *ISME Communications* 2022; 2: 62.

Azarbad H, Constant P, Giard-Laliberté C *et al.* Water stress history and wheat genotype modulate rhizosphere microbial response to drought. *Soil Biology and Biochemistry* 2018; 126:228–36.

Azarbad H, Tremblay J, Giard-Laliberté C *et al.* Four decades of soil water stress history together with host genotype constrain the response of the wheat microbiome to soil moisture. *FEMS Microbiol Ecol* 2020;96: fiaa098.

B. Sohn M, Li H. A GLM-based latent variable ordination method for microbiome samples. *Biometrics* 2018;74:448–57.

Babin D, Deubel A, Jacquiod S *et al.* Impact of long-term agricul- tural management practices on soil prokaryotic communi- ties. *Soil Biol Biochem* 2019; 129:17–28.

Bais HP, Weir TL, Perry LG *et al.* The role of root exudates in rhizosphere interactions with plants and other organisms. *Annual Review of Plant Biology* 2006; 57:233–66.

Banerjee S, van der Heijden MGA. Soil microbiomes and one health. *Nature Reviews Microbiology*

2022;19, DOI: 10.1038/s41579-022-00779-w.

Barahona E, Navazo A, Martínez-Granero F *et al.* Pseudomonas fluorescens F113 Mutant with Enhanced Competitive Colonization Ability and Improved Biocontrol Activity against Fungal Root Pathogens. *Applied and Environmental Microbiology* 2011; 77:5412–9.

Bargmann I, Martens R, Rillig MC *et al.* Hydrochar amendment promotes microbial immobilization of mineral nitrogen. *Journal of Plant Nutrition and Soil Science* 2014; 177:59–67.

Barnard RL, Osborne CA, Firestone MK. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *The ISME Journal* 2013; 7:2229–41.

Barneix AJ. Physiology and biochemistry of source-regulated protein accumulation in the wheat grain. *Journal of Plant Physiology* 2007; 164:581–90.

Baudoin E, Benizri E, Guckert A. Impact of artificial root exudates on the bacterial community structure in bulk soil and maize rhizosphere. *Soil Biology and Biochemistry* 2003; 35:1183–92.

Beeckman F, Motte H, Beeckman T. Nitrification in agricultural soils: impact, actors and mitigation. *Current Opinion in Biotechnology* 2018;50:166–73.

Bell TH, Cloutier-Hurteau B, Al-Otaibi F *et al.* Early rhizosphere microbiome composition is related to the growth and Zn uptake of willows introduced to a former landfill. *Environ Microbiol* 2015; 17:3025–38.

Bell TH, El-Din Hassan S, Lauron-Moreau A *et al.* Linkage between bacterial and fungal rhizosphere communities in hydrocarbon-contaminated soils is related to plant phylogeny. *ISME Journal* 2014; 8:331–43.

Bell TH, Yergeau E, Maynard C *et al.* Predictable bacterial composition and hydrocarbon degradation in Arctic soils following diesel and nutrient disturbance. *ISME Journal* 2013; 7:1200–10.

Berendsen RL, Pieterse CMJ, Bakker PAHM. The rhizosphere microbiome and plant health. *Trends in Plant Science* 2012; 17:478–86.

Berg G. Plant--microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. *Applied Microbiology and Biotechnology* 2009; 84:11–8.

Björk RG, Klemedtsson L, Molau U *et al.* Linkages between N turnover and plant community structure in a tundra landscape. *Plant and Soil* 2007; 294:247–61.

Blagodatskaya E, Kuzyakov Y. Active microorganisms in soil: Critical review of estimation criteria and approaches. *Soil Biology and Biochemistry* 2013;67:192–211.

Blanchet M, Pringault O, Bouvy M *et al.* Changes in bacterial community metabolism and composition during the degradation of dissolved organic matter from the jellyfish Aurelia aurita in a Mediterranean coastal lagoon. *Environmental Science and Pollution Research* 2015; 22:13638–53.

Blazewicz SJ, Schwartz E, Firestone MK. Growth and death of bacteria and fungi underlie rainfall-induced carbon dioxide pulses from seasonally dried soil. *Ecology* 2014; 95:1162–72.

Bodelier PLE, Steenbergh AK. Interactions between methane and the nitrogen cycle in light of climate change. *Current Opinion in Environmental Sustainability* 2014, DOI: 10.1016/j.cosust.2014.07.004.

Bodenhausen N, Bortfeld-Miller M, Ackermann M *et al.* A Synthetic Community Approach Reveals Plant Genotypes Affecting the Phyllosphere Microbiota. *PLoS Genetics* 2014;10, DOI: 10.1371/journal.pgen.1004283.

Bodenhausen N, Horton MW, Bergelson J. Bacterial Communities Associated with the Leaves and the Roots of Arabidopsis thaliana. *PLoS ONE* 2013;8, DOI: 10.1371/journal.pone.0056329.

Bogard M, Allard V, Brancourt-Hulmel M *et al.* Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. *Journal of Experimental Botany* 2010; 61:4303–12.

Bonfante P, Genre A. Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications* 2010;1:48.

Bonkowski M, Brandt F. Do soil protozoa enhance plant growth by hormonal effects? *Soil Biology and Biochemistry* 2002;34:1709–15.

Bouché F, D'Aloia M, Tocquin P *et al.* Integrating roots into a whole plant network of flowering time genes in Arabidopsis thaliana. *Scientific Reports* 2016 ;6 :29042.

Bourceret A, Guan R, Dorau K *et al.* Maize Field Study Reveals Covaried Microbiota and Metabolic Changes in Roots over Plant Growth. *mBio* 2022;13, DOI: 10.1128/mbio.02584-21.

Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology* 2010;8:218–30.

Broeckling CD, Broz AK, Bergelson J *et al*. Root Exudates Regulate Soil Fungal Community Composition and Diversity. *Applied and Environmental Microbiology* 2008; 74:738–44.

Bru D, Ramette A, Saby NPA *et al.* Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *The ISME Journal* 2011;5:532–42.

Bruce KD, Hiorns WD, Hobman JL *et al.* Amplification of DNA from native populations of soil bacteria

by using the polymerase chain reaction. *Appl Environ Microbiol* 1992;58:3413 LP–6.

Bulgarelli D, Garrido-Oter R, Münch PC *et al.* Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host and Microbe* 2015; 17:392–403.

Butterbach-Bahl K, Baggs EM, Dannenmann M *et al.* Nitrous oxide emissions from soils: how well do we understand the processes and their controls? *Philosophical Transactions of the Royal Society B: Biological Sciences* 2013;368:20130122.

Calder ´on K, Spor A, Breuil MC *et al.* Effectiveness of ecological rescue for altered soil microbial communities and functions. *ISME J* 2017 ;11 :272–83.

Callahan BJ, McMurdie PJ, Rosen MJ *et al.* DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016 ; 13 :581–3.

Callens K, Fontaine F, Sanz Y *et al.* Microbiome-based solutions to address new and existing threats to food security, nutrition, health and agrifood systems' sustainability. *Frontiers in Sustainable Food Systems* 2022;6, DOI: 10.3389/fsufs.2022.1047765.

Callens K, Fontaine F, Sanz Y *et al.* Microbiome-based solutions to address new and existing threats to food security, nutrition, health and agrifood systems' sustainability. *Frontiers in Sustainable Food Systems* 2022;6, DOI: 10.3389/fsufs.2022.1047765.

Caporaso JG, Lauber CL, Walters WA *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal* 2012 ;6 :1621–4.

Cardinale M, Grube M, Erlacher A *et al.* Bacterial networks and co-occurrence relationships in the lettuce root microbiota. *Environmental Microbiology* 2015; 17:239–52.

Caro-Quintero A, Konstantinidis KT. Bacterial species may exist, metagenomics reveal. *Environmental Microbiology* 2012 ;14 :347–55.

Carrillo Y, Bell C, Koyama A *et al.* Plant traits, stoichiometry and microbes as drivers of decomposition in the rhizosphere in a temperate grassland. *Journal of Ecology* 2017;105:1750–65.

Castanheira NL, Dourado AC, Pais I *et al.* Colonization and beneficial effects on annual ryegrass by mixed inoculation with plant growth promoting bacteria. *Microbiological Research* 2017;**198**:47–55

Castillo P, Escalante M, Gallardo M *et al.* Effects of bacterial single inoculation and co-inoculation on growth and phytohormone production of sunflower seedlings under water stress. *Acta Physiologiae Plantarum* 2013 ; 35:2299–309.

Chable V, Lemichez S, Hohmann P *et al.* Report on the holobiont as promising selection target to

improve resilience and product quality. 2021.

Chang H-X, Haudenshield JS, Bowen CR *et al.* Metagenome-Wide Association Study and Machine Learning Prediction of Bulk Soil Microbiome and Crop Productivity. *Frontiers in Microbiology* 2017;8, DOI: 10.3389/fmicb.2017.00519.

Chaparro JM, Badri D V., Bakker MG *et al.* Root Exudation of Phytochemicals in Arabidopsis Follows Specific Patterns That Are Developmentally Programmed and Correlate with Soil Microbial Functions. *PLoS One* 2013; 8:1–10.

Chaparro JM, Badri D V., Vivanco JM. Rhizosphere microbiome assemblage is affected by plant development. *ISME Journal* 2014 ;8 :790–803.

Chemidlin Prévost-Bouré N, Dequiedt S, Thioulouse J *et al.* Similar processes but different environmental filters for soil bacterial and fungal community composition turnover on a broad spatial scale. *PLoS ONE* 2014;9: e111667.

Chen D, Cheng J, Chu P *et al.* Regional-scale patterns of soil microbes and nematodes across grasslands on the Mongolian plateau: relationships with climate, soil, and plants. *Ecography* 2015; 38:622–31.

Chen L, Gao Y. Global Climate Change Effects on Soil Microbial Biomass Stoichiometry in Alpine Ecosystems. *Land* 2022;11, DOI: 10.3390/land11101661.

Chen S, Waghmode TR, Sun R *et al.* Root-associated microbiomes of wheat under the combined effect of plant development and nitrogen fertilization. *Microbiome* 2019;7:136.

Chernin L, Ismailov Z, Haran S *et al.* Chitinolytic *Enterobacter agglomerans* Antagonistic to Fungal Plant Pathogens. Applied and Environmental Microbiology 1995; 61:1720–6.

Chung OK, Pomeranz Y, Finney KF. Wheat flour lipids in breadmaking. *Cereal Chem* 1978;55:598–618.

Clarholm M. Bacteria and protozoa as integral components of the forest ecosystem--their role in creating a naturally varied soil fertility. *Antonie van Leeuwenhoek* 2002; 81:309–18.

Classen AT, Sundqvist MK, Henning JA *et al.* Direct and indirect effects of climate change on soil microbial and soil microbial-plant interactions: What lies ahead? *Ecosphere* 2015;6: art130.

Clemmensen KE, Bahr A, Ovaskainen O *et al.* Roots and Associated Fungi Drive Long-Term Carbon Sequestration in Boreal Forest. *Science* 2013; 339:1615–8.

Correa-Garcia S, Constant P, Yergeau E. The forecasting power of the microbiome. *Trends in Microbiology* 2022, DOI: 10.1016/j.tim.2022.11.013.

Cruz de Carvalho MH. Drought stress and reactive oxygen species. *Plant Signaling & Behavior* 2008; 3:156–65.

de la Porte A, Schmidt R, Yergeau É *et al.* A Gaseous Milieu: Extending the Boundaries of the Rhizosphere. *Trends in Microbiology* 2020;28:536–42.

de Vargas C, Audic S, Henry N *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* 2015;348, DOI: 10.1126/science.1261605.

de Vries FT, Griffiths RI, Bailey M *et al.* Soil bacterial networks are less stable under drought than fungal networks. *Nature Communications* 2018; 9:3033.

Deka H, Deka S, Baruah CK. Plant Growth Promoting Rhizobacteria for Value Addition: *Mechanism of Action* 2015, 305–21.

Dequiedt S, Thioulouse J, Jolivet C *et al.* Biogeographical patterns of soil bacterial communities. *Environmental Microbiology Reports* 2009; 1:251–5.

Dimkpa C. Microbial siderophores: Production, detection and application in agriculture and environment Microbial siderophores: Production, detection and application in agriculture and environment. 2016.

Dombrowski N, Schlaeppi K, Agler MT *et al.* Root microbiota dynamics of perennial Arabis alpina are dependent on soil residence time but independent of flowering time. *ISME Journal* 2017 ;11:43–55.

Dong M, Li L, Chen M *et al.* Predictive analysis methods for human microbiome data with application to Parkinson's disease. *PLoS One* 2020; 15:1–18.

Donn S, Kirkegaard JA, Perera G *et al.* Evolution of bacterial communities in the wheat crop rhizosphere. *Environmental Microbiology* 2015 ;17:610–21.

Edgar RC, Haas BJ, Clemente JC *et al.* UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011; 27:2194–200.

Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* 2013;**10**:996–8.

Edwards J, Johnson C, Santos-Medellín C *et al.* Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences* 2015;112: E911–20.

Edwards U, Rogall T, Blöcker H *et al.* Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research* 1989;**17**:7843–53.

Ellermann M, Arthur JC. Siderophore-mediated iron acquisition and modulation of host-bacterial interactions. *Free radical biology & medicine* 2017; 105:68–78.

Enebe MC, Babalola OO. Soil fertilization affects the abundance and distribution of carbon and nitrogen cycling genes in the maize rhizosphere. *AMB Express* 2021; 11:24.

Engelbrektson A, Kunin V, Engelbrektson A *et al*. Defining the core Arabidopsis thaliana root microbiome. *Nature* 2012; 488:86–90.

Evans SE, Wallenstein MD. Climate change alters ecological strategies of soil bacteria. *Ecology Letters* 2014; 17:155–64.

Fan K, Delgado-Baquerizo M, Guo X *et al.* Suppressed N fixation and diazotrophs after four decades of fertilization. *Microbiome* 2019; 7:143.

Farhat MR, Sultana R, Iartchouk O *et al.* Genetic Determinants of Drug Resistance in *Mycobacterium tuberculosis* and Their Diagnostic Value. *American Journal of Respiratory and Critical Care Medicine* 2016;194:621–30.

Faucon MP, Houben D, Lambers H. Plant Functional Traits: Soil and Ecosystem Services. *Trends in Plant Science* 2017; 22:385–94.

Ferrocino I, Rantsiou K, Lange L *et al.* The need for an integrated multi-OMICs approach in microbiome science in the food system. 2023:1–22.

Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* 2017;15:579–90.

Figuerola ELM, Guerrero LD, Rosa SM *et al.* Bacterial Indicator of Agricultural Management for Soil under No-Till Crop Production. *PLoS One* 2012; 7:1–12.

Fisher K, Newton WE. Nitrogen Fixation: An Historical Perspective. In: Smith BE, Richards RL, Newton WE (eds.). Catalysts for Nitrogen Fixation: Nitrogenases, Relevant Chemical Models and Commercial Processes. *Dordrecht: Springer Netherlands*, 2004, 1–31.

Fitriani wangsa putrie R, Tri wahyudi A, Asih Nawangsih A *et al.* Screening of Rhizobacteria for Plant Growth Promotion and Their Tolerance to Drought Stress. *Microbiology Indonesia* 2013;7:94–104.

Flavel TC, Murphy D V. Carbon and Nitrogen Mineralization Rates after Application of Organic Amendments to Soil. *Journal of Environmental Quality* 2006; 35:183–93.

Franklin RB, Mills AL. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiology Ecology* 2003; 44:335–46.

Fraterrigo JM, Rusak JA. Disturbance-driven changes in the variability of ecological patterns and processes. *Ecology Letters* 2008; 11:756–70.

French E, Kaplan I, Iyer-Pascuzzi A *et al.* Emerging strategies for precision microbiome management in diverse agroecosystems. *Nature Plants* 2021 ;7 :256–67.

Freund W, Kim MY. Determining the Baking Quality of Wheat and Rye Flour. *Future of Flour – A compendium of Flour Improvement* 2006:101–16.

Friedman J, Hastie T, Simon N *et al.* Package 'glmnet.' *Journal of statistical software* 2017;33:1–22.

Fry EL, De Long JR, Álvarez Garrido L *et al.* Using plant, microbe, and soil fauna traits to improve the predictive power of biogeochemical models. *Methods in Ecology and Evolution* 2019; 10:146–57.

Funk JL, Larson JE, Ames GM *et al.* Revisiting the Holy Grail: Using plant functional traits to understand ecological processes. *Biological Review*s 2017; 92:1156–73.

Geisseler D, Horwath WR, Joergensen RG *et al.* Pathways of nitrogen utilization by soil microorganisms – A review. *Soil Biology and Biochemistry* 2010; 42:2058–67.

Gianinazzi S, Schüepp H. Impact of Arbuscular Mycorrhizas on Substainable Agriculture and Natural Ecosystems. *Springer Science & Business Media*, 1994.

Giauque H, Hawkes C V. Historical and current climate drive spatial and temporal patterns in fungal endophyte diversity. *Fungal Ecology* 2016 ;20:108–14.

Glavina T, Jones CD, Lebels SI *et al.* Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* 2015; 349:860–4.

Goel S, Singh M, Grewal S *et al.* Wheat Proteins: A Valuable Resources to Improve Nutritional Value of Bread. *Frontiers in Sustainable Food Systems* 2021;5, DOI: 10.3389/fsufs.2021.769681.

Goodswen SJ, Barratt JLN, Kennedy PJ *et al.* Machine learning and applications in microbiology. *FEMS Microbiology Reviews* 2021;45:1–19.

Graham EB, Wieder WR, Leff JW *et al.* Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes. *Soil Biology and Biochemistry* 2014; 68:279–82.

Griffiths RI, Thomson BC, James P *et al.* The bacterial biogeography of British soils. *Environmental Microbiology* 2011;13:1642–54.

Grimvall A. changes in four societal drivers and their potential to reduce Swedish nutrient inputs into the sea. 2016.

Grobelak A, Napora A, Kacprzak M. Using plant growth-promoting rhizobacteria (PGPR) to improve plant growth. *Ecological Engineering* 2015; 84:22–8.

Grzyb A, Wolna-Maruwka A, Niewiadomska A. The Significance of Microbial Transformation of Nitrogen Compounds in the Light of Integrated Crop Management. *Agronomy* 2021;11:1415.

Gu S, Wei Z, Shao Z *et al.* Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. *Nature Microbiology* 2020, DOI: 10.1038/s41564-020-0719-8.

Gu Y, Banerjee S, Dini-Andreote F *et al.* Small changes in rhizosphere microbiome composition predict disease outcomes earlier than pathogen density variations. *The ISME Journal* 2022;16:2448–56.

Guarda G, Padovan S, Delogu G. Grain yield, nitrogen-use effi- ciency and baking quality of old and modern Italian bread- wheat cultivars grown at different nitrogen levels. *Eur J Agron* 2004; 21:181–92.

Gubry-Rangin C, Nicol GW, Prosser JI. Archaea rather than bacteria control nitrification in two agricultural acidic soils. *FEMS Microbiol Ecol* 2010 ;74:566–74.

Gururani MA, Upadhyaya CP, Baskar V *et al.* Plant growth-promoting rhizobacteria enhance abiotic stress tolerance in solanum tuberosum through inducing changes in the expression of ROS-scavenging enzymes and improved photosynthetic performance. *Journal of Plant Growth Regulation* 2013; 32:245–58.

Halfeld-Vieira B de A, Vieira Júnior JR, Romeiro R da S *et al.* Induction of systemic resistance in tomato by the autochthonous phylloplane resident Bacillus cereus. *Pesquisa Agropecuária Brasileira* 2006; 41:1247–52.

Hannula SE, Morriën E, van der Putten WH *et al.* Rhizosphere fungi actively assimilating plant-derived carbon in a grassland soil. *Fungal Ecology* 2020;48, DOI: 10.1016/j.funeco.2020.100988.

Hannula SE, Zhu F, Heinen R *et al.* Foliar-feeding insects acquire microbiomes from the soil rather than the host plant. *Nature Communications* 2019;10:1254.

Hartman K, van der Heijden MGA, Wittwer RA *et al.* Cropping practices manipulate abundance patterns of root and soil microbiome members paving the way to smart farming. *Microbiome* 2018; 6:14.

Hawkesford MJ. Reducing the reliance on nitrogen fertilizer for wheat production. *Journal of Cereal Science* 2014; 59:276–83.

Haynes RJ. The decomposition process: Mineralization, immobilization, humus formation. *Mineral nitrogen in the plant-soil systems* 1986:52–126.

Henry S, Bru D, Stres B *et al.* Quantitative detection of the nosZ gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, narG, nirK, and nosZ genes in soils. *Applied and Environmental Microbiology* 2006;72:5181–9.

Herrera Paredes S, Gao T, Law TF *et al.* Design of synthetic bacterial communities for predictable plant phenotypes. *PLoS Biology* 2018;16, DOI: 10.1371/journal.pbio.2003962.

Hills L, Id BZ, Id JH *et al.* Contrast in soil microbial metabolic func- tional diversity to fertilization and crop rotation under rhizo- sphere and non-rhizosphere in the coal gangue landfill reclamation area of Loess Hills. *PLoS ONE* 2020. DOI: 10.1371/journal. pone.0229341.

Hohmann P, Schlaeppi K, Sessitsch A. miCROPe 2019 – emerging research priorities towards microbe-assisted crop production. *FEMS Microbiology Ecology* 2020;96, DOI: 10.1093/femsec/fiaa177.

Horemans B, Breugelmans P, Saeys W *et al.* Soil-bacterium compatibility model as a decision-making tool for soil bioremediation. *Environmental Science and Technology* 2017; 51:1605–15.

Hu HW, Macdonald CA, Trivedi P *et al.* Water addition regulates the metabolic activity of ammonia oxidizers responding to environmental perturbations in dry subhumid ecosystems. *Environ Microbiol* 2015; 17:444–61.

Hucl P, Briggs C, Shirtliffe S *et al.* Increasing grain yield while maintaining baking quality in Canada Western Red Spring wheat. *Canadian Journal of Plant Science* 2022;102:973–83.

Hünninghaus M, Koller R, Kramer S *et al.* Changes in bacterial community composition and soil respiration indicate rapid successions of protist grazers during mineralization of maize crop residues. *Pedobiologia* 2017; 62:1–8.

Husson F, Josse J, Le S *et al.* Package 'factominer.' *An R package* 2016;96:698.

Ichihashi Y, Date Y, Shino A *et al.* Multi-omics analysis on an agroecosystem reveals the significant role of organic nitrogen to increase agricultural crop yield. *Proceedings of the National Academy of Sciences* 2020;117:14552–60.

James G, Witten D, Hastie T *et al. An Introduction to Statistical Learning*. Springer, 2013.

Janssens IA, Dieleman W, Luyssaert S *et al.* Reduction of forest soil respiration in response to nitrogen deposition. *Nature Geoscience* 2010;3:315–22.

Jeanne T, Parent S-É, Hogue R. Using a soil bacterial species balance index to estimate potato crop productivity. *Plos One* 2019;14: e0214089.

Jones DL, Kielland K. Soil amino acid turnover dominates the nitrogen flux in permafrost-dominated

taiga forest soils. *Soil Biol Biochem* 2002; 34:209–19.

Kanters C, Anderson IC, Johnson D. Chewing up the wood-wide web: Selective grazing on ectomycorrhizal fungi by collembola. Forests 2015; 6:2560–70.

Kassambara A, Kassambara MA. Package 'ggpubr.' *R package version 01* 2020;6.

Kassambara A. Package 'rstatix.' *R topics documented* 2020.

Kaushal M, Wani SP. Plant-growth-promoting rhizobacteria: drought stress alleviators to ameliorate crop production in drylands. *Annals of Microbiology* 2016; 66:35–42.

Kavamura VN, Hayat R, Clark IM *et al.* Inorganic nitrogen application affects both taxonomical and predicted functional structure of wheat rhizosphere bacterial communities. *Frontiers in Microbiology* 2018; 9:1–15.

Khan MS, Gao J, Chen X *et al.* Isolation and Characterization of Plant Growth-Promoting Endophytic Bacteria *Paenibacillus polymyxa* SK1 from *Lilium lancifolium*. *BioMed Research International* 2020;2020:1–17.

Khdhiri M, Hesse L, Elena M *et al.* Soil carbon content and relative abundance of high affinity H_2 -oxidizing bacteria predict atmospheric H_2 soil uptake activity better than soil microbial community composition. *Soil Biology & Biochemistry* 2015; 85:1–9.

Khoiri AN, Cheevadhanarak S, Jirakkakul J *et al.* Comparative Metagenomics Reveals Microbial Signatures of Sugarcane Phyllosphere in Organic Management. *Frontiers in Microbiology* 2021;12, DOI: 10.3389/fmicb.2021.623799.

Kim SY, Veraart AJ, Meima-franke M *et al.* Geoderma Combined effects of carbon, nitrogen and phosphorus on CH 4 production and denitri fi cation in wetland sediments. *Geoderma* 2015;259–260:354–61.

Knief C, Dunfield PF. Response and adaptation of different methanotrophic bacteria to low methane mixing ratios. *Environmental Microbiology* 2005;7:1307–17.

Kong AYY, Hristova K, Scow KM *et al.* Impacts of different N management regimes on nitrifier and denitrifier communities and N cycling in soil microenvironments. *Soil Biology and Biochemistry* 2010; 42:1523–33.

Kooijman AM, Mourik JM Van, Schilder MLM. The relationship between N mineralization or microbial biomass N with micromorphological properties in beech forest soils with different texture and pH. 2009:449–59.

Kostic AD, Gevers D, Siljander H *et al.* The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes. *Cell Host & Microbe* 2015 ;17:260–73.

Kour D, Rana KL, Kaur T *et al.* Microbe-mediated alleviation of drought stress and acquisition of phosphorus in great millet (Sorghum bicolour L.) by drought-adaptive and phosphorus-solubilizing microbes. *Biocatalysis and Agricultural Biotechnology* 2020; 23:101501.

Kourtev PS, Ehrenfeld JG, Häggblom M. Experimental analysis of the effect of exotic and native plant species on the structure and function of soil microbial communities. *Soil Biology and Biochemistry* 2003; 35:895–905.

Kowalchuk GA, Stephen JR. Ammonia- Oxidizing Bacteria: *A Model for Molecular Microbial Ecology*. 2001:485–529.

Kracmarova M, Uhlik O, Strejcek M *et al.* Soil microbial communities following 20 years of fertilization and crop rotation practices in the Czech Republic. *Environmental Microbiome* 2022;17:13.

Kramer J, Özkaya Ö, Kümmerli R. Bacterial siderophores in community and host interactions. *Nature Reviews Microbiology* 2020; 18:152–63.

Krome K, Rosenberg K, Dickler C *et al.* Soil bacteria and protozoa affect root branching via effects on the auxin and cytokinin balance in plants. *Plant and Soil* 2010;328:191–201.

Kuenen JG. Anammox bacteria: from discovery to application. *Nature Reviews Microbiology* 2008; 6:320.

Kuramae EE, Zhou JZ, Kowalchuk GA *et al.* Soil-borne microbial functional structure across different land uses. *Scientific World Journal* 2014;2014, DOI: 10.1155/2014/216071.

Kusstatscher P, Adam E, Wicaksono WA *et al.* Microbiome-Assisted Breeding to Understand Cultivar-Dependent Assembly in Cucurbita pepo. *Frontiers in Plant Science* 2021;12, DOI: 10.3389/fpls.2021.642027.

L´opez-Bellido L, L ´opez-Bellido RJ, Castillo JE *et al.* Effects of long-term tillage, crop rotation and nitrogen fertilization on bread-making quality of hard red spring wheat. *Field Crops Res* 2001; 72:197–210.

La Favre JS, Focht DD. Conservation in Soil of H_2 Liberated from N_2 Fixation by Hup- Nodules. *Applied and environmental microbiology* 1983.

Lauber CL, Hamady M, Knight R *et al.* Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Applied and Environmental*

Microbiology 2009;75:5111–20.

Leilei X, Baohua X, Jinchao L *et al.* Science of the Total Environment Stimulation of long-term ammonium nitrogen deposition on methanogenesis by *Methanocellaceae* in a coastal wetland. *Science of the Total Environment* 2017; 595:337–43.

Leininger S, Urich T, Schloter M *et al.* Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 2006;442:806–9.

Lesk C, Rowhani P, Ramankutty N. Influence of extreme weather disasters on global crop production. *Nature* 2016; 529:84–7.

Levy-Booth DJ, Prescott CE, Grayston SJ. Microbial functional genes involved in nitrogen fixation, nitrification and denitrification in forest ecosystems. *Soil Biology and Biochemistry* 2014;75:11–25.

Li J, Delgado-Baquerizo M, Wang J-T *et al.* Fungal richness contributes to multifunctionality in boreal forest soil. *Soil Biology and Biochemistry* 2019; 136:107526.

Li J, Guo C, Jian S *et al.* Nitrogen fertilization elevated spatial heterogeneity of soil microbial biomass carbon and nitrogen in switchgrass and gamagrass croplands. *Scientific Reports* 2018;8:1–16.

Li J, Wang J, Singh BK *et al.* Application of microbial inoculants significantly enhances crop productivity: A meta-analysis of studies from 2010 to 2020. 2022:1–10.

Li X, Li Z, Zhang X *et al.* Disentangling immobilization of nitrate by fungi and bacteria in soil to plant residue amendment. *Geoderma* 2020; 374:114450.

Li Y, Bazghaleh N, Vail S *et al.* Root and rhizosphere fungi associated with the yield of diverse Brassica napus genotypes. *Rhizosphere* 2023a;25:100677.

Li Y, Vail SL, Arcand MM *et al.* Contrasting Nitrogen Fertilization and *Brassica napus* (Canola) Variety Development Impact Recruitment of the Root-Associated Microbiome. *Phytobiomes Journal* 2023b;**7**:125–37.

Lindsay EA, Colloff MJ, Gibb NL *et al.* The Abundance of Microbial Functional Genes in Grassy Woodlands Is Influenced More by Soil Nutrient Enrichment than by Recent Weed Invasion or Livestock Exclusion. *Applied and Environmental Microbiology* 2010; 76:5547 LP – 5555.

Liu DY, Ding WX, Jia ZJ *et al.* Relation between methanogenic archaea and methane production potential in selected natural wetland ecosystems across China. 2011:329–38.

LiuX, LiQ, LiY *et al.* Paenibacillus strains with nitrogen fixation and multiple beneficial properties for promoting plant growth. *Peer J* 2019 ;7 : e7445.

Longepierre M, Widmer F, Keller T *et al.* Limited resilience of the soil microbiome to mechanical compaction within four growing seasons of agricultural management. *ISME Communications* 2021 ;1 :44.

Luo Y, Wang F, Huang Y *et al.* Sphingomonas sp. Cra20 Increases Plant Growth Rate and Alters Rhizosphere Microbial Community Structure of Arabidopsis thaliana Under Drought Stress. *Frontiers in Microbiology* 2019;**10**, DOI: 10.3389/fmicb.2019.01221.

Maimaiti J, Zhang Y, Yang J *et al.* Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. *Environmental Microbiology* 2007 ; 9 :435–44.

Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T *et al.* Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. *Frontiers in Microbiology* 2021;12, DOI: 10.3389/fmicb.2021.634511.

Marschner P, Rengel Z. Chapter 12 - nutrient availability in soils. In: Petra M (ed). Marschner's Mineral Nutrition of Higher Plants (Third Edition). *Academic Press*, 2012, 315–30.

Martin KJ, Rygiewicz PT. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC Microbiol* 2005; 5:28.

Marulanda A, Barea J-M, Azcón R. Stimulation of Plant Growth and Drought Tolerance by Native Microorganisms (AM Fungi and Bacteria) from Dry Environments: Mechanisms Related to Bacterial Effectiveness. *Journal of Plant Growth Regulation* 2009; 28:115–24.

Matocha CJ, Dhakal P, Pyzola SM. Chapter Four - The Role of Abiotic and Coupled Biotic/Abiotic Mineral Controlled Redox Processes in Nitrate Reduction. In: Sparks DLBT-A in A (ed.). Vol 115. *Academic Press*, 2012, 181–214.

Mazza Rodrigues JL, Melotto M. Naturally engineered plant microbiomes in resource-limited ecosystems. *Trends in Microbiology* 2023:1–3.

McHugh TA, Compson Z, Gestel N van *et al.* Climate controls prokaryotic community composition in desert soils of the southwestern United States. *FEMS Microbiology Ecology* 2017;93, DOI: 10.1093/femsec/fix116.

McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* 2013;8:e61217.

Medina RH, Kutuzova S. Machine learning and deep learning applications in microbiome research.

2022:1–7.

Meisner A, Jacquiod S, Snoek BL *et al.* Drought legacy effects on the composition of soil fungal and prokaryote communities. *Frontiers in Microbiology* 2018; 9:1–12.

Melillo JM, Frey SD, DeAngelis KM *et al.* Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science* 2017; 358:101–5.

Meyer KM, Porch R, Muscettola IE *et al.* Plant neighborhood shapes diversity and reduces interspecific variation of the phyllosphere microbiome. *The ISME Journal* 2022;16:1376–87.

Midani FS, Weil AA, Chowdhury F *et al.* Human Gut Microbiota Predicts Susceptibility to Vibrio cholerae Infection. *The Journal of Infectious Diseases* 2018;218:645–53.

Moitinho-Silva L, Steinert G, Nielsen S *et al.* Predicting the HMA-LMA Status in Marine Sponges by Machine Learning. Frontiers in Microbiology 2017;8, DOI: 10.3389/fmicb.2017.00752.

Moran MA, Satinsky B, Gifford SM *et al.* Sizing up metatranscriptomics. *The ISME Journal* 2013;7:237–43.

Moreau D, Bardgett RD, Finlay RD *et al.* A plant perspective on nitrogen cycling in the rhizosphere. *Funct Ecol* 2019;33: 540–52.

Moroenyane I, Tremblay J, Yergeau É. Temporal and spatial interactions modulate the soybean microbiome. *FEMS Microbiol Ecol* 2021; 97:1–12.

Morriën E, Hannula SE, Snoek LB *et al.* Soil networks become more connected and take up more carbon as nature restoration progresses. *Nature Communications* 2017; 8:14349.

Mummey D, Holben W, Six J *et al.* Spatial Stratification of Soil Bacterial Populations in Aggregates of Diverse Soils. *Microbial Ecology* 2006; 51:404–11.

Naseem H, Ahsan M, Shahid MA *et al.* Exopolysaccharides producing rhizobacteria and their role in plant growth and drought tolerance. *Journal of Basic Microbiology* 2018; 58:1009–22.

Navarro-Noya YE, Chávez-Romero Y, Hereira-Pacheco S *et al.* Bacterial Communities in the Rhizosphere at Different Growth Stages of Maize Cultivated in Soil Under Conventional and Conservation Agricultural Practices. *Microbiol Spectr* 2022;10: e01834-21.

Nelson AG, Quideau S, Frick B *et al.* Spring wheat genotypes differentially alter soil microbial communities and wheat breadmaking quality in organic and conventional systems. *Canadian Journal of Plant Science* 2011;91:485–95.

Nelson MB, Berlemont R, Martiny AC *et al.* Nitrogen Cycling Potential of a Grassland Litter Microbial

Community. Kostka JE (ed.). *Applied and Environmental Microbiology* 2015;81:7012 LP – 7022.

Newlands NK, Zamar DS, Kouadio LA *et al.* An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. *Frontiers in Environmental Science* 2014;2, DOI: 10.3389/fenvs.2014.00017.

Ngumbi E, Kloepper J. Bacterial-mediated drought tolerance: Current and future prospects. *Applied Soil Ecology* 2016; 105:109–25.

Nguyen BAT, Dumack K, Trivedi P *et al.* Plant associated protists—Untapped promising candidates for agrifood tools. *Environmental Microbiology* 2022 :1–12.

Nguyen TH, Chevaleyre Y, Prifti E *et al*. Deep Learning for Metagenomic Data: using 2D Embeddings and Convolutional Neural Networks. 2017.

Nilsson RH, Larsson KH, Taylor AFS *et al.* The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res* 2019;47: D259–64.

O'Sullivan CA, Fillery IRP, Roper MM *et al.* Identification of several wheat landraces with biological nitrification inhibition capacity. *Plant Soil* 2016;404: 61-74.

Oh M, Pruden A, Chen C *et al.* MetaCompare: a computational pipeline for prioritizing environmental resistome risk. *FEMS Microbiology Ecology* 2018;94, DOI: 10.1093/femsec/fiy079.

Oksanen J, Blanchet FG, Kindt R *et al.* Package 'vegan.' *Community ecology package, version* 2013;2:1–295.

Olanrewaju OS, Glick BR, Babalola OO. Mechanisms of action of plant growth promoting bacteria. *World Journal of Microbiology and Biotechnology* 2017; 33:197.

Ortiz A, Sansinenea E. The Role of Beneficial Microorganisms in Soil Quality and Plant Health. *Sustainability (Switzerland)* 2022 ;14, DOI: 10.3390/su14095358.

Ouyang Y, Norton JM, Stark JM. Ammonium availability and temperature control contributions of ammonia oxidizing bacteria and archaea to nitrification in an agricultural soil. *Soil Biology and Biochemistry* 2017;113:161–72.

Oyetunde T, Liu D, Martin HG *et al.* Machine learning framework for assessment of microbial factory performance. *PLoS ONE* 2019;14:e0210558.

P. White E, B. Adler P, K. Lauenroth W *et al.* A comparison of the species-time relationship across ecosystems and taxonomic groups. *Oikos* 2006; 112:185–95.

Pajares S, Bohannan BJM. Ecology of Nitrogen Fixing, Nitrifying, and Denitrifying Microorganisms in *Tropical Forest Soils.* 2016; 7:1–20.

Parent LE, Jamaly R, Atucha A *et al.* Current and next-year cranberry yields predicted from local features and carryover effects. *PLoS ONE* 2021;16:1–16.

Parnell JJ, Berka R, Young HA *et al.* From the Lab to the Farm: An Industrial Perspective of Plant Beneficial Microorganisms. *Frontiers in Plant Science* 2016; 7:1–12.

Patten CL, Glick BR. Bacterial biosynthesis of indole-3-acetic acid. *Canadian Journal of Microbiology* 1996; 42:207–20.

Pausch J, Kramer S, Scharroba A *et al.* Small but active - pool size does not matter for carbon incorporation in below-ground food webs. *Functional Ecology* 2016; 30:479–89.

Paz-Ferreiro J, Fu S. Biological indices for soil quality evaluation: perspectives and limitations. *Land Degrad* Dev 2016; 27:14–25.

Pechanek U, Karger A, Gröger S *et al.* Effect of Nitrogen Fertilization on Quantity of Flour Protein Components, Dough Properties, and Breadmaking Quality of Wheat. *Cereal Chemistry Journal* 1997; 74:800–5.

Pepe-Ranney C, Campbell AN, Koechli CN *et al.* Unearthing the Ecology of Soil Microorganisms Using a High Resolution DNA-SIP Approach to Explore Cellulose and Xylose Metabolism in Soil. *Frontiers in Microbiology* 2016;7, DOI: 10.3389/fmicb.2016.00703.

Pignataro A, Moscatelli MC, Mocali S *et al.* Assessment of soil microbial functional diversity in a coppiced forest system. *Applied Soil Ecology* 2012; 62:115–23.

Poore GD, Kopylova E, Zhu Q *et al.* Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* 2020;579:567–74.

Prosser JI, Nicol GW. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends in Microbiology* 2012;20:523–31.

Putten WH, Bradford MA, Pernilla Brinkman E *et al.* Where, when and how plant–soil feedback matters in a changing world. *Functional Ecology* 2016; 30:1109–21.

Qian X, Gu J, Pan HJ et al. Effects of living mulches on the soil nutrient contents, enzyme activities, and bacterial community diversities of apple orchard soils. *European Journal of Soil Biology* 2015; 70:23–30.

Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data

processing and web-based tools. *Nucleic Acids Res* 2013;41: D590–6.

Quiza L, St-Arnaud M, Yergeau E. Harnessing phytomicrobiome signalling for rhizosphere microbiome engineering. *Front Plant Sci* 2015 ;6 :507.

Quiza L, Tremblay J, Pagé AP *et al.* The effect of wheat genotype on microbiome composition is more evident in roots than rhizosphere and is strongly influenced by time. *ISME comms.* 2022; submitted.

Raaijmakers JM, Mazzola M. Diversity and Natural Functions of Antibiotics Produced by Beneficial and Plant Pathogenic Bacteria. *Annual Review of Phytopathology* 2012; 50:403–24.

Ramirez KS, Knight CG, de Hollander M *et al.* Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nature Microbiology* 2017;3:189–96.

Ranjard L, Poly F, Combrisson J *et al.* Heterogeneous Cell Density and Genetic Structure of Bacterial Pools Associated with Various Soil Microenvironments as Determined by Enumeration and DNA Fingerprinting Approach (RISA). *Microbial Ecology* 2000; 39:263–72.

Rasche F, Cadisch G. The molecular microbial perspective of organic matter turnover and nutrient cycling in tropical agroecosystems - What do we know? *Biology and Fertility of Soils* 2013;49:251–62.

Rashid MI, Mujawar LH, Shahzad T *et al.* Bacteria and fungi can contribute to nutrients bioavailability and aggregate formation in degraded soils. *Microbiological Research* 2016;183:26–41.

Reiman D, Metwally AA, Sun J *et al.* PopPhy-CNN: A Phylogenetic Tree Embedded Architecture for Convolutional Neural Networks to Predict Host Phenotype From Metagenomic Data. *IEEE Journal of Biomedical and Health Informatics* 2020;24:2993–3001.

Řezáčová V, Czakó A, Stehlík M *et al.* Organic fertilization improves soil aggregation through increases in abundance of eubacteria and products of arbuscular mycorrhizal fungi. *Scientific Reports* 2021; 11:12548.

Rillig MC, Muller LA, Lehmann A. Soil aggregates as massively concurrent evolutionary incubators. The *ISME Journal* 2017; 11:1943–8.

Robertson GP, Groffman PM. Nitrogen Transformations: *Soil Microbiology, Ecology and Biochemistry*; Chapter 14, Denitrification. 2015.

Romero-Olivares AL, Allison SD, Treseder KK. Soil microbes and their response to experimental warming over time: A meta-analysis of field studies. *Soil Biology and Biochemistry* 2017; 107:32–40.

Rothschild D, Leviatan S, Hanemann A *et al.* An atlas of robust microbiome associations with phenotypic traits based on large-scale cohorts from two continents. *PLoS ONE* 2022;17:e0265756.

Rousk J, Bååth E. Fungal biomass production and turnover in soil estimated using the acetate-in-ergosterol technique. *Soil Biology and Biochemistry* 2007; 39:2173–7.

S´anchez ´A, Vila JCC, Chang C-Y *et al.* Directed evolution ofmicro- bial communities. *Annu Rev Biophys 2021*; 50:323–41.

Saleem M, Fetzer I, Dormann CF *et al.* Predator richness increases the effect of prey diversity on prey yield. *Nat Commun* 3: 1305. 2012.

Santoyo G. How plants recruit their microbiome? New insights into beneficial interactions. *Journal of Advanced Research* 2022;40:45–58.

Schimel J, Balser TC, Wallenstein M. Microbial stress-response physiology and its implications for ecosystem function. *Ecology* 2007; 88:1386–94.

Schindlbacher A, Rodler A, Kuffner M *et al.* Experimental warming effects on the microbial community of a temperate mountain forest soil. *Soil Biology and Biochemistry* 2011; 43:1417–25.

Schloter M, Nannipieri P, Sørensen SJ *et al.* Microbial indicators for soil quality. *Biol. Fert. Soils.* 2018: 54:1-10.

Schmidt JE, Kent AD, Brisson VL *et al.* Agricultural management and plant selection interactively affect rhizosphere microbial community structure and nitrogen cycling. *Microbiome* 2019; 7:146.

Schmidt R, Wang XB, Garbeva P *et al.* The nitrification inhibitor nitrapyrin has non-target effects on the soil microbial community structure, composition, and functions. *Applied Soil Ecology* 2022;171, DOI: 10.1016/j.apsoil.2021.104350.

Schnoes AM, Brown SD, Dodevski I *et al.* Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Computational Biology* 2009;5:e1000605.

Schulz R, Makary T, Hubert S *et al.* Is it necessary to split nitrogen fertilization for winter wheat? On-farm research on Luvisols in South-West Germany. *The Journal of Agricultural Science* 2015; 153:575–87.

Sengupta A, Fansler SJ, Chu RK *et al.* Disturbance triggers non-linear microbe-environment feedbacks. *Biogeosciences* 2021;18:4773–89.

Sessitsch A, Pfaffenbichler N, Mitter B. Microbiome Applications from Lab to Field: Facing Complexity. *Trends in Plant Science* 2019;24:194–8.

Shaharoona B, Arshad M, Zahir ZA. Effect of plant growth promoting rhizobacteria containing ACC-deaminase on maize (*Zea mays L.*) growth under axenic conditions and on nodulation in mung bean (*Vigna radiata L.*). *Letters in Applied Microbiology* 2006; 42:155–9.

Sharma D, Paterson AD, Xu W. TaxoNN: ensemble of neural networks on stratified microbiome data for disease prediction. *Bioinformatics* 2020;36:4544–50.

Shawy LJ, Burns RG. *Enzyme Activity Profiles and Soil Quality*., 2009.

Shen X, Hu H, Peng H *et al.* Comparative genomic analysis of four representative plant growth-promoting rhizobacteria in Pseudomonas. *BMC Genomics 2013*; 14:271.

Sherr BF, Sherr EB, Berman T. Grazing, Growth, and Ammonium Excretion Rates of a Heterotrophic Microflagellate Fed with Four Species of Bacteria. *Applied and Environmental Microbiology* 1983;45:1196–201.

Sheth RU, Cabral V, Chen SP *et al.* Manipulating bacterial com- munities by in situ microbiome engineering. *Trends Genet* 2016;32:189–200.

Shi Y, Xiang X, Shen C *et al.* Vegetation-Associated Impacts on Arctic Tundra Bacterial and Microeukaryotic Communities. *Applied and Environmental Microbiology* 2015; 81:492–501.

Simon J, Klotz MG, Nap N. Biochimica et Biophysica Acta Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations ☆. *BBA - Bioenergetics* 2013;1827:114–35.

Sitaula BK, Hansen S, Sitaula JIB *et al.* Effects of soil compaction on N_2O emission in agricultural soil. Chemosphere - *Global Change Science* 2000b; 2:367–71.

Sitaula BK, Hansen S, Sitaula JIB *et al.* Methane oxidation potentials and fluxes in agricultural soil:Effects of fertilisation and soil compaction. *Biogeochemistry* 2000;48:323–39.

Six J, Feller C, Denef K et al. Soil organic matter, biota and aggregation in temperate and tropical soils - Effects of no-tillage. *Agronomie* 2002; 22:755–75.

Skinner C, Gattinger A, Krauss M *et al*. The impact of long-term organic farming on soil-derived greenhouse gas emissions. *Scientific Reports* 2019; 9:1702.

Sorensen PO, Templer PH, Finzi AC. Contrasting effects of winter snowpack and soil frost on growing season microbial biomass and enzyme activity in two mixed-hardwood forests. *Biogeochemistry* 2016; 128:141–54.

Spaepen S, Vanderleyden J, Remans R. Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiology Reviews* 2007; 31:425–48.

Sparacino-Watkins C, Stolz JF, Basu P. Nitrate and periplasmic nitrate reductases. *Chemical Society Reviews* 2014; 43:676–706.

St. Luce M, Whalen JK, Ziadi N *et al.* Chapter two - Nitrogen Dynamics and Indices to Predict Soil Nitrogen Supply in Humid Temperate Soils. In: Sparks DLBT-A in A (ed.). Vol 112. *Academic Press*, 2011, 55–102.

Šťovíček A, Kim M, Or D *et al.* Microbial community response to hydration-desiccation cycles in desert soil. *Scientific Reports* 2017; 7:45735.

Subbarao, G. V, Kishii, M., Bozal-leorri, A., Ortiz-monasterio, I., & Gao, X. (2021). *Enlisting wild grass genes to combat nitrification in wheat farming: A nature-based solution*. 1–9. https://doi.org/10.1073/pnas.2106595118/-/DCSupplemental.Published

Sun A, Jiao X-Y, Chen Q *et al.* Microbial communities in crop phyllosphere and root endosphere are more resistant than soil microbiota to fertilization. *Soil Biology and Biochemistry* 2021; 153:108113.

Sun S, Li S, Avera BN *et al.* Soil Bacterial and Fungal Communities Show Distinct Recovery Patterns during Forest Ecosystem Restoration. *Applied and Environmental Microbiology* 2017;83, DOI: 10.1128/AEM.00966-17.

Techen A-K, Helming K, Brüggemann N *et al.* Soil research challenges in response to emerging agricultural soil management practices. 2020, 179–240.

Tiessen H, Cuevas E, Chacon P. The role of soil organic matter in sustaining soil fertility. *Nature* 1994; 371:783–5.

Tin Kam Ho. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition 1995*. IEEE Comput. Soc. Press, 278–82.

Ton J, Van Pelt JA, Van Loon LC *et al.* Differential Effectiveness of Salicylate-Dependent and Jasmonate/Ethylene-Dependent Induced Resistance in *Arabidopsis*. *Molecular Plant-Microbe Interactions*® 2002;15:27–34.

Torrecilha RBP, Utsunomiya YT, Bosco AM *et al.* Correlations between peripheral parasite load and common clinical and laboratory alterations in dogs with visceral leishmaniasis. *Preventive Veterinary Medicine* 2016;132:83–7.

Tourna M, Freitag TE, Nicol GW *et al.* Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environmental Microbiology* 2008;10:1357–64.

Town JR, Dumonceaux T, Tidemann B *et al.* Crop rotation significantly influences the composition of soil, rhizosphere, and root microbiota in canola (Brassica napus L.). *Environmental Microbiome* 2023;**18**:40.

Tremblay J, Yergeau E. Systematic processing of ribosomal RNA gene amplicon sequencing data. *Giga Science* 2019 ;8: giz146.

Treseder KK, Berlemont R, Allison SD *et al.* Drought increases the frequencies of fungal functional genes related to carbon and nitrogen acquisition. *PLoS ONE* 2018 ;13: e0206441.

Trivedi P, Delgado-Baquerizo M, Jeffries TC *et al.* Soil aggregation and associated microbial communities modify the impact of agricultural management on carbon content. *Environmental Microbiology* 2017; 19:3070–86.

Trivedi P, Leach JE, Tringe SG *et al.* Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology* 2020;18:607–21.

Ukalska-Jaruga A, Siebielec G, Siebielec S *et al.* The impact of exogenous organic matter on wheat growth and mineral nitrogen availability in soil. Agronomy 2020; 10:1314.

Unger S, Máguas C, Pereira JS *et al.* The influence of precipitation pulses on soil respiration – Assessing the "Birch effect" by stable carbon isotopes. *Soil Biology and Biochemistry* 2010; 42:1800–10.

Van der Putten WH. Climate Change, Aboveground-Belowground Interactions, and Species' Range Shifts. *Annual Review of Ecology, Evolution, and Systematics* 2012;43:365–83.

Vanasse A. Les céréales à paille. Québec (Québec) : Centre de référence en agriculture et agroalimentaire du Québec (CRAAQ), 2012.

Vannette RL, Fukami T. Dispersal enhances beta diversity in nectar microbes. *Ecology Letters* 2017;20:901–10.

Vardharajula S, Zulfikar Ali S, Grover M *et al.* Drought-tolerant plant growth promoting Bacillus spp.: effect on growth, osmolytes, and antioxidant status of maize under drought stress. *Journal of Plant Interactions* 2011; 6:1–14.

Wagner MR, Roberts JH, Balint-Kurti P *et al.* Heterosis of leaf and rhizosphere microbiomes in field-grown maize. *New Phytologist* 2020;228:1055–69.

Walters WA, Jin Z, Youngblut N *et al.* Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proceedings of the National Academy of Sciences of the United States of America* 2018; 115:7368–73.

Wan X, Huang Z, He Z *et al.* Soil C:N ratio is the major determinant of soil microbial community structure in subtropical coniferous and broadleaf forest plantations. *Plant Soil* 2015; 387:103–16.

Wang C, Mao G, Liao K *et al.* Machine learning approach identifies water sample source based on microbial abundance. *Water Research* 2021a; 199:117185.

Wang J, Wang X, Liu G *et al.* Bacterial richness is negatively related to potential soil multifunctionality in a degraded alpine meadow. *Ecological Indicators* 2021b; 121:106996.

Wang Q, Garrity GM, Tiedje JM *et al.* Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* 2007; 73:5261 LP – 5267.

Wang X-B, Azarbad H, Leclerc L *et al.* A Drying-Rewetting Cycle Imposes More Important Shifts on Soil Microbial Communities than Does Reduced Precipitation. *mSystems* 2022;7, DOI: 10.1128/msystems.00247-22.

Wardle DA, Bardgett RD, Klironomos JN *et al.* Ecological Linkages Between Aboveground and Belowground Biota. *Science* 2004; 304:1629–33.

Wassermann B, Müller H, Berg G. An Apple a Day: Which Bacteria Do We Eat With Organic and Conventional Apples? *Frontiers in Microbiology* 2019;10, DOI: 10.3389/fmicb.2019.01629.

Wassermann B, Rybakova D, Adam E *et al.* Studying Seed Microbiomes. 2021, 1–21.

Webster G, Embley TM, Freitag TE *et al.* Links between ammonia oxidizer species composition, functional diversity and nitrification kinetics in grassland soils. *Environmental Microbiology* 2005;7:676–84.

Wen G, Anne BM, Claude V *et al.* Machine learning-based canola yield prediction for site- specific nitrogen recommendations. *Nutrient Cycling in Agroecosystems* 2021;121:241–56.

Widder S, Allen RJ, Pfeiffer T *et al.* Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME Journal* 2016;10:2557–68.

Wieder WR, Bonan GB, Allison SD. Global soil carbon projections are improved by modelling microbial processes. *Nature Climate Change* 2013; 3:909–12.

Wilhelm RC, van Es HM, Buckley DH. Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biology and Biochemistry* 2021;164:108472.

Wilpiszeski RL, Aufrecht JA, Retterer ST *et al.* Soil Aggregate Microbial Communities: Towards Understanding Microbiome Interactions at Biologically Relevant Scales. *Applied and Environmental Microbiology* 2019;85, DOI: 10.1128/AEM.00324-19.

Winkelmann G. Ecology of siderophores with special reference to the fungi. *Bio Metals* 2007; 20:379–92.

Wittwer RA, Bender SF, Hartman K *et al.* Organic and conservation agriculture promote ecosystem multifunctionality. *Science Advances* 2021 ;7, DOI: 10.1126/sciadv. abg6995.

Worldwide RCT and C. Package'Stats'. 2020.

Xiao J, Chen L, Yu Y *et al.* A Phylogeny-Regularized Sparse Regression Model for Predictive Modeling of Microbial Community Data. *Frontiers in Microbiology* 2018;9:1–12.

Xiong C, He J, Singh BK *et al.* Rare taxa maintain the stability of crop mycobiomes and ecosystem functions. *Environmental Microbiology* 2021a; 23:1907–24.

Xiong C, Lu Y. Microbiomes in agroecosystem: Diversity, function and assembly mechanisms. *Environmental Microbiology Reports* 2022;14:833–49.

Xiong C, Zhu Y, Wang J *et al.* Host selection shapes crop microbiome assembly and network complexity. *New Phytologist* 2021b; 229:1091–104.

Xu J, Li X-L, Luo L. Effects of Engineered Sinorhizobium meliloti on Cytokinin Synthesis and Tolerance of Alfalfa to Extreme Drought Stress. *Applied and Environmental Microbiology* 2012; 78:8056–61.

Xu L, Coleman-Derr D. Causes and consequences of a conserved bacterial root microbiome response to drought stress. *Current Opinion in Microbiology* 2019; 49:1–6.

Xu Y, Xu Z, Cai Z. Review of denitrification in tropical and subtropical soils of terrestrial ecosystems. 2013:699–710.

Xu Z, Zhang T, Wang S et al. Soil pH and C/N ratio determines spatial variations in soil microbial communities and enzymatic activities of the agricultural ecosystems in Northeast China: Jilin Province case. *Applied Soil Ecology* 2020 ; 155:103629.

Yao Q, Li Z, Song Y *et al.* Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil. *Nature Ecology & Evolution* 2018;2:499–509.

Ye R, Jin Q, Bohannan B *et al.* Soil Biology & Biochemistry pH controls over anaerobic carbon mineralization, the efficiency of methane production , and methanogenic pathways in peatlands across an ombrotrophic eminerotrophic gradient. *Soil Biology and Biochemistry* 2012;54:36–47.

Yergeau E, Bell TH, Champagne J *et al.* Transplanting soil microbiomes leads to lasting effects on willow growth, but not on the rhizosphere microbiome. *Frontiers in Microbiology* 2015;6:1–14.

Yergeau E, Bezemer TM, Hedlund K *et al.* Influences of space, soil, nematodes, and plants on microbial community composition of chalk grassland soils. *Environmental Microbiology* 2010a ;12:2096–106.

Yergeau E, Labour K, Hamel C *et al.* Patterns of Fusarium community structure and abundance in relation to spatial, abiotic and biotic factors in soil. *FEMS Microbiology Ecology* 2010b, DOI: 10.1111/j.1574-6941.2009.00777. x.

Yergeau É, Quiza L, Tremblay J. Microbial indicators are better predictors of wheat yield and quality than N fertilization. *FEMS Microbiology Ecology* 2020;96:1–13.

Yergeau E, Sanschagrin S, Maynard C *et al.* Microbial expression profiles in the rhizosphere of willows depend on soil contamination. *The ISME Journal* 2014; 8:344–58.

Yergeau E, Tremblay J, Joly S *et al.* Soil contamination alters the willow root and rhizosphere metatranscriptome and the root-rhizosphere interactome. *ISME Journal* 2018 ;12:869–84.

Yu X, Chen X, Wang L *et al.* Novel insights into the effect of nitrogen on storage protein biosynthesis and protein body development in wheat caryopsis. *Journal of Experimental Botany* 2017; 68:2259–74.

Yuan MM, Guo X, Wu L *et al.* Climate warming enhances microbial network complexity and stability. *Nature Climate Change* 2021;11:343–8.

Zhalnina K, Dias R, de Quadros PD *et al.* Soil pH Determines Microbial Diversity and Composition in the Park Grass Experiment. *Microbial Ecology* 2015;69:395–406.

Zhan C, Matsumoto H, Liu Y *et al.* Pathways to engineering the phyllosphere microbiome for sustainable crop production. *Nature Food* 2022;3:997–1004.

Zhang J, Zhang N, Liu Y-X *et al*. Root microbiota shift in rice correlates with resident time in the field and developmental stage. *Science China Life Sciences* 2018; 61:613–21.

Zhou X, Wang J, Liu F *et al.* Cross-kingdom synthetic microbiota supports tomato suppression of Fusarium wilt disease. *Nature Communications* 2022;13:7890.

Zhu T, Meng T, Zhang J *et al.* Nitrogen mineralization, immobilization turnover, heterotrophic nitrification, and microbial groups in acid forest soils of subtropical China. *Biology and Fertility of Soils* 2013; 49:323–31.

Zörb C, Ludewig U, Hawkesford MJ. Perspective on Wheat Yield and Quality with Reduced Nitrogen Supply. *Trends in Plant Science* 2018;23:1029–37.

6. ANNEX 1: PUBLICATION SUPPLEMENTARY MATERIALS FOR CHAPTER 2

Table S1: Metadata for wheat field surveys across Quebec wheat farms. Region, wheat variety sowed, geographical coordinates and soil analyses (if available) for the 80 fields sampled.

Sample	Region	Variety	Latitude (°N)	Longitude (°W)	pH	Water content (%)	Total C (%)	Total N (%)	C:N ratio
10	Lac.St-Jean	Variety trial	47.59	67.01					
2	Lac.St-Jean	Variety trial	47.59	67.01					
24A	Monteregie	Warthog	45.14	73.30	6.92	0.232	2.12	0.18	11.80
24B	Monteregie	Warthog	45.15	73.31	6.97	0.189	3.82	0.36	10.60
24D	Monteregie	Warthog	45.14	73.31	5.77	0.198	4.07	0.38	10.70
3	Lac.St-Jean	Variety trial	47.59	67.01					
4	Lac.St-Jean	Variety trial	47.59	67.01					
5	Lac.St-Jean	Variety trial	47.59	67.01					
6	Lac.St-Jean	Variety trial	47.59	67.01					
7	Lac.St-Jean	Variety trial	47.59	67.01					
8	Lac.St-Jean	Variety trial	47.59	67.01					
9	Lac.St-Jean	Variety trial	47.59	67.01					
BON225	Mauricie	Walton	46.01	73.24	6.53	0.143			
BOU48	Monteregie	Walton	45.12	73.82	5.01	0.138			
CEROM11	Mauricie	Walton	45.35	73.15	7.13	0.156			
EBRER33	Monteregie	Walton	45.17	73.12	6.31	0.401	2.65	0.21	12.60
EBRER5	Monteregie	Walton	45.16	73.13	6.47	0.224	2.32	0.19	12.20
EBRER6	Monteregie	Walton	45.17	73.13	6.65	0.202	2.50	0.21	11.90
EBRER7	Monteregie	Walton	45.16	73.13	6.81	0.204	1.07	0.10	10.70
Fongiminus1	Monteregie	Walton	45.12	73.82					
Fongiminus2	Monteregie	Walton	45.12	73.82					
Fongiminus3	Monteregie	Walton	45.12	73.82					
Fongiplus1	Monteregie	Walton	45.12	73.82					
Fongiplus2	Monteregie	Walton	45.12	73.82					
Fongiplus3	Monteregie	Walton	45.12	73.82					
GDF1	Lanaudiere	Dakosta	46.83	73.17		0.145			

GDF2	Lanaudiere	Dakosta	46.84	73.17		0.161			
GDF3	Lanaudiere	Dakosta	46.85	73.16		0.141			
IJ	Monteregie	Walton	45.11	73.56	6.14	0.133	1.82	0.17	10.70
JAM1	Lanaudiere	Harvard	46.06	73.30	6.06	0.151	2.39	0.21	11.40
JAM2	Lanaudiere	Harvard	46.18	73.29	5.95	0.153	2.42	0.22	11.00
JAM3	Lanaudiere	Harvard	46.12	73.29		0.147	2.62	0.20	12.00
JAM4	Lanaudiere	Harvard	46.22	73.31	5.29	0.139	3.05	0.24	13.10
JAM5	Lanaudiere	Harvard	46.23	73.32		0.156	3.05	0.24	12.70
MAS1	Lac.St-Jean	Walton	48.50	72.28	5.64	0.204			
MASBEL1	Estrie	Warthog	45.63	71.55	5.46	0.172			
MASBEL2	Estrie	Warthog	45.62	71.55	5.99	0.244			
MASBEL3	Estrie	Warthog	45.62	71.55	5.99	0.214			
MASLAP1	Estrie	Warthog	45.23	71.42		0.186			
MJ18	Lac.St-Jean	Walton	48.25	71.47	6.64	0.165			
MJ2	Lac.St-Jean	Walton	48.51	72.27	5.81	0.173			
MJ20B	Lac.St-Jean	Walton	48.24	71.46	5.45	0.212			
MJ23	Lac.St-Jean	Walton	48.34	72.18	5.59	0.214			
NIC32	Centre.du.Qc	Orge	46.16	72.33	6.68	0.159			
NIC33	Centre.du.Qc	Walton	46.16	72.33	6.90	0.133	4.08	0.30	13.60
NJP1	Lanaudiere	Helios	46.53	73.19	5.85	0.168	1.18	0.11	10.70
NJP2	Lanaudiere	Helios	46.53	73.19	5.67	0.370	1.55	0.14	11.10
NJP3	Lanaudiere	Helios	46.53	73.19	6.05	0.159	1.46	0.13	11.20
NJP4	Lanaudiere	Helios	46.53	73.19	6.13	0.147	1.36	0.12	11.30
NOR2818	Mauricie	Scotia	46.52	72.25	5.44	0.148	4.06	0.28	14.50
NOR2823A	Mauricie	Scotia	46.51	72.23	5.79	0.113	2.33	0.17	13.70
NOR2829	Mauricie	Scotia	46.52	72.25	5.36	0.168	2.61	0.18	14.50
NUT10	Lac.St-Jean	Touran	48.29	71.37	5.97	0.171			
NUT16	Lac.St-Jean	Touran	48.28	71.37	5.96	0.205			
NUT21	Lac.St-Jean	Touran	48.25	71.48	6.52	0.186			
NUT22	Lac.St-Jean	Touran	48.25	71.48	6.99	0.174			
NUT23	Lac.St-Jean	Touran	48.25	71.48	6.49	0.146			
NUTAG18	Lac.St-Jean	Touran	48.29	71.48	5.84				
NUTCR10	Lac.St-Jean	Touran	48.35	72.20	6.38	0.164	3.04	0.23	13.20
NUTNN	Lac.St-Jean	Touran	48.24	71.48	6.23	0.182			
PRI2	Centre.du.Qc	Scotia	46.92	71.50	6.24	0.167	2.54	0.21	12.10
PRI3	Centre.du.Qc	Scotia	46.10	71.51	6.88	0.260	5.25	0.32	16.40
PRI4	Centre.du.Qc	Scotia	47.10	72.51		0.201	2.58	0.23	
PRI42	Centre.du.Qc	Scotia	46.20	72.17	6.94	0.117	3.22	0.26	11.20
PRO1	Estrie	Warthog	45.10	71.45	5.96	0.399	3.54	0.34	12.40
PRO2	Estrie	Warthog	45.10	71.46	5.97	0.217			10.40
QUI 2825B	Mauricie	Scotia	46.51	72.23	5.20	0.102	4.53	0.31	14.60
QUI2812	Mauricie	Scotia	46.50	72.24	5.67	0.126	4.09	0.28	14.60
QUI2807	Mauricie	Scotia	46.49	72.24	5.62	0.107			

QUI2809	Mauricie	Scotia	46.49	72.23	5.44	0.126	3.04	0.22	13.80
QUI2808	Mauricie	Scotia	46.50	72.23	5.11	0.130	3.92	0.29	13.50
QUI2823	Mauricie	Scotia	46.50	72.24	4.82	0.126			
QUI2833A	Mauricie	Scotia	46.49	72.23	5.38	0.109	5.89	0.43	13.70
QUI2833B	Mauricie	Scotia	46.49	72.23	5.61	0.152	4.70	0.37	12.70
RAY2E	Monteregie	Warthog	45.14	73.31	7.24	0.214			
T1B	Lac.St-Jean	Touran	48.35	72.20	5.96	0.397			
TI	Lac.St-Jean	Touran	48.35	72.20	6.48	0.191			
TI.SR92	Lac.St-Jean	Touran	48.19	68.60	4.96	0.409			
TI1	Lac.St-Jean	Touran	48.06	69.03	5.70	0.141			
TI22	Lac.St-Jean	Touran	47.59	69.02	5.75	0.355			

Samples highlighted in grey: samples from fields for which we received yield data and a grain sample.

Table S2: Number of raw read counts for each of the 80 soil samples. Number of raw read counts for each of the 80 soil samples following 16S rRNA gene and ITS region amplicon sequencing.

	16S rRNA gene		ITS region	
Sample	count	% total	count	% total
10	141,273	1.57%	88,470	1.08%
2	73,875	0.82%	130,942	1.59%
24A	27,958	0.31%	73,963	0.90%
24B	180,248	2.00%	140,677	1.71%
24D	171,708	1.91%	117,533	1.43%
3	164,404	1.83%	133,260	1.62%
4	124,151	1.38%	161,047	1.96%
5	116,660	1.30%	146,851	1.79%
6	101,196	1.13%	117,010	1.42%
7	124,068	1.38%	120,774	1.47%
8	46,012	0.51%	129,695	1.58%
9	123,114	1.37%	160,998	1.96%
BON225	112,857	1.26%	114,345	1.39%
BOU48	114,610	1.27%	40,636	0.49%
CEROM11	77,318	0.86%	90,937	1.11%
EBRER33	94,522	1.05%	33,685	0.41%
EBRER5	92,357	1.03%	114,284	1.39%
EBRER6	142,217	1.58%	70,102	0.85%
EBRER7	113,146	1.26%	78,638	0.96%
Fongiminus1	125,446	1.40%	83,978	1.02%

Fongiminus2	141,278	1.57%	67,717	0.82%
Fongiminus3	81,185	0.90%	90,760	1.10%
Fongiplus1	88,416	0.98%	41,833	0.51%
Fongiplus2	80,556	0.90%	84,463	1.03%
Fongiplus3	117,942	1.31%	84,313	1.03%
GDF1	111,759	1.24%	633	0.01%
GDF2	106,566	1.19%	99,204	1.21%
GDF3	146,648	1.63%	163,094	1.98%
IJ	90,228	1.00%	126,331	1.54%
JAM1	122,233	1.36%	99,891	1.21%
JAM2	85,031	0.95%	101,610	1.24%
JAM3	122,263	1.36%	150,829	1.83%
JAM4	163,830	1.82%	79,360	0.96%
JAM5	122,248	1.36%	1,007	0.01%
MAS1	87,301	0.97%	52,053	0.63%
MASBEL1	130,836	1.46%	102,891	1.25%
MASBEL2	125,176	1.39%	85,740	1.04%
MASBEL3	140,849	1.57%	97,324	1.18%
MASLAP1	86,089	0.96%	105,795	1.29%
MJ18	114,610	1.27%	92,307	1.12%
MJ2	90,726	1.01%	106,843	1.30%
MJ20B	128,806	1.43%	108,314	1.32%
MJ23	138,845	1.54%	190,772	2.32%
NIC32	84,666	0.94%	115,783	1.41%
NIC33	116,268	1.29%	165,221	2.01%
NJP1	152,815	1.70%	88,459	1.08%
NJP2	117,476	1.31%	76,790	0.93%
NJP3	149,234	1.66%	108,184	1.32%
NJP4	133,777	1.49%	86,052	1.05%
NOR2818	114,215	1.27%	143,849	1.75%
NOR2823A	136,670	1.52%	105,979	1.29%
NOR2829	117,650	1.31%	109,961	1.34%
NUT10	98,229	1.09%	71,645	0.87%
NUT16	117,410	1.31%	102,037	1.24%
NUT21	144,081	1.60%	120,217	1.46%
NUT22	97,011	1.08%	112,985	1.37%
NUT23	41,625	0.46%	123,661	1.50%
NUTAG18	116,666	1.30%	85,995	1.05%
NUTCR10	128,674	1.43%	100,100	1.22%
NUTNN	122,806	1.37%	52,633	0.64%
PRI2	67,243	0.75%	135,248	1.64%
PRI3	84,891	0.94%	99,164	1.21%
PRI4	101,840	1.13%	92,463	1.12%

PRI42	83,493	0.93%	123,270	1.50%
PRO1	117,938	1.31%	102,964	1.25%
PRO2	98,121	1.09%	127,850	1.55%
QUI.2825B	99,052	1.10%	112,093	1.36%
QUI2812	102,488	1.14%	114,570	1.39%
QUI2807	172,633	1.92%	78,893	0.96%
QUI2809	93,521	1.04%	151,444	1.84%
QUI2808	106,961	1.19%	2,920	0.04%
QUI2823	173,131	1.93%	98,498	1.20%
QUI2833A	108,806	1.21%	88,170	1.07%
QUI2833B	101,668	1.13%	97,974	1.19%
RAY2E	113,622	1.26%	105,233	1.28%
T1B	76,026	0.85%	238,404	2.90%
TI	89,821	1.00%	142,331	1.73%
TI.SR92	108,433	1.21%	105,003	1.28%
TI1	97,936	1.09%	61,124	0.74%
TI22	114,257	1.27%	96,736	1.18%

Tables S3: Amplification protocols for qPCR quantifications of N-cycle functional genes. Amplification protocols for qPCR quantifications of N-cycle functional genes, 16S rRNA gene and ITS region.

	amoA (bact)	amoA (arch), nirK, nosZ	16S rRNA gene, ITS region
Enzyme activation	95°C for 3 min	95°C for 3min	95°C for 5min
Number of cycles	40	40	30
Denaturation	95°C for 15s	95°C for 20s	95°C for 30s
Annealing	51.7°C for 45s	62°C for 30s	57 °C for 30s
Elongation	72°C for 60s	72°C for 20s	72 °C for 30s
Fluorescence acquisition	72°C, after elongation	72°C, after elongation	72°C, after elongation