Etude des interactions bactériennes dans la neige du Svalbard et de leur dynamique liée aux acides organiques

Benoit Bergk Pinto

THÈSE

Présentée devant :

ÉCOLE CENTRALE DE LYON Pour obtenir le grade de :

DOCTEUR

de l'école doctorale Électronique, électrotechnique, automatique

UNIVERSITÉ DE LYON

Spécialité : Ingénierie du vivant

par

Benoît Bergk Pinto

Étude des interactions bactériennes dans la neige du Svalbard et de leur dynamique liée aux acides organiques

Numéro de thèse NTT : 2020LYSEC16

Dr. Cene Gostinčar: chercheur – Department of Biology – Biotechnical Faculty – Ljubljana University (Slovénie)	**Rapporteur**
Dr. Beat Frey: directeur de laboratoire – Institut fédéral suisse de recherche pour la forêt, la neige et les paysages WSL (Suisse)	**Rapporteur**
Dr. Purification Lopez-Garcia: directrice de recherche CNRS – ESE (Ecologie Systémique Evolution) – Université Paris Sud (France)	**Membre externe du jury**
Dr. Graeme Nicol: directeur de recherche à Groupe de Génomique Microbienne Environnementale - Ecole Centrale de Lyon - Lyon (France)	**Président du jury**
Dr. Catherine Larose: chargée de Recherche CNRS à Groupe de Génomique Microbienne Environnementale - Ecole Centrale de Lyon - Lyon (France)	**Directrice de thèse**
Pr. Timothy M. Vogel : professeur à Groupe de Génomique Microbienne Environnementale - Ecole Centrale de Lyon - Lyon (France)	**Co-encadrant**

THESIS

Presented at:

ÉCOLE CENTRALE DE LYON

For the title of:

Doctor

of doctoral school: Électronique, électrotechnique, automatique

UNIVERSITÉ DE LYON

Speciality: Bioengineering

By

Benoît Bergk Pinto

Organic acid driven bacterial interactions in Arctic snow

Dr. Cene Gostinčar: researcher – Department of Biology – Biotechnical Faculty – Ljubljana University (Slovenia)	**Reviewer**
Dr. Beat Frey: groupleader and senior scientist – Swiss federal institute for forest, snow and landscape WSL (Switzerland)	**Reviewer**
Dr. Purification Lopez-Garcia : CNRS research director – ESE (Ecologie Systémique Evolution) – Université Paris Sud (France)	**External member of the jury**
Dr. Graeme Nicol : research director at Environmental Microbial Genomics Group - Ecole Centrale de Lyon - Lyon (France)	**President of the jury**
Dr. Catherine Larose: in charge of research (CNRS) at Environmental Microbial Genomics Group - Ecole Centrale de Lyon - Lyon (France)	**Thesis advisor**
Pr. Timothy M. Vogel : professor at Environmental Microbial Genomics Group - Ecole Centrale de Lyon - Lyon (France)	**Co-supervisor**

Table of Contents

Remerciements

Et voilà, c'est bientôt la fin d'une aventure débutée il y a un peu plus de trois ans et appelée doctorat. Et quelle aventure ! Avec elle s'achève mon séjour en France et dans un laboratoire où j'ai vécu de très beaux moments. Et c'est aux personnes qui ont contribué à faire de cette aventure un si beau souvenir que je tiens à dire merci au cours de ces quelques lignes.

Tout d'abord, un tout grand merci à mes directeurs de thèse, Catherine Larose et Timothy M. Vogel. Merci de m'avoir fait confiance pour mener à bien cette thèse. Merci aussi pour votre soutient, votre patience et toutes les discussions et critiques constructives que vous avez pu me faire durant ces années de thèse. Non seulement ma thèse s'en est retrouvée améliorée mais ceci m'a également permis d'évoluer dans mon rapport au travail scientifique et à, tantôt la joie, tantôt la frustration que peuvent procurer certains résultats inattendus.

Ensuite, je tiens à remercier mes compagnons engagés dans la même aventure, … , et dans le même bureau! Merci donc à Mia Sungeung, Romie Perrier, Adrien Boniface, Rose Layton et Conception Sanchez-Cid Torres pour tous ces bons moments partagés durant nos pauses déjeuner, café et les diverses sorties que nous avons pu effectuer! Je tiens aussi à remercier Mia et Romie pour l'aide qu'elles ont pu m'apporter dans le laboratoire tant pour m'y retrouver avec les produits que pour leur conseils d'organisation qui m'ont permi de mener à bien toutes mes expériences. Je remercie aussi Rose et Conception pour l'aide qu'elles m'ont fournie pour certaines expériences (comme la préparation de plaque de séquençage ou l'utilisation de l'unité de filtration !).

I will now switch in English for a bit to thank warmly the English speakers from my lab! I want first to thank warmly Christoph Keuschning for his great help concerning the sequencing and also the troubleshooting of my qPCR issues (BSA saved my life !). I also want to thank again Rose Layton for her great help during some part of my experiments but also during the trip in Svalbard. Thanks to some of her social skills, I was able to go by boat very close to the icebergs and this is a very great memory of my trip there ! Thanks also to Graeme Nicol and Christina Hazard for all the good moments spent and the karaoke evenings organized during conclave. I would also like to thank Cecile Thion, Linda Hink and Damien Finn as excellent people who also contributed to my spending a great time in this lab even if they didn't stay until the end. I would also like to thank Frederik Bak, a visiting PhD student, for all the nice scientific talk (I switched from OTU clustering to ASV thanks to you !) and for all the fun times that I spent running with him and improving my sport skills.

Je tiens également à remercier d'autres membres du laboratoire. Merci à Laure Franqueville et Cécille Thion pour leur bonne gestion du laboratoire et sans qui les commandes auraient été un parcours du combattant. Je remercie également Edith Bergeroux pour sa gentillesse et son aide pour toute la paperasse des ordres de missions et des inscriptions ! Merci aussi à Marie-Christine pour son assistance déterminante dans toutes les procédures administratives liée à la thèse. Merci aussi à Richard pour sa capacité à réparer tout ce qui peut l'être et sans qui j'aurais eu quelques soucis pour partir au Svalbard avec une unité de filtration fonctionnelle. Merci aussi à Alice, notre petit rayon de soleil réunionnais et tornade blanche du labo pour sa gentillesse et ses agréables conversations qui ramènent le sourire même

lorsque la journée a mal commencé. Merci à Pascal pour ses discussions et réflexions intéressantes sur les antibiotiques et… ses barbecues d'anthologie! Enfin, merci aussi à tous les autres membres du laboratoire qui ne sont pas dans ce texte, misi pas oublié, ils sont nombreux et ont également contribués à faire de ce lieu un agréable environnement de travail.

I would also like to thank warmly all the members (PI and PhD students) from the Microarctic ITN for the fantastic time that I had working with them. Specifically, I would like to thank Antonio for creating a great atmosphere each time we met! Thank you also Laura, Antonio, Zohaib and Alex for all the great time we had during the meeting but also outside as friends. I would also like to thank warmly Zohaib for the nice time that I spent in Roskilde during the secondment and his highly appreciated help to improve the pipeline presented in this thesis. I want also to thank Alex Anesio and Carsten Suhr Jacobsen for their kindness during the secondment in Roskilde and the stimulating discussions that we had concerning this work. Last, but not least, in the project, I want to thank Clare and Alex for organizing all those meetings and making this project run so well.

I'd like also to thank all the members of the jury to have accepted to take part to this PhD defense.

Je tiens aussi à remercier les personnes du laboratoire IGE de Genoble et notamment Aurélien Dommergue, pour les mesures chimiques réalisées sur les échantillons de neige de cette thèse et qui ont donc rendu possible le présent travail.

Je termine ce long remerciement en remerciant chaleureusement mes amis de longue date Tatyana, Jérome, Cindy, Maxence et Anne-Elisabeth pour leur gentillesse et cette amitié qui n'a jamais cessé malgré le fait que je n'était plus beaucoup en Belgique durant tout ce temps!

Enfin je remercie mes parents et ma soeur Nicole pour m'avoir si bien entouré durant toute ma vie mais aussi et surtout durant la fin de cette thèse qui a été bien plus éprouvante que je ne l'aurais imaginé!

Un grand merci à vous tous!

Résumé (français)

Les interactions positives entre microorganismes ont longtemps été minoritairement étudiées par les microbiologistes. Mais aujourd'hui, la coopération entre bactéries gagne en importance depuis qu'il a été mis en évidence qu'un grand nombre de bactéries sont auxotrophes et nécessitent la présence d'autres microorganismes pour se développer. Ce nouveau domaine d'étude s'est principalement développé grâce à des co-cultures de microorganismes en laboratoire.

Dans cette thèse, nous désirions valider certaines prédictions issues de ces études en laboratoire en les testant sur une communauté bactérienne d'un environnement naturel. Nous nous sommes focalisés sur l'impact des acides organiques et avons émis l'hypothèse qu'une augmentation de la concentration en acide organique causerait une augmentation de la compétition entre bactéries tout en diminuant la coopération dans cette communauté bactérienne. Afin de tester cette hypothèse, nous avons d'abord utilisé le séquençage métagénomique afin de détecter les gènes rapportés dans la littérature scientifique comme étant un indice de coopération (plasmides) ou de compétition (gènes de résistance aux antibiotiques) entre bactéries. De plus, nous avons également évalué les interactions bactériennes en construisant des réseaux de co-variance basés sur des données issues du séquençage 16S rRNA. Cette approche hybride fut ensuite appliquée sur une communauté bactérienne de neige arctique.

Au cours de notre première étude, réalisée sur une série temporelle de neige, nous avons pu appliquer avec succès notre méthode afin d'étudier l'impact des acides organiques sur les interactions bactériennes dans la neige. Nous avons mis en évidence que les échantillons présentant une concentration plus importante en acides organiques présentaient également un nombre plus important de gènes de résistance aux antibiotiques. Ce résultat supporte l'hypothèse qu'une augmentation de la concentration d'acides organiques dans la neige augmente la compétition entre bactéries. A l'inverse, les échantillons de neige possédant de fortes concentrations d'acides organiques présentaient un plus faible nombre de gènes structuraux de plasmides dans leurs métagénomes. Ceci étaye ainsi l'hypothèse qu'une augmentation de la concentration d'acides organiques dans la neige diminue également la coopération. La comparaison des réseaux de co-variance a conforté cette interprétation. Suite à ces résultats encourageants, nous avons décidé de valider de manière plus minutieuse notre hypothèse en tentant de reproduire nos résultats dans des microcosmes de neiges amendés au moyen d'un des acides organiques majoritairement identifié dans notre précédente étude (acétate). En parallèle, nous avons également développé un pipeline personnalisé pour traiter nos métagénomes en améliorant la fiabilité de l'annotation fonctionnelle des séquences qui ne peuvent pas être assemblées (en déduisant une valeur seuil d'annotation suivant sa distribution observée dans les séquences assemblées). En utilisant cette nouvelle méthode d'annotation pour traiter nos métagénomes, nous avons reproduit la même approche hybride pour étudier les interactions bactériennes. Nous avons pu confirmer le fait que l'augmentation de la concentration d'acides organiques augmentait la compétition entre bactéries mais nous n'avons pas observé d'impact significatif sur la coopération qui ne différait pas beaucoup du niveau observé dans les microcosmes contrôles. Nous en avons déduit que la concentration d'acides organique dans la neige affectait principalement la compétition entre bactéries mais n'avait pas ou peu d'effet sur la

collaboration dans la communauté bactérienne de la neige arctique. Malgré ces résultats encourageants, le présent travail a également mis en lumière la difficulté de pouvoir interpréter de manière univoque les gènes qui participent aux interactions bactériennes et suggère la mise en place d'une base de données spécialement dédiée à ce type de gènes.

Summary

Microbial interactions are ubiquitous in the environment, but microbiologists mainly focused on negative interactions (mainly competition) between microorganisms as natural selection would only select for the most individually adapted bacteria in the environment. However, bacterial cooperation is starting to attract more and more attention as microbiologists realized that a significant number of microorganisms are auxotrophic for one or more biomolecules and require the presence of other microorganisms in order to grow. This new field of microbiology has been mostly developing in laboratory controlled co-cultures. Thus we now face the challenge of validating the acquired knowledge from the wet lab experiments in the environment.

In this thesis, we wanted to validate observations that had been made in lab controlled co-cultures at the level of an environmental bacterial community. We focused on the effect of organic acids (a carbon source) and hypothesized that an increase in their concentration would augment bacterial competition and reduce bacterial cooperation. To test this hypothesis, we selected two methods to assess bacterial interactions. First, we tracked genes reported as being proxies of cooperation (plasmids) and competition (antibiotics resistance genes-ARG) in metagenomes. We also used co-variance networks to assess microbial interactions. This hybrid approach was then used on a bacterial community from the Arctic snow. The Arctic snow environment was used as a model because it is reported as being dynamic with seasonal increases in organic acids.

During our first study, carried out on snow samples collected in Svalbard, we successfully applied our methodology to track bacterial interactions and how they were influenced by the increase in organic acid concentrations. In the snow metagenomes, the ARGs were detected in higher abundance in the snow samples with higher organic acid concentrations, which we considered as a signal of an increase of competition. In contrast, plasmid backbone genes were retrieved in higher abundance in the snow metagenomes from samples with low organic carbon concentrations as cooperation should decrease when organic acids increase. The co-variance networks showed a decrease of connectivity in the networks in high organic acid concentration snow samples. To validate our observations, we carried out a time series experiment in lab-controlled snow microcosms amended with acetate, which is one of the most abundant organic acids found in the field. In parallel, we developed a custom bioinformatics pipeline to process the functional annotation of our metagenomes in a more accurate way for of the reads that could not be assembled. This pipeline improved the degree of accuracy of read annotation (with an error rate of 5%) and improved our annotation using an e-value threshold on the assembled reads. We applied this new annotation method on our snow microcosm metagenomes and repeated our methodology to track microbial interactions. We confirmed that organic acids triggered bacterial competition in our microcosms, but showed little effect on reducing bacterial collaboration. This work highlighted the difficulties of using genes reported as proxies of cooperation or competition and that more effort is still needed to build an appropriate reference database for such genes.

List of publications

Published articles:

Bergk Pinto, Benoît, Lorrie Maccario, Aurélien Dommergue, Timothy M. Vogel, and Catherine Larose. 2019. **"Do Organic Substrates Drive Microbial Community Interactions in Arctic Snow?"** *Frontiers in Microbiology* 10. https://doi.org/10.3389/fmicb.2019.02492.

Alexandra T. Holland, **Benoît Bergk Pinto**, Rose Layton, Christopher J. Williamson, Alexandre M. Anesio, Timothy M. Vogel, Catherine Larose, Martyn Tranter **Over winter microbial processes in a Svalbard snow pack: an experimental approach**, *Frontiers in Microbiology 11.* https://doi.org/10.3389/fmicb.2020.01029

Zhu, Chengsheng, Maximilian Miller, Nicholas Lusskin, **Benoît Bergk Pinto**, Lorrie Maccario, Max Haggblom, Timothy Vogel, Catherine Larose, and Yana Bromberg. 2020. **"Snow microbiome functional analyses reveal novel aspects of microbial metabolism of complex organic compounds."** MicrobiologyOpen, In press.

Submitted articles:

D.R. Finn, **B. Bergk Pinto**, J. Cole, H. Cadillo-Quiroz, T.M. Vogel **Functional trait-based modeling supports the copiotroph-oligotroph hypothesis at the species and community scale of terrestrial microorganisms**, FEMS, submitted

Avant-propos

L'introduction suivante est destinée à donner le contexte général nécessaire pour comprendre le projet de doctorat qui sera décrit. Il commence par introduire les interactions bactériennes qui est le principal sujet d'étude de cette thèse. Ensuite, l'environnement choisi pour l'étude est décrit ainsi que le projet dans lequel ce doctorat est intégré. Les hypothèses de travail seront introduites, à partir des différents éléments présentés en introduction. Un résumé des autres chapitres de la thèse incluant les résultats les plus marquants est également inclus.

Synthèse (Français)

Les bactéries, comme tous les autres organismes vivants, interagissent avec leur environnement mais aussi avec d'autres bactéries et micro-organismes. Ces interactions peuvent être classées en différentes catégories selon qu'elles sont neutres, nuisibles ou bénéfiques pour l'un ou les deux partenaires (Faust and Raes 2012). Le résultat de l'interaction est déterminé par l'effet de l'interaction sur la valeur sélective (le fitness en anglais) des bactéries impliquées. La valeur sélective d'un organisme peut être définie comme sa capacité à survivre dans un environnement particulier et à se reproduire pour participer au pool génétique de l'espèce dans la prochaine génération. Étant donné que les bactéries se développent par clonalité, nous pourrions définir la valeur sélective des bactéries comme le nombre de cellules filles qu'elles seront capables de produire. Si la bactérie parvient à augmenter sa progéniture grâce à l'interaction, elle sera qualifiée de positive. Au contraire, si l'interaction diminue ce montant, elle sera qualifiée de négative. Le résultat de l'interaction peut être bénéfique (augmentation de la valeur sélective = positif), préjudiciable (diminution de la valeur sélective = négatif) ou neutre (pas d'impact sur le fitness = nul) pour les espèces en interaction. Par exemple. Si nous considérons la prédation, le prédateur augmente sa valeur sélective dans l'interaction parce qu'il mange sa proie mais la proie diminue sa valeur sélective car elle est consommée par le prédateur. Au niveau des espèces, l'interaction est positive pour les espèces prédatrices et négative pour les espèces proies.

En pensant aux concepts écologiques ayant considéré les interactions biologiques, nous pourrions considérer la théorie de l'évolution de Charles Darwin (Darwin 1859) où les espèces non adaptées à leur environnement finiraient par disparaître. En effet, par environnement, il faut aussi inclure les interactions avec les autres espèces qui peuvent augmenter ou diminuer la valeur sélective des espèces impliquées dans l'interaction. Cette importance des interactions dans le processus évolutif a été formalisée pour la première fois par Van Valen (1973) dans son hypothèse de la Reine Rouge. Il a déclaré que pour survivre, les espèces devaient s'adapter en permanence à leur environnement mais aussi aux espèces avec lesquelles elles étaient en compétition. Le fait que cette « course évolutive » ne se terminerait jamais pourrait s'expliquer par le fait que chaque espèce tente d'atteindre un optimum mutuellement incompatible avec celui de ses concurrents. En outre, cette théorie soutenait que le principal type d'interactions dans la nature était principalement la compétition ou le parasitisme : `` La Reine Rouge propose que les événements de mutualisme, au moins au

même niveau trophique, aient peu d'importance dans l'évolution par rapport aux interactions négatives (…) »(Van Valen 1973). C'est pourquoi en microbiologie, jusqu'à récemment, la grande majorité des travaux sur les communautés microbiennes ne recherchaient pas trop les interactions, car on pensait que la composante principale était la compétition comme le déclarent Foster and Bell (2012) par exemple. En conséquence, les changements dans la structure de la communauté bactérienne ne pourraient résulter que d'une diminution de la valeur sélective due à un taux de croissance plus faible dans les nouvelles conditions environnementales (changement de pH par exemple) ou à des interactions compétitives (par exemple, compétition pour l'espace ou accès préférentiels aux nutriments). C'est pourquoi de nombreux articles ont considéré principalement l'environnement chimique et ont oublié de s'intéresser aux interactions bactériennes…

De nos jours, l'importance des interactions bactériennes positives commence à être considérée. En effet, de plus en plus de publications se concentrent sur les interactions bactériennes et, fait intéressant, les publications liées aux interactions positives croissent plus rapidement que les interactions négatives. Ces tendances pourraient être liées à la publication d'une nouvelle hypothèse évolutive soutenant l'apparition d'une collaboration à partir d'une interaction compétitive. Cette hypothèse, appelée `` hypothèse de la reine noire '', soutient la théorie selon laquelle la perte de gènes de fonctions coûteuses fuyantes (fonctions métaboliques qui libèrent leurs produits finaux par diffusion dans l'environnement, produisant ainsi des biens publics) peut entraîner un accroissement de la valeur sélective au niveau individuel si le nombre de producteurs dans la communauté (également appelés «aides») est encore suffisamment important pour soutenir la croissance de la communauté. Ces bactéries qui perdent de tels gènes deviennent des «bénéficiaires» des «aides» (Morris 2015). On a également émis l'hypothèse que ce couplage métabolique permettait aux bactéries d'éviter une compétition exclusive en faisant en sorte que les deux partenaires de ce couplage atteignent un état stable dans l'environnement, évitant ainsi la disparition de l'une des deux espèces et réduisant la compétition (Mas et al. 2016).

Une deuxième conséquence de cette théorie est que le faible taux de micro-organismes pouvant croître en culture, connue sous le nom d '«anomalie du grand nombre de plaques» pourrait également s'expliquer par le fait qu'un nombre très élevé de micro-organismes partagent des métabolites de bien public, étant ainsi auxotrophe pour un ou plusieurs métabolites spécifiques (Zengler and Zaramela 2018). Cette hypothèse est soutenue à un rythme accru par des expériences de culture réussissant à isoler des micro-organismes non cultivés auparavant en ajoutant par exemple des sidérophores (D'Onofrio et al. 2010). De plus, des expériences récentes de co-culture de Ren et al. (2015) ont montré un nombre sans précédent de synergies positives de souches bactériennes isolées des sols, soutenant le fait que les couplages métaboliques pourraient également être importants dans les communautés bactériennes environnementales. D'Souza and Kost (2016), ont également montré dans une expérience de culture en laboratoire que la culture de bactéries dans un milieu de culture où un nutriment qu'elles étaient capables de produire était déjà présent pouvait sélectionner la perte des gènes de synthèse de cette voie, confirmant l'hypothèse de Morris (2015).
Les interactions bactériennes peuvent être très diverses et impliquer également des interactions avec une grande variété d'autres taxons en interaction comme les plantes (Finkel

et al. 2019), les champignons (van Overbeek and Saikkonen 2016) ou les animaux (Rønn, Vestergård, and Ekelund 2015).
L'importance de ces interactions est de plus en plus mise en évidence par des études récentes. Par exemple, il a été démontré que le développement racinaire des plantes pouvait être affecté par des interactions bactériennes spécifiques (Finkel et al. 2019) ou que le rendement du maïs pouvait être corrélé à des interactions bactériennes spécifiques (Tao et al. 2018). Il a également été démontré que les interactions bactériennes modulent le succès de reproduction des diatomées qui constituent une partie très importante du phytoplancton, les principaux producteurs des océans (Torres-Monroy and Ullrich 2018). Enfin, les interactions bactériennes peuvent également moduler les cycles biogéochimiques (Ho et al. 2016). Si les deux espèces bactériennes impliquées dans l'interaction bactérienne modulent des réactions de différents cycles biogéochimiques, un couplage biogéochimique peut être observé dans l'environnement conduisant à des dépendances métaboliques imprévues entre les cycles (Schlesinger et al. 2011; Burgin et al. 2011). Par exemple, Beal, House, and Orphan (2009) ont découvert que l'oxydation du méthane pouvait être couplée au manganèse ou même au fer au lieu de la réduction du sulfate dans les sédiments marins à suintement de méthane. Un autre exemple impressionnant de couplage de processus biogéochimiques a été trouvé dans une ancienne saumure marine isolée dans l'Antarctique. Mikucki et al. (2009) ont observé qu'une communauté microbienne active faisait circuler le soufre par couplage avec une réduction de fer en Fe (II). Un tel couplage a permis à la communauté microbienne de se développer malgré son isolement (pas de photosynthèse entraînant un apport limité de carbone dans la communauté) (Mikucki et al. 2009).

Les interactions bactériennes attirant de plus en plus l'attention, la stabilité d'une telle interaction a également commencé à être étudiée. Une grande variété de facteurs environnementaux a été étudiés principalement au cours d'expériences en laboratoire. Il a été démontré que chacun de ces paramètres sélectionne les interactions positives (coopération), ou au contraire, les interactions compétitives. Ainsi, en pensant à l'environnement où les concentrations de nutriments, le pH, l'humidité et l'homogénéité de l'environnement peuvent varier dans le temps, nous pourrions commencer à penser que les interactions bactériennes peuvent être considérées comme des liens dynamiques entre les bactéries. Un tel concept peut être illustré par l'expérience de co-culture de Benomar et al. (2015) qui ont observé que lorsque *Desulfovibrio vulgaris*, une bactérie réductrice du sulfate, était co-cultivée avec *Clostridium acetobutylicum*, certains échanges métaboliques entre les espèces pouvaient démarrer lorsqu'une pénurie de nutriments (manque de sulfate) se produisait dans le milieu de culture. La conséquence d'une telle dynamique est que le résultat de la structure d'une communauté bactérienne définie ne peut pas être prédit uniquement en examinant leurs métabolismes individuels mais aussi en examinant tous les couplages et potentiels d'interaction possibles qui se trouvent entre ses différents membres. Konopka, Lindemann, and Fredrickson (2015) ont modélisé une telle dynamique et ont observé que les interactions bactériennes pouvaient générer une dynamique endogène affectant la communauté bactérienne même en l'absence de perturbations exogènes. Ils ont également remarqué que le réseau d'interactions pouvait améliorer la résistance de la communauté bactérienne contre les perturbations et devrait être pris en compte en plus de la redondance métabolique dans la communauté bactérienne (Konopka, Lindemann, and Fredrickson 2015). L'étude de l'effet de chacune des variables environnementales sur une communauté bactérienne est toujours en cours, de nouveaux effets sont donc encore découverts et,

comme nous pouvons le constater en regardant la littérature scientifique, toutes les études ne s'accordent pas sur l'effet de chaque paramètre. Ces résultats contrastés peuvent s'expliquer par le fait que la coopération et la compétition peuvent également se produire à différentes échelles au sein d'une communauté bactérienne. Par exemple, Cordero et al. (2012) ont observé que les antibiotiques étaient sécrétés par certains membres d'un consortium bactérien tandis que les autres souches collaboratives partageaient les gènes de résistance aux antibiotiques. Ils ont conclu que la synthèse d'antibiotiques pouvait être considérée comme un trait collaboratif dans cette partie de la communauté et était utilisée pour rivaliser avec d'autres membres de la communauté. D'un autre côté, les bactéries modèles utilisées pour ces études peuvent également influencer les résultats et montrer ainsi que différentes stratégies de collaboration ou de compétition pourraient être sélectionnées dans différents environnements.

Dans cette thèse, nous avons choisi de considérer principalement les interactions bactériennes à travers les interactions liées à la nutrition et non liées à la compétition pour l'espace donc nous nous concentrerons principalement sur les interactions suivantes pour nos interprétations :

(A) Compétition nutritionnelle : Il s'agit d'un cas spécial de compétition (négatif-négatif) où les deux partenaires sont en compétition un nutriment particulier. Même si finalement l'une des deux espèces peut remporter la compétition, l'interaction est négative pour les deux espèces car chaque espèce prend une partie du pool de nutriments qui est ainsi perdu pour la seconde espèce.
(B) Syntrophie: Il s'agit d'un cas particulier de mutualisme (Positif-Positif). Deux bactéries coopèrent pour dégrader un composant qu'elles ne pourraient pas dégrader seules. Ainsi, l'interaction est positive pour les deux bactéries, car elles peuvent métaboliser un pool de nutriments qu'elles ne pourraient pas traiter sans la présence de leur partenaire.
(C) Alimentation croisée (appelée cross-feeding en anglais) : il s'agit d'un cas particulier de comensalisme (neutre-positif). La présence de la première espèce produisant un nutriment essentiel, tel qu'une vitamine, permet à une bactérie incapable de produire ce composé (l'auxotrophe) de survivre. La bactérie produisant la vitamine ne tire aucun avantage de cette interaction, mais l'auxotrophe est capable de survivre dans cet environnement grâce à cette deuxième espèce, de sorte que sa valeur sélective est considérablement augmentée par la présence de cette autre espèce.

L'Arctique

L’Arctique et le réchauffement global

L’Arctique est généralement défini comme la partie du globe localisée au-delà de 66◦33 ’de latitude dans l’hémisphère Nord. Les températures les plus chaudes de l'année sont généralement inférieures à 10 degrés Celsius. Le temps est donc très froid. Cela peut s'expliquer par le fait que les pôles obtiennent un taux d’ensoleillement atténué en raison de l’inclinaison de la Terre par rapport à l'incidence des rayons solaires. Les jours et la nuit varient

considérablement au cours de l'année : les deux extrêmes représentant 24 heures de nuit (nuit polaire) en hiver et 24 heures de jour en été.

Pendant longtemps, ces régions ont été considérées comme très pauvres en termes de biodiversité et de chaîne alimentaire mais, de nos jours, cette vision a considérablement évolué au fur et à mesure que la connaissance de cet environnement augmentait. Malgré les conditions extrêmes, une riche biodiversité et des chaînes alimentaires complexes sont présentes.
Néanmoins, cette biodiversité est désormais en danger du fait du réchauffement climatique et les chaînes alimentaires commencent à être impactées par l'effet de cette augmentation rapide de la température (Post et al. 2009). Les effets du changement climatique peuvent être multiples. Ce changement de gradient de température peut par exemple permettre aux espèces envahissantes de se propager à travers un nouvel écosystème et d'avoir un impact significatif sur les espèces indigènes. Il peut également modifier la disponibilité des nutriments (= la nourriture), les échanges de gaz et les bilans carbone des écosystèmes (Post et al. 2009). Mais tous les effets des changements climatiques ne sont pas aussi évidents que les précédents, car ils peuvent être masqués par l'effet tampon des écosystèmes ou des interactions entre les espèces (Post et al. 2009). La dynamique de l'Arctique peut également avoir un impact indirect sur d'autres écosystèmes et avoir une rétroaction positive sur le changement climatique lui-même, car la fonte de la couverture de neige et de glace de mer diminue l'albédo mondial de la Terre. C'est pour cette raison qu'il est urgent de caractériser davantage ces écosystèmes afin de pouvoir évaluer plus précisément les impacts réels à long terme du changement climatique sur les écosystèmes arctiques.

Présentation du projet Microarctic

Ce doctorat fait partie du réseau innovant de formation (ITN) Microarctic soutenu par une subvention du programme d'actions Marie Sklowdowska Curie de la Commission européenne. Les objectifs de ce projet sont de former la prochaine génération d'experts en microbiologie et biogéochimie de l'Arctique qui, grâce à leur compréhension unique de l'environnement arctique en évolution rapide et des facteurs qui influent sur la réponse des écosystèmes et des organismes au réchauffement de l'Arctique, seront en mesure de répondre aux besoins de gouvernance et de leadership dans divers aspects liés aux intérêts publics, politiques et commerciaux.
Le réseau Microarctic est composé de 15 doctorats réalisés dans 13 universités et entreprises à travers l'Europe. Ce grand réseau étudiera les différents écosystèmes composant l'Arctique (air, pergélisol, glace, ...) de différentes manières. Le projet en lui-même est divisé en 7 lots de travaux interconnectés (WP). Ce doctorat fait partie du WP 1 qui est un WP axé sur l'étude de l'effet du temps, de la saison et du réchauffement de l'Arctique sur les communautés bactériennes des écosystèmes terrestres de l'Arctique. Au cours de cette thèse, j'étudierai la dynamique saisonnière des interactions bactériennes dans la neige de l'Arctique. Le choix d'étudier les interactions bactériennes dans un tel environnement a été motivé par plusieurs critères que je vais essayer de mettre en évidence en présentant l'environnement de la neige.

La neige est une composante majeure de la cryosphère terrestre (régions polaires et alpines du globe). Il couvre environ 46 millions de kilomètres carrés pendant l'hiver. Plus de 98% de la neige est localisée dans l'hémisphère Nord. Parmi tous les différents biomes de l'Arctique, la neige peut être considérée comme très importante. Elle est colonisée par une communauté diversifiée d'algues des neiges, de bactéries et de champignons. La neige a un impact significatif sur les sols de l'Arctique de plusieurs façons. Pendant l'hiver, elle agit comme isolant sur le sol en le protégeant des vents froids (Vincent et al. 2009). De plus, au printemps, l'enneigement commence à fondre et les nutriments de la communauté microbienne de la neige sont transférés dans le sol (Vincent et al. 2011). Un changement de la couverture neigeuse peut également avoir un impact significatif sur l'hydrologie de la région arctique et, par conséquent, sur les écosystèmes aquatiques tels que les lacs, les rivières et les zones humides, car c'est l'une des principales sources d'eau pour certains de ces écosystèmes (Vincent et al. 2009). Ainsi, la vulnérabilité de cet écosystème est une préoccupation majeure et la caractérisation des communautés microbiennes de la neige est vraiment cruciale car elles peuvent interagir et avoir un impact également sur d'autres communautés.

La neige arctique pourrait être considérée comme un environnement extrême. En effet, la température est très basse et la disponibilité en eau est faible. De plus, pendant la saison printanière, le rayonnement UV peut être très élevé à sa surface (Maccario et al. 2015). Pour survivre dans un tel environnement, les bactéries ont développé une gamme de stratégies et d'adaptations. Pour survivre aux stress photo-oxydants induits par une forte irradiation UV, les bactéries peuvent par exemple produire des enzymes anti-oxydantes capables de réagir de manière croisée avec les ROS (= Reactive Oxygen Species) générées par les UV et réparer leur ADN endommagé (Sinha and Häder 2002; Ziegelhoffer and Donohue 2009). Les bactéries sont également exposées à une concentration élevée en sel car la majeure partie de la communauté bactérienne pourrait être concentrée dans des micro-canaux riches en sel à l'intérieur de la neige (Maccario et al. 2015). Nous désignons les bactéries adaptées pour vivre dans des environnements secs ou très salés comme étant psychrophiles. Néanmoins, au printemps, une augmentation rapide des nutriments peut être observée à l'intérieur de la neige et l'environnement peut devenir assez riche en nutriments par rapport à la neige d'hiver. Cet environnement est donc assez dynamique et présente une large gamme de variations environnementales. Cette propriété est donc très intéressante puisque la neige peut être utilisée comme environnement modèle pour comparer les interactions bactériennes dans un environnement oligotrophe (pauvre en nutriments) et le même environnement enrichi en nutriments à la fin de la saison printanière. Étant donné le fait que lors de cette thèse, nous nous sommes focalisés sur les interactions liées à la nutrition des bactéries, ce milieu fut choisi principalement pour cette propriété remarquable.

Après cette brève revue du milieu d'étude nous allons maintenant passer en revue les outils d'étude utilisés pour les interactions bactériennes.
Comme nous l'avons vu, la grande majorité des études menées sur les interactions bactériennes ont été réalisées dans des expériences basées sur la culture. De telles méthodes présentent un biais majeur pour étudier les interactions bactériennes à l'échelle de la communauté car leur représentativité du système d'origine est loin d'être exhaustive. Les principaux avantages de ces systèmes sont qu'ils sont très faciles à suivre dans le temps et

présentent une complexité réduite permettant d'utiliser la protéomique pour suivre les métabolites sécrétés par certaines souches spécifiques (Chignell et al. 2018; Herschend et al. 2017) ou visualiser physiquement les interactions par microscopie ou via des techniques plus complexes telles que le nanoSIM (Musat et al. 2016).

Les méthodes de culture incluent des co-cultures où l'observation de taux de croissance différentiels par rapport à la culture pure peuvent être utilisés pour déduire si l'interaction est positive ou négative. De telles méthodes peuvent également être utilisées sur des co-cultures d'organismes modèles où leurs génomes respectifs sont déjà connus afin d'évaluer comment une interaction positive ou négative peut affecter leurs profils d'expression génique en générant des profils transcriptomiques différentiels (Hansen et al. 2017; Molina-Santiago et al. 2017; Khan et al. 2018; McClure et al. 2018). Enfin, la modélisation peut également être appliquée afin de prédire, sur la base d'expériences de cultures antérieures et d'analyses de génomes, leurs réseaux métaboliques et comment une interaction ou une perturbation du système (par exemple une augmentation des nutriments) pourrait les affecter (e.g Zeng and Yang 2019) . Ces analyses sont appelées Flux Balance Analysis (FBA) et peuvent prédire comment les flux métaboliques seraient affectés par une perturbation. Cette analyse repose sur le fait que chaque réaction métabolique est connue et peut être estimée par des systèmes d'équations qui peuvent être résolus. Sur la base d'une telle approche, Zelezniak et al. (2015) ont développé un outil pour estimer quels métabolites pourraient être échangés et prédire, sur la base de tels systèmes d'équations, si les interactions entre les espèces considérées pouvaient être positives ou négatives. Néanmoins, actuellement, ce système est limité par le nombre d'espèces pouvant être calculées par le programme (<100), ce qui le rend actuellement inadapté aux études de communautés environnementales (com. Pers. De l'auteur de l'outil).

L'utilisation de méthodes sans culture est récente mais a tendance à se développer très rapidement car elle permet de s'affranchir du biais de représentativité des méthodes précédentes. Les deux principales méthodes sans culture utilisées de nos jours pour étudier les interactions bactériennes comprennent les réseaux de co-variance / co-occurrence et les métagénomes ou analyses de métatranscriptomes. Les deux techniques ont actuellement des limites.
L'approche des réseaux repose principalement sur l'hypothèse que les taxons qui covarient positivement dans le temps coopèrent et ceux qui covarient négativement sont en compétition. Cette approche a été utilisée pour les communautés microbiennes des océans (Ruan 2006; Lima-Mendez et al. 2015) les sols (Barberán et al. 2012; Ding et al. 2015), les microbiomes humains (Faust et al. 2012) et des sédiments pollués par les métaux lourds(Yin et al. 2015). Ces réseaux utilisent souvent la co-variance pour déduire des interactions bactériennes positives (coopératives) et négatives (compétitives) (par exemple Ruan 2006), mais la co-variance pourrait également indiquer que les populations répondent simultanément à d'autres stimuli.

Une deuxième stratégie pour suivre les interactions bactériennes consiste à suivre les gènes connus comme étant caractéristique des interactions bactériennes. Il faut ensuite rechercher s'ils augmentent ou diminuent dans les échantillons avec des tendances similaires à ce qui peut être observé dans les réseaux d'interaction. La limitation actuelle de cette approche, c'est qu'il n'y a actuellement aucun consensus clair sur les gènes qui sont indubitablement de

fiables pour la coopération et la compétition. En effet, nous pouvons remarquer qu'un grand nombre de ces gènes sont impliqués dans des processus liés à la fois à la concurrence et à la coopération. Par exemple, la sécrétion d'exopolysaccharides (EPS) est souvent considérée comme un trait coopératif car les biofilms présentent de nombreuses synergies entre eux (Faust et al. 2012). Néanmoins, Oliveira et al. (2015) ont observé que la sécrétion d'EPS pouvait également être déclenchée par une exposition à des concentrations d'antibiotiques non-létales, montrant un lien avec la compétition. C'est pourquoi nous avons choisi de nous concentrer sur un nouveau type de gène caractéristique pour la coopération qui sont des gènes de structure des plasmidies. En effet, plusieurs articles ont montré que la collaboration pouvait être maintenue par des échanges génétiques et des scientifiques ont également observé que les gènes codant pour les biens publics étaient préférentiellement localisés sur les éléments mobiles et les points chauds de recombinaison dans les génomes bactériens (Dimitriu et al. 2014; 2015; 2016; Nogueira et al. 2009). Concernant le type de gènes utilisé pour évaluer la compétition, nous avons décidé de sélectionner des gènes de résistance aux antibiotiques car ils sont également considérés dans la littérature comme seuls signes de compétition d'interférence. Cordero et al. (2012) ont observé que la sécrétion d'antibiotiques peut également être un bien public. C'est pourquoi il est également important, lors de la comparaison de l'abondance des gènes de résistance aux antibiotiques dans le temps, de suivre l'augmentation de la diversité de ces gènes. Si le nombre de gènes différents augmente, nous pouvons facilement exclure le fait que la sécrétion d'antibiotiques est un bien public car la communauté ne partage pas une petite quantité de résistance mais affiche une augmentation du nombre de composés toxiques sécrétés qui est plus compatible avec l'hypothèse d'une concurrence accrue entre les différents membres de la communauté bactérienne.

Comme chacune des deux méthodes d'études ne nécessitant pas de cultures sont limitées, nous avons décidé de les utiliser toutes les deux en simultané afin de renforcer les résultats de notre étude.

Hypothèses de la thèse

Sur la base des théories précédentes que nous avons exposées dans cette introduction, nous avons émis l'hypothèse qu'une augmentation des acides organiques dans le réchauffement de la neige de printemps augmenterait la concurrence (et réduirait la collaboration).
Cette hypothèse est appuyée par le fait qu'une augmentation du carbone pourrait augmenter la compétition d'interférence comme observé par Hol et al. (2014). Nous nous attendons également à ce qu'une croissance bactérienne provoquant une augmentation du stress bactérien augmente également la compétition, comme le soutient la théorie de la détection de la concurrence de Cornforth and Foster (2013).
En opposition, la coopération pourrait être plus élevée dans un environnement nutritif plus limité, comme l'ont montré Benomar et al. (2015) que les échanges de métabolites pourraient être déclenchés par des stress nutritionnels.

Résumé des différents chapitres de la thèse

Le premier chapitre de cette thèse établi une brève revue de la littérature scientifique concernant les interactions bactériennes et a déjà été résumé plus largement au cours des parties précédentes de la synthèse. En outre, nous avons également présenté les différents outils permettant de suivre les interactions bactériennes et justifié nos choix méthodologiques.

Au cours du second chapitre, nous avons examiné l'effet des changements de carbone organique sur les communautés microbiennes de neige in situ sur deux mois. Nous avons comparé les communautés bactériennes de neige d'une période à faible teneur en carbone organique à celles d'une période de carbone organique plus élevée. Nous avons émis l'hypothèse qu'une augmentation de la teneur en carbone ferait passer l'interaction microbienne dominante de la collaboration à la compétition. Pour évaluer les interactions microbiennes, nous avons construit des réseaux taxonomiques de co-variance à partir d'OTU obtenus à partir du séquençage du gène de l'ARNr 16S. De plus, nous avons suivi les gènes marqueurs de la coopération microbienne (gènes du squelette plasmidique) et de la compétition (gènes de résistance aux antibiotiques) à travers les deux périodes d'échantillonnage dans les métagénomes et les métatranscriptomes. Nos résultats ont montré une diminution de la connectivité moyenne du réseau à la fin du printemps par rapport au début du printemps que nous avons interprété comme une diminution de la coopération. Cette observation a été renforcée par les gènes de squelette plasmidique significativement plus abondants dans les métagénomes du début du printemps. La modularité du réseau à partir de la fin du printemps s'est également avérée supérieure à celle du début du printemps, ce qui est un autre indicateur possible d'une concurrence accrue. Les gènes de résistance aux antibiotiques étaient significativement plus abondants dans les métagénomes de la fin du printemps. De plus, les gènes de résistance aux antibiotiques étaient également positivement corrélés à la teneur en carbone organique de la neige au cours des deux saisons. La teneur en carbone organique de la neige pourrait être responsable de ce changement dans les interactions bactériennes dans la communauté de neige de l'Arctique.

En parallèle de ces investigations concernant les interactions bactériennes, nous avons également réalisé un pipeline permettant d'améliorer la qualité des annotations métagénomiques réalisées dans la partie suivante de cette thèse. La motivation principale étant de pouvoir quantifier plus précisément avec une plus grande certitude les gènes impliqués dans les interactions bactériennes. La limitation dans notre design expérimental étant que nous utilisons une technique de séquençage ayant un faible débit (miSeq) comparativement au standard de métagénomique (hiSeq) car nous réalisons un échantillonnage assez important (près d'une centaine d'échantillons) en séries temporelle. Cette contrainte méthodologique liée à notre étude a pour conséquence que l'assemblage des séquences obtenues par séquençage n'est pas très exhaustif (moins de 50% des séquences parviennent à être assemblées). Dans ce chapitre, nous présenterons donc un nouveau pipeline conçu pour traiter spécifiquement un tel ensemble de données. Nous avons recouru au co-assemblage et utilisé une stratégie d'annotation de séquences pour compléter l'exhaustivité des annotations afin de récupérer les séquences qui ne pouvaient pas être cartographiées sur les contigs assemblés. De plus, afin d'éviter d'ajouter trop de bruit lors du

sauvetage des séquences en utilisant l'annotation de lecture, nous avons construit un algorithme pour définir un seuil de valeur e basé sur le bruit de l'annotation de séquences appris des séquences utilisées dans l'assemblage.

Pour concevoir un pipeline, nous avons sélectionné plusieurs outils récents connus pour être efficaces pour effectuer l'assemblage, la cartographie(mapping), le regroupement et l'annotation de ces données. De plus, ce pipeline a également été construit dans le but d'être très convivial en termes d'installation. Très souvent, les pipelines pour la métagénomique nécessitent d'installer de nombreux outils ou dépendances séparément. La conséquence étant que certaines connaissances préalables en informatique sont nécessaires pour utiliser de tels outils. De plus, parfois, la reproductibilité d'un tel outil peut être délicate si trop de différences dans l'installation sont faites par les utilisateurs. Dans cette idée de reproductibilité, d'accessibilité et de transparence, nous avons également conçu un script d'installation pour permettre à chaque utilisateur d'installer chaque outil nécessaire au pipeline de manière simple et reproductible. Concernant les performances de ce pipeline, nous avons pu montrer que le taux d'erreur attendu (False discovery rate) pour l'annotation était proche de 5%. Enfin, nous avons également utilisé un jeu de données réel concernant un site de bioreméditation et montré que la représentabilité des échantillons semblait bien meilleure lorsque nous utilisions notre pipeline que lorsque nous utilisions une stratégie d'assemblage de métagénomes classique. Néanmoins il reste encore du chemin avant de pouvoir publier cet outil car une validation plus poussée devrait être effectuée afin de mieux caractériser les performances de notre outil.

Notre première tentative d'étude des interactions bactériennes dans la neige Arctique a montré combien il était difficile d'établir de manière fiable un effet de la concentration d'acide organique sur les interactions bactériennes en raison du niveau élevé de facteurs de confusion possibles lors d'une étude *in situ (Bergk Pinto et al. 2019)*. Pour cette raison, nous avons décidé de valider davantage notre hypothèse originale en étudiant l'effet de la concentration d'acides organiques dans une expérience de microcosme de neige. Nous avons construit une expérience de séries temporelles où nous avons comparé l'évolution d'une communauté bactérienne de l'Arctique des neiges exposée à une forte concentration d'acides organiques à sa population d'origine dans une série temporelle de contrôle. Afin de suivre et comparer les interactions bactériennes, nous avons décidé de continuer à appliquer notre double approche. Suite à un manque de signal, nous avons dû quelque peu adapter notre méthodologie en étudiant également d'autres gènes indicateurs d'interactions bactériennes. Nous avons pu confirmer le fait que l'augmentation de la concentration d'acides organiques augmentait la compétition entre bactéries mais nous n'avons pas observé d'impact significatifs sur la coopération qui ne différait pas beaucoup du niveau observé dans les microcosmes contrôles. Cette tendance à également été confirmée dans les réseaux de co-variance où le pourcentage d'interactions négatives détecté dans le réseau des microcosmes amendés avec l'acide organique était près de quatre fois supérieur à celui observé dans le réseau des microcosmes contrôles. Nous en avons déduit que la concentration d'acides organique dans la neige affectait principalement la compétition entre bactéries mais n'avait pas ou peu d'effet sur la collaboration dans la communauté bactérienne de la neige arctique.

Malgré ces résultats encourageants, le présent travail a également mis en lumière la difficulté de pouvoir interpréter de manière univoque les gènes qui participent aux interactions bactériennes et suggère la mise en place d'une base de données spécialement dédiée à ce

type de gènes. En effet, le recoupement des définitions des gènes impliqués dans le métabolisme lié aux antibiotiques diffère par exemple entre la Gene Onthology et la base de donnée KEGG ce qui rend parfois laborieux l'analyse des données et ajoute une certaine difficulté à l'interprétation des résultats.

Bibliographie

Barberán, Albert, Scott T Bates, Emilio O Casamayor, and Noah Fierer. 2012. "Using Network Analysis to Explore Co-Occurrence Patterns in Soil Microbial Communities." *The ISME Journal* 6 (2): 343–51. https://doi.org/10.1038/ismej.2011.119.

Beal, Emily J., Christopher H. House, and Victoria J. Orphan. 2009. "Manganese- and Iron-Dependent Marine Methane Oxidation." *Science* 325 (5937): 184–87. https://doi.org/10.1126/science.1169984.

Benomar, Saida, David Ranava, María Luz Cárdenas, Eric Trably, Yan Rafrafi, Adrien Ducret, Jérôme Hamelin, Elisabeth Lojou, Jean-Philippe Steyer, and Marie-Thérèse Giudici-Orticoni. 2015. "Nutritional Stress Induces Exchange of Cell Material and Energetic Coupling between Bacterial Species." *Nature Communications* 6 (February): 6283. https://doi.org/10.1038/ncomms7283.

Bergk Pinto, Benoît, Lorrie Maccario, Aurélien Dommergue, Timothy M. Vogel, and Catherine Larose. 2019. "Do Organic Substrates Drive Microbial Community Interactions in Arctic Snow?" *Frontiers in Microbiology* 10. https://doi.org/10.3389/fmicb.2019.02492.

Burgin, Amy J., Wendy H. Yang, Stephen K. Hamilton, and Whendee L. Silver. 2011. "Beyond Carbon and Nitrogen: How the Microbial Energy Economy Couples Elemental Cycles in Diverse Ecosystems." *Frontiers in Ecology and the Environment* 9 (1): 44–52. https://doi.org/10.1890/090227.

Chignell, J. F., S. Park, C. M. R. Lacerda, S. K. De Long, and K. F. Reardon. 2018. "Label-Free Proteomics of a Defined, Binary Co-Culture Reveals Diversity of Competitive Responses Between Members of a Model Soil Microbial System." *Microbial Ecology* 75 (3): 701–19. https://doi.org/10.1007/s00248-017-1072-1.

Cordero, Otto X., Hans Wildschutte, Benjamin Kirkup, Sarah Proehl, Lynn Ngo, Fatima Hussain, Frederique Le Roux, Tracy Mincer, and Martin F. Polz. 2012. "Ecological Populations of Bacteria Act as Socially Cohesive Units of Antibiotic Production and Resistance." *Science* 337 (6099): 1228–31. https://doi.org/10.1126/science.1219385.

Cornforth, Daniel M., and Kevin R. Foster. 2013. "Competition Sensing: The Social Side of Bacterial Stress Responses." *Nature Reviews. Microbiology* 11 (4): 285–93. https://doi.org/10.1038/nrmicro2977.

Darwin, Charles. 1859. *On The Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Dimitriu, Tatiana, Chantal Lotton, Julien Bénard-Capelle, Dusan Misevic, Sam P. Brown, Ariel B. Lindner, and François Taddei. 2014. "Genetic Information Transfer Promotes Cooperation in Bacteria." *Proceedings of the National Academy of Sciences of the United States of America* 111 (30): 11103–8. https://doi.org/10.1073/pnas.1406840111.

Dimitriu, Tatiana, Dusan Misevic, Ariel B Lindner, and François Taddei. 2015. “Mobile Genetic Elements Are Involved in Bacterial Sociality.” *Mobile Genetic Elements* 5 (1): 7–11. https://doi.org/10.1080/2159256X.2015.1006110.

Dimitriu, Tatiana, Dusan Misevic, Chantal Lotton, Sam P. Brown, Ariel B. Lindner, and François Taddei. 2016. “Indirect Fitness Benefits Enable the Spread of Host Genes Promoting Costly Transfer of Beneficial Plasmids.” *PLOS Biology* 14 (6): e1002478. https://doi.org/10.1371/journal.pbio.1002478.

Ding, Junjun, Yuguang Zhang, Ye Deng, Jing Cong, Hui Lu, Xin Sun, Caiyun Yang, et al. 2015. “Integrated Metagenomics and Network Analysis of Soil Microbial Community of the Forest Timberline.” *Scientific Reports* 5 (January): 7994. https://doi.org/10.1038/srep07994.

D’Onofrio, Anthony, Jason M. Crawford, Eric J. Stewart, Kathrin Witt, Ekaterina Gavrish, Slava Epstein, Jon Clardy, and Kim Lewis. 2010. “Siderophores from Neighboring Organisms Promote the Growth of Uncultured Bacteria.” *Chemistry & Biology* 17 (3): 254–64. https://doi.org/10.1016/j.chembiol.2010.02.010.

D’Souza, Glen, and Christian Kost. 2016. “Experimental Evolution of Metabolic Dependency in Bacteria.” *PLOS Genetics* 12 (11): e1006364. https://doi.org/10.1371/journal.pgen.1006364.

Faust, Karoline, and Jeroen Raes. 2012. “Microbial Interactions: From Networks to Models.” *Nature Reviews Microbiology* 10 (8): 538–50. https://doi.org/10.1038/nrmicro2832.

Faust, Karoline, J. Fah Sathirapongsasuti, Jacques Izard, Nicola Segata, Dirk Gevers, Jeroen Raes, and Curtis Huttenhower. 2012. “Microbial Co-Occurrence Relationships in the Human Microbiome.” *PLOS Computational Biology* 8 (7): e1002606. https://doi.org/10.1371/journal.pcbi.1002606.

Finkel, Omri M., Isai Salas-González, Gabriel Castrillo, Theresa F. Law, Jonathan M. Conway, Corbin D. Jones, and Jeffery L. Dangl. 2019. “Root Development Is Maintained by Specific Bacteria-Bacteria Interactions within a Complex Microbiome.” *BioRxiv*, May, 645655. https://doi.org/10.1101/645655.

Foster, Kevin R., and Thomas Bell. 2012. “Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species.” *Current Biology* 22 (19): 1845–50. https://doi.org/10.1016/j.cub.2012.08.005.

Hansen, Lea Benedicte Skov, Dawei Ren, Mette Burmølle, and Søren J. Sørensen. 2017. “Distinct Gene Expression Profile of *Xanthomonas Retroflexus* Engaged in Synergistic Multispecies Biofilm Formation.” *The ISME Journal* 11 (1): 300–303. https://doi.org/10.1038/ismej.2016.107.

Herschend, Jakob, Zacharias B. V. Damholt, Andrea M. Marquard, Birte Svensson, Søren J. Sørensen, Per Hägglund, and Mette Burmølle. 2017. “A Meta-Proteomics Approach to Study the Interspecies Interactions Affecting Microbial Biofilm Development in a Model Community.” *Scientific Reports* 7 (1): 16483. https://doi.org/10.1038/s41598-017-16633-6.

Ho, Adrian, Roey Angel, Annelies J. Veraart, Anne Daebeler, Zhongjun Jia, Sang Yoon Kim, Frederiek-Maarten Kerckhof, Nico Boon, and Paul L. E. Bodelier. 2016. “Biotic Interactions in Microbial Communities as Modulators of Biogeochemical Processes: Methanotrophy as a Model System.” *Frontiers in Microbiology* 7 (August). https://doi.org/10.3389/fmicb.2016.01285.

Hol, Felix JH, Mathias J. Voges, Cees Dekker, and Juan E. Keymer. 2014. “Nutrient-Responsive Regulation Determines Biodiversity in a Colicin-Mediated Bacterial

Community." *BMC Biology* 12 (August): 68. https://doi.org/10.1186/s12915-014-0068-2.

Holland, Alexandra T., Benoît Bergk Pinto, Rose Layton, Christopher J. Williamson, Alexandre M. Anesio, Timothy M. Vogel, Catherine Larose, and Martyn Tranter. 2020. "Over Winter Microbial Processes in a Svalbard Snow Pack: An Experimental Approach." *Frontiers in Microbiology* 11. https://doi.org/10.3389/fmicb.2020.01029.

Khan, Nymul, Yukari Maezato, Ryan S. McClure, Colin J. Brislawn, Jennifer M. Mobberley, Nancy Isern, William B. Chrisler, et al. 2018. "Phenotypic Responses to Interspecies Competition and Commensalism in a Naturally-Derived Microbial Co-Culture." *Scientific Reports* 8 (1): 297. https://doi.org/10.1038/s41598-017-18630-1.

Konopka, Allan, Stephen Lindemann, and Jim Fredrickson. 2015. "Dynamics in Microbial Communities: Unraveling Mechanisms to Identify Principles." *ISME J* 9 (7): 1488–95.

Lima-Mendez, Gipsi, Karoline Faust, Nicolas Henry, Johan Decelle, Sébastien Colin, Fabrizio Carcillo, Samuel Chaffron, et al. 2015. "Determinants of Community Structure in the Global Plankton Interactome." *Science* 348 (6237): 1262073. https://doi.org/10.1126/science.1262073.

Maccario, Lorrie, Laura Sanguino, Timothy M. Vogel, and Catherine Larose. 2015. "Snow and Ice Ecosystems: Not so Extreme." *Research in Microbiology*, Special issue on Microbial diversity, adaptation and evolution, 166 (10): 782–95. https://doi.org/10.1016/j.resmic.2015.09.002.

Mas, Alix, Shahrad Jamshidi, Yvan Lagadeuc, Damien Eveillard, and Philippe Vandenkoornhuyse. 2016. "Beyond the Black Queen Hypothesis." *ISME Journal* 10 (9): 2085–91. https://doi.org/10.1038/ismej.2016.22.

McClure, Ryan S., Christopher C. Overall, Eric A. Hill, Hyun-Seob Song, Moiz Charania, Hans C. Bernstein, Jason E. McDermott, and Alexander S. Beliaev. 2018. "Species-Specific Transcriptomic Network Inference of Interspecies Interactions." *The ISME Journal* 12 (8): 2011. https://doi.org/10.1038/s41396-018-0145-6.

Mikucki, Jill A., Ann Pearson, David T. Johnston, Alexandra V. Turchyn, James Farquhar, Daniel P. Schrag, Ariel D. Anbar, John C. Priscu, and Peter A. Lee. 2009. "A Contemporary Microbially Maintained Subglacial Ferrous 'Ocean.'" *Science* 324 (5925): 397–400. https://doi.org/10.1126/science.1167350.

Molina-Santiago, Carlos, Zulema Udaondo, Baldo F. Cordero, and Juan L. Ramos. 2017. "Interspecies Cross-Talk between Co-Cultured Pseudomonas Putida and Escherichia Coli." *Environmental Microbiology Reports* 9 (4): 441–48. https://doi.org/10.1111/1758-2229.12553.

Morris, J. Jeffrey. 2015. "Black Queen Evolution: The Role of Leakiness in Structuring Microbial Communities." *Trends in Genetics* 31 (8): 475–82. https://doi.org/10.1016/j.tig.2015.05.004.

Musat, Niculina, Florin Musat, Peter Kilian Weber, and Jennifer Pett-Ridge. 2016. "Tracking Microbial Interactions with NanoSIMS." *Current Opinion in Biotechnology*, Analytical biotechnology, 41 (October): 114–21. https://doi.org/10.1016/j.copbio.2016.06.007.

Nogueira, Teresa, Daniel J. Rankin, Marie Touchon, François Taddei, Sam P. Brown, and Eduardo P. C. Rocha. 2009. "Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence." *Current Biology* 19 (20): 1683–91. https://doi.org/10.1016/j.cub.2009.08.056.

Oliveira, Nuno M., Esteban Martinez-Garcia, Joao Xavier, William M. Durham, Roberto Kolter, Wook Kim, and Kevin R. Foster. 2015. "Biofilm Formation As a Response to

Ecological Competition." *PLOS Biology* 13 (7): e1002191. https://doi.org/10.1371/journal.pbio.1002191.
Overbeek, Leonard S. van, and Kari Saikkonen. 2016. "Impact of Bacterial–Fungal Interactions on the Colonization of the Endosphere." *Trends in Plant Science*, Special Issue: Unravelling the Secrets of the Rhizosphere, 21 (3): 230–42. https://doi.org/10.1016/j.tplants.2016.01.003.
Post, Eric, Mads C. Forchhammer, M. Syndonia Bret-Harte, Terry V. Callaghan, Torben R. Christensen, Bo Elberling, Anthony D. Fox, et al. 2009. "Ecological Dynamics across the Arctic Associated with Recent Climate Change." *Science (New York, N.Y.)* 325 (5946): 1355–58. https://doi.org/10.1126/science.1173113.
Ren, Dawei, Jonas S Madsen, Soren J Sorensen, and Mette Burmolle. 2015. "High Prevalence of Biofilm Synergy among Bacterial Soil Isolates in Cocultures Indicates Bacterial Interspecific Cooperation." *ISME J* 9 (1): 81–89.
Rønn, Regin, Mette Vestergård, and Flemming Ekelund. 2015. "Interactions Between Bacteria, Protozoa and Nematodes in Soil." *Acta Protozoologica* 51 (3): 223–35.
Ruan, Q. 2006. "Local Similarity Analysis Reveals Unique Associations among Marine Bacterioplankton Species and Environmental Factors." *Bioinformatics* 22: 2532–38. https://doi.org/10.1093/bioinformatics/btl417.
Schlesinger, William H, Jonathan J Cole, Adrien C Finzi, and Elisabeth A Holland. 2011. "Introduction to Coupled Biogeochemical Cycles." *Frontiers in Ecology and the Environment* 9 (1): 5–8. https://doi.org/10.1890/090235.
Sinha, Rajeshwar P., and Donat-P. Häder. 2002. "UV-Induced DNA Damage and Repair: A Review." *Photochemical & Photobiological Sciences* 1 (4): 225–36. https://doi.org/10.1039/B201230H.
Tao, Jiemeng, Delong Meng, Chong Qin, Xueduan Liu, Yili Liang, Yunhua Xiao, Zhenghua Liu, Yabing Gu, Juan Li, and Huaqun Yin. 2018. "Integrated Network Analysis Reveals the Importance of Microbial Interactions for Maize Growth." *Applied Microbiology and Biotechnology* 102 (8): 3805–18. https://doi.org/10.1007/s00253-018-8837-4.
Torres-Monroy, Ingrid, and Matthias S. Ullrich. 2018. "Identification of Bacterial Genes Expressed During Diatom-Bacteria Interactions Using an in Vivo Expression Technology Approach." *Frontiers in Marine Science* 5. https://doi.org/10.3389/fmars.2018.00200.
Van Valen, Lee. 1973. "A New Evolutionary Law." *Evolutionary Theory* 1: 1–30.
Vincent, Warwick F., Terry V. Callaghan, Dorthe Dahl-Jensen, Margareta Johansson, Kit M. Kovacs, Christine Michel, Terry Prowse, James D. Reist, and Martin Sharp. 2011. "Ecological Implications of Changes in the Arctic Cryosphere." *Ambio* 40 (Suppl 1): 87–99. https://doi.org/10.1007/s13280-011-0218-5.
Vincent, Warwick F., Lyle G. Whyte, Connie Lovejoy, Charles W. Greer, Isabelle Laurion, Curtis A. Suttle, Jacques Corbeil, and Derek R. Mueller. 2009. "Arctic Microbial Ecosystems and Impacts of Extreme Warming during the International Polar Year." *Polar Science*, MERGE, 3 (3): 171–80. https://doi.org/10.1016/j.polar.2009.05.004.
Yin, Huaqun, Jiaojiao Niu, Youhua Ren, Jing Cong, Xiaoxia Zhang, Fenliang Fan, Yunhua Xiao, et al. 2015. "An Integrated Insight into the Response of Sedimentary Microbial Communities to Heavy Metal Contamination." *Scientific Reports* 5 (September): srep14266. https://doi.org/10.1038/srep14266.
Zelezniak, Aleksej, Sergej Andrejev, Olga Ponomarova, Daniel R. Mende, Peer Bork, and Kiran Raosaheb Patil. 2015. "Metabolic Dependencies Drive Species Co-Occurrence

in Diverse Microbial Communities." *Proceedings of the National Academy of Sciences* 112 (20): 6449–54. https://doi.org/10.1073/pnas.1421834112.
Zeng, Hong, and Aidong Yang. 2019. "Modelling Overflow Metabolism in Escherichia Coli with Flux Balance Analysis Incorporating Differential Proteomic Efficiencies of Energy Pathways." *BMC Systems Biology* 13 (1): 3. https://doi.org/10.1186/s12918-018-0677-4.
Zengler, Karsten, and Livia S. Zaramela. 2018. "The Social Network of Microorganisms — How Auxotrophies Shape Complex Communities." *Nature Reviews Microbiology* 16 (6): 383. https://doi.org/10.1038/s41579-018-0004-5.
Ziegelhoffer, E.C., and T.J. Donohue. 2009. "Bacterial Responses to Photo-Oxidative Stress." *Nature Reviews Microbiology* 7 (12): 856–63. https://doi.org/10.1038/nrmicro2237.

Chapter I - Introduction: Interactions matter in microbiology

This thesis will focus on microbial interactions in the Arctic snow and how organic acid concentrations affect them. We will begin by formally introducing interactions in biology in an evolutionary context. We will summarize the different theories and hypotheses published by microbiologists to explain how cooperative interactions can be selected for in the environment and also explore their importance in microbiology. We will then discuss the factors that influence these bacterial interactions create dynamic changes in this relationship. The main focus will be on the links between nutrition strategies and bacterial interactions as this thesis centers on the effects of organic acids on microbial communities. We will also discuss the state of the art and the methodology used for studying interactions in natural communities, given that the vast majority of the current knowledge has been generated in lab experiments. We will finally introduce the environment chosen for this study (the arctic snow) and detail our hypotheses.

1 Definition and biological context for interactions in microbiology

The word “interaction” is defined by the Cambridge dictionary as “an occasion when two or more people or things communicate with or react to each other”. When applying this definition to the biological field, it would define any communication or reaction between two or more organisms.

Bacteria, as all other living organisms, interact with their environment but also with each other and other microorganisms (Faust and Raes 2012). These interactions can be classified in different categories depending on whether it is neutral, detrimental or beneficial for one or both partners. Lidicker (1979) summarized all these possible interactions into an “intra-action compass” (Figure 1). The outcome of the interaction is determined by the effect of the interaction on the fitness of the bacteria involved (Faust and Raes 2012). The fitness of an organism can be defined as its ability to survive in a particular environment and its capacity to reproduce and participate in the next generation’s gene pool (Barker 2009; Orr 2009). Since bacteria grow by clonal division, their fitness can be defined as the number of child cells they are able to produce. If the bacterium increases its offspring due to an interaction, then it will be termed positive. If the interaction decreases this amount, it will be termed negative.

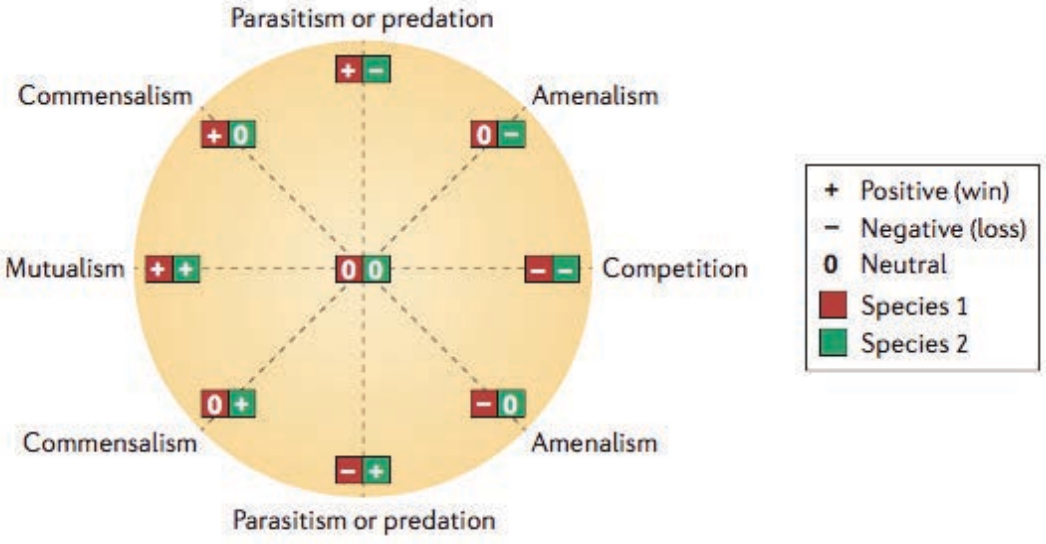


Figure 1: The intra-action compass figure from Lidicker (1979) modified by Faust and Raes (2012) to display all the possible pairwise interactions. The outcome of the interaction can be beneficial (win fitness = positive), detrimental (lose fitness = negative) or neutral (no impact on the fitness = null) for the species interacting. For example, if we consider predation, the predator wins fitness in the interaction because it consumes its prey and the prey loses fitness because it's consumed by the predator. At the species level, the interaction is positive for the predator species and negative for the prey species.

Several ecological concepts have considered biological interactions, including Charles Darwin's theory of evolution (Darwin 1859), which postulates that species not adapted to their environment would go extinct. When considering the environment, interactions with other species that can increase or decrease a specie's fitness should also be included (e.g. Gillott 1995). The importance of interactions in the evolutionary process was first formalized by Van Valen (1973) in his Red Queen hypothesis. He stated that to survive, species had to continuously adapt to their environment but also to the species with which they were competing. This "evolutionary run" would never end, because each species is trying to reach an optimum which is mutually incompatible with the one(s) of its competitor(s). This theory also states that the main kind of interactions in nature are competition or parasitism: "The Red Queen proposes that events of mutualism, at least on the same trophic level, are of little importance in evolution in comparison to negative interactions (…)" (Van Valen 1973). This statement may have contributed to a focus on studying competitive interactions in microbiology, which were considered to be dominant among the different forms, as also stated by Foster and Bell (2012). As a consequence, changes in bacterial community structure were hypothesized to result mainly from a decrease in fitness due to a lower growth rate under changing environmental conditions (e.g. change of pH) or to competitive interactions (e.g. antibiotics, competition for space or preferential nutrient access).

2 Bacterial interactions: cooperation matters

The importance of positive bacterial interactions has recently become a focus of research. Publications related to bacterial interactions, and especially positive interactions, are growing faster than those on negative interactions (Figure 2). This shift could be related to the release of a new evolutionary hypothesis supporting the appearance of collaboration from competitive interactions. This hypothesis, called the "Black Queen's hypothesis", supports the theory that gene loss of costly leaky functions (metabolic functions that release their end product by diffusion in the environment, thus producing public goods) can result in an increased fitness at the individual level if the number of producers in the community (also called helpers) are still big enough to support community growth. Bacteria losing such genes become beneficiaries of the helpers (Morris 2015). This metabolic coupling was also hypothesized to enable bacteria to avoid exclusive competition by making the two partners reach a steady state in the environment, thus avoiding a disappearance of one of the two species and reducing competition (Mas et al. 2016).

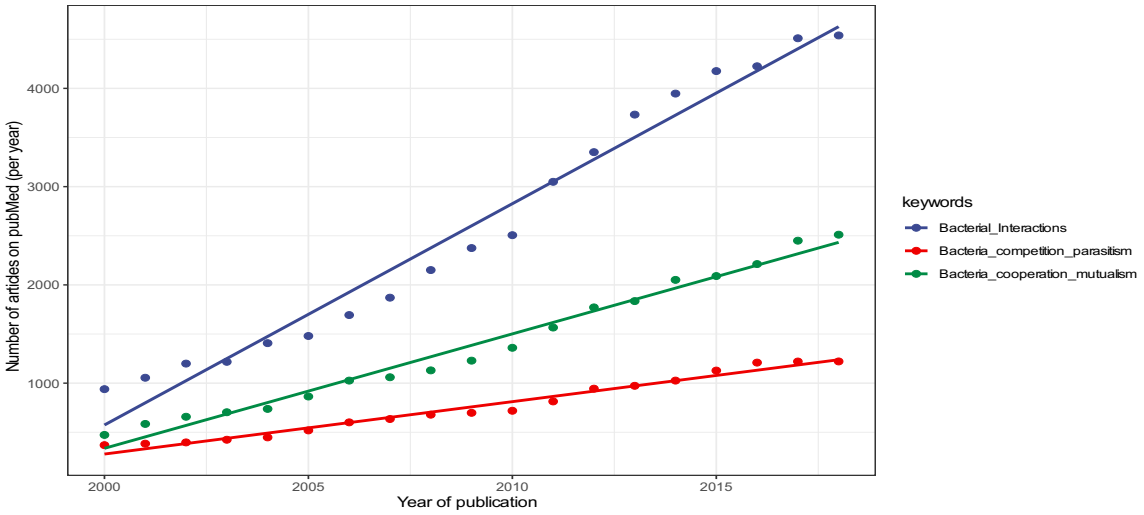


Figure 2: plot showing the number of articles published between the year 2000 until 2018 (by year) on PubMed using keywords related to bacterial interactions (blue). The number of articles published with keywords related to positive interactions (green) is growing faster than the number of articles with keywords related to negative interactions (red).

This theory might also explain the low cultivability of microorganisms, known as the "great plate count anomaly", because if a significant number of microorganisms require public good metabolites, they could be auxotroph for one or several specific metabolites (Zengler and Zaramela 2018) and therefore be unable to grow without their microbial partners. Recent culture experiments that were able to isolate previously uncultured microorganisms by adding siderophores (D'Onofrio et al. 2010; Vartoukian et al. 2016) support this hypothesis. In addition, recent co-culture experiments on bacterial strains isolated from soils by Ren et al. (2015) showed a number of positive synergies (higher growth rates in co-culture than in pure culture), supporting the fact that metabolic couplings could be important in bacterial communities. D'Souza and Kost (2016) also showed that bacterial strains lost biosynthesis genes for metabolites present within the culture media by selection, which confirms the hypothesis that organisms can lose the ability to perform functions whose products are available from the environment (Morris, 2015).

2.1 Importance of bacterial interactions

Bacterial interactions can be very diverse and can also involve other interacting taxa like plants (Finkel et al. 2019), fungi (van Overbeek and Saikkonen 2016) or animals (Rønn, Vestergård, and Ekelund 2015). The significance of these interactions is becoming more apparent through recent studies. For example, plant root development has been shown to be affected by bacteria-bacteria interactions (Finkel et al. 2019) and that maize yield could be correlated to specific bacterial interactions (Tao et al. 2018). Bacterial interactions have also been shown to modulate the reproductive success of diatoms (Torres-Monroy and Ullrich 2018). Bacterial interactions can also modulate biogeochemical cycles (Ho et al. 2016). If two interacting bacterial species mediate reactions from different biogeochemical cycles, biogeochemical coupling can be observed, leading to unpredicted metabolic dependencies between cycles (Schlesinger et al. 2011; Burgin et al. 2011). For example, Beal, House, and Orphan (2009) discovered that methane oxidation could be coupled to manganese or even iron instead of sulfate reduction in marine methane-seep sediments. Another example of biogeochemical coupling was discovered in an ancient marine brine system isolated from the atmosphere in the Antarctic. Mikucki et al. (2009) observed that an active microbial community was cycling sulfur through coupling with iron reduction to Fe (II). This enabled the microbial community to grow despite its isolation (no photosynthesis causing a limited input of carbon in the community) (Mikucki et al. 2009).

3 Bacterial interactions can change

As bacterial interactions attract more and more attention, the stability of these interactions has also begun to be investigated. Several environmental parameters have been studied mainly during lab experiments (Table 1). Each of these parameters has been shown to select for cooperation or counter-select for it by increasing competition. Therefore, given that nutrient concentrations, pH, water content and the homogeneity of the environment can vary across time, bacterial interactions can be seen as dynamic links between organisms. This concept can be illustrated by the co-culture experiment carried out by Benomar et al. (2015), in which *Desulfovibrio vulgaris*, a Gram-positive sulfate reducing bacterium, was co-cultured with *Clostridium acetobutylicum, a* Gram-negative bacterium. The authors showed that metabolic exchanges between species only began once sulfate was depleted in the culture medium. The consequence of dynamic interactions is that the outcome of the structure of a defined bacterial community cannot be predicted only by the individual metabolisms of the organisms present, but also by possible couplings and interaction potentials among the different members. Konopka, Lindemann, and Fredrickson (2015) modelled these dynamics and observed that bacterial interactions could generate endogenous dynamics affecting the bacterial community even in the absence of exogeneous perturbations. They also showed that the interaction network could improve the resistance of the community to perturbation and should be considered in addition to metabolic redundancy (Konopka, Lindemann, and Fredrickson 2015).

The study of the effect of each of the environmental variables on bacterial interactions are still ongoing and new effects are still being discovered. As observed in Table 1, not all the studies agree on the effect of each parameter. Such contrasting results can be explained by

the fact that cooperation and competition can also occur at different scales within a bacterial community. For example, Cordero et al. (2012) observed that antibiotic were secreted by some members of a bacterial consortium while the other collaborative strains shared the antibiotic resistance genes. They concluded that antibiotic synthesis could be considered as a collaborative trait in this part of the community and was used to compete against other members of the community. On the other hand, the model bacteria used for these studies can also influence the results and thus show that different collaborative or competitive strategies could be selected for across different environments.

Table 1: Summary of biotic and abiotic factors found to select for competition or cooperation among the bacterial communities. Most of those results have been generated during lab experiments results and/or modelling.

Type of factor	**Factor impacting interaction**	**Selects for / indicates**	**Type of study**	**Article**
Biotic Factors	Antibiotics	Competition Cooperation	Lab experiment	(Vasse et al. 2017; Cordero et al. 2012)
	Genetic information transfer	Cooperation	Lab experiment Bioinformatics	(Dimitriu et al. 2014; 2015; 2016; Nogueira et al. 2009)
	High bacterial density	Cooperation	Lab experiment	(Darch et al. 2012)
	Low bacterial density	Cooperation	Lab experiment	(Ross-Gillespie et al. 2009; Ross-Gillespie et al. 2007)
	Increased mutation rate	Competition (decreases cooperation)	Lab experiment	(Harrison and Buckling 2005)
Abiotic Factors	High nutrient richness	Competition Cooperation	Lab experiment	(Brockhurst et al. 2010; Ponce-Soto et al. 2015; Lambert et al. 2011; Ravindran 2017; F. J. Hol et al. 2014)
	Low nutrient richness	Competition Cooperation	Lab experiment	(F. J. Hol et al. 2014; Ponce-Soto et al. 2015; Velez et al. 2018; Lambert et al. 2011; Lambert, Vyawahare, and Austin 2014; Benomar et al.

				2015; Ravindran 2017; Pande et al. 2015)
	high substrate complexity	Cooperation	Lab experiment	(Deng and Wang 2016; Tecon and Or 2017)
	Oxidative stress	Cooperation	Lab experiment	(John et al. 2017)
	Increased structure of environment	Cooperation	Lab experiment modelling	(Kümmerli et al. 2009; F. J. H. Hol et al. 2013; 2015; Mc Ginty, Rankin, and Brown 2011; Tecon et al. 2018)

4 Nutritional strategies and bacterial interactions

Based on Table 1, nutrient concentrations or composition can influence bacterial interactions. The effect of changes in nutrient concentrations on the bacterial interactions is the main focus of this thesis. Therefore, we will focus on interactions related to nutrients in the following sections. The main types of interactions are defined below:

(A) **Nutritional competition:** This is a special case of competition (Negative-Negative) where both partners compete for a particular nutrient (fig.2.A). Even though one of the two species can win the competition, the interaction is negative for both species, because each one consumes a part of the nutrient pool that is then lost for the other species.

(B) **Syntrophy:** This is a particular case of mutualism (Positive-Positive) (fig.2.B). Two microorganisms cooperate to degrade a compound that they would not be able to degrade alone. Thus, the interaction is positive for both bacteria, since they can metabolize a pool of nutrients that they would not be able to process without the presence of their partner.

(C) **Cross-feeding:** This is a special case of commensalism (Neutral-Positive). The presence of the first species producing an essential nutrient such as a vitamin allows a bacterium unable to produce this compound (auxotroph) to survive (fig.2C). The bacterium producing the nutrient does not benefit from this interaction, but the auxotroph is able to survive (improved fitness) due to the second species.

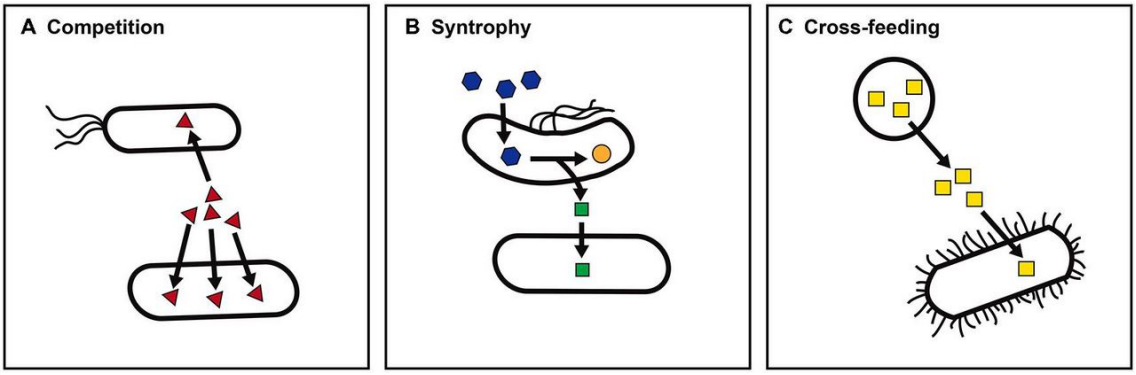


Figure 3: Illustration showing three kinds of nutritional interactions that can be observed among microorganisms. **A** The competition for nutrient uptake will lead to the increase of one bacterial strain and the decrease/disappearance of the bacterial strain that lost the competition. **B** The consumption of a by- product (green square) e.g. hydrogen generated by fermentation of ethanol to acetate by a second strain that reduces CO_2 to CH_4, thus making the first reaction more thermodynamically favorable. **C** A bacterial strain producing an essential nutrient (e.g. vitamin) allows the growth of auxotrophs. Figure from the article of Seth and Taga (2014).

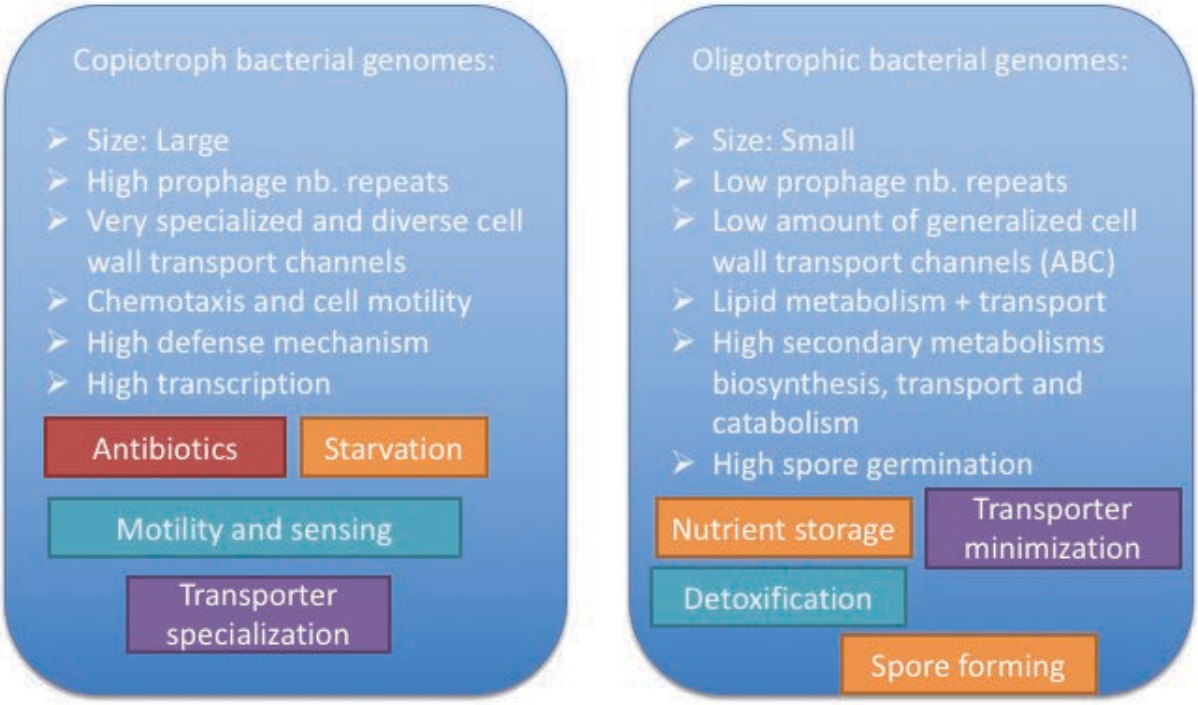
4.1 Copiotrophs versus oligotrophs: is lifestyle linked to bacterial interactions?

A first question arising from these interactions is whether they could result from an adaptation to a specific lifestyle, by selecting a specific kind of interaction (competition or cooperation) based on the nutrient load, for example. Here, we review the different attempts to classify bacteria based on the trophic strategies.

A lot of effort has been invested to classify microorganisms based on their ability to survive in copiotrophic (nutrient rich) or oligotrophic (nutrient poor) environments (see review by Ho et al. 2017). Some researchers have tried to link the taxonomy of the bacterial community members to bacterial properties such as carbon mineralization in soils, for example (Fierer, Bradford, and Jackson 2007). This approach has been difficult to generalize across studies due to the lack of consensus related to using a taxonomic based approach. For example, some organisms previously identified as copiotrophs were identified as oligotrophs and vice versa (Ho et al. 2017). In another attempt to classify bacteria based on lifestyle, researchers have begun using genomic features as proxies for copiotrophic or oligotrophic organisms. For example, Klappenbach, Dunbar, and Schmidt (2000) classified bacterial taxa based on their rRNA copy numbers and observed a dominance of high copy number bacteria in rich media (responding quicker to amendment) and a dominance of low copy number bacteria in the unamended media. In a more complex attempt to classify copiotrophic and oligotrophic bacteria, Lauro et al. (2009) used several bacteria already classified as copiotrophs and oligotrophs and compared their genomes to find categories of genes that were differentially abundant in one of the two groups of genomes. The advantage of this approach is that only metabolic potential is needed to classify bacteria, which can easily be generalized and used to predict lifestyles from metagenomes by comparing gene abundance, for example. Antibiotic related genes were found to be significantly more abundant in copiotroph genomes, suggesting that they could be found in environments with intense competition. Another hypothesis is that these organisms are less nutrient limited and could thus invest more energy and resources into competition-related metabolisms. The second hypothesis

could be supported by the study of F. J. Hol et al. (2014) who observed that an antibiotic sensitive *E.coli* strain could co-exist with a colicin-secreting *E.coli* strain when co-cultivated on a poor growth medium (sugars), but not on a rich medium (amino acids and peptides), where the colicin-secreting *E.coli* strain released antibiotics.

Figure 2: short summary of genomic features detected by (Lauro et al. 2009) as being more abundant in copiotrophs or oligotrophs bacterial genomes.



However, considering competitive interactions as resulting only from adaptations to a given nutrient level in the environment is likely inaccurate since other factors, such as those summarized in Table 1, have also shown to be involved.

4.2 Bacterial competition sensing: how bacteria can use environmental signals to interact

A theory to explain bacterial competition as a consequence of stress, mainly related to nutritional state, but also to cell damages caused by potential competitors, has recently been proposed by Cornforth and Foster (2013). They postulate that bacteria are able to sense competition (i.e. competition sensing) through a physiological response that detects harm caused by other organisms. They suggest that many stress responses in bacteria detect ecological competition by sensing changes in nutrient levels (exploitative competition) or direct cell damage (interference competition).

Competition sensing has the advantage that it reduces the number of factors to consider and creates a framework where competition is mainly dependent on two environmental factors (nutrient stress and cell damage). One limitation of this system is that the selection of cooperation is not discussed as a strategy since the authors suggest that, in their own words, bacteria exist in mainly a "microbe-kill-microbe" world (Cornforth and Foster 2013a).

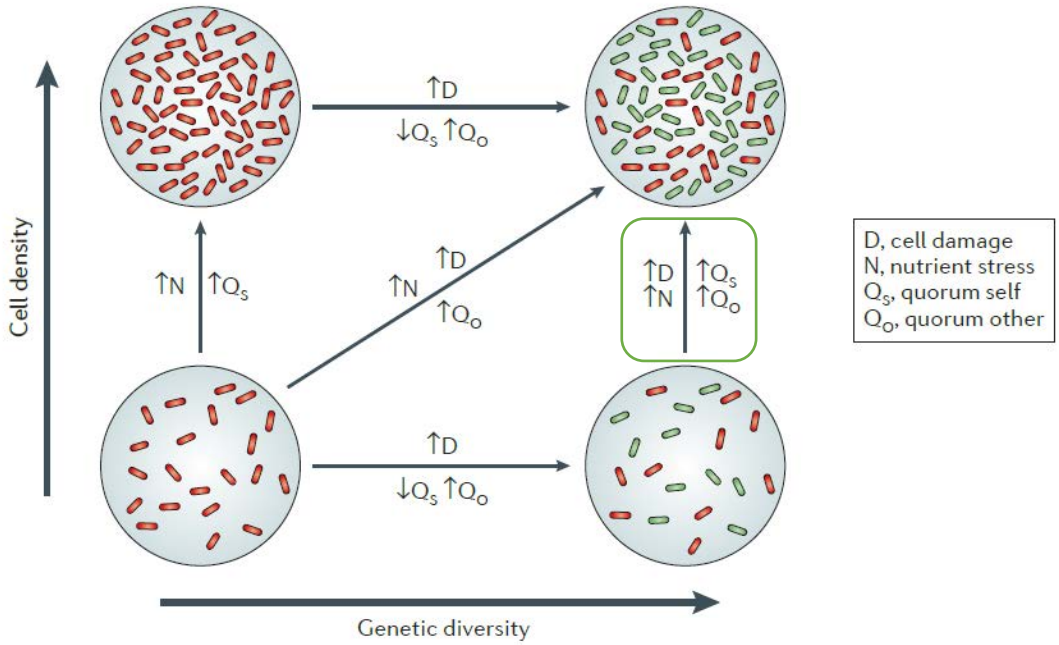


Figure 4: figure from Cornforth and Foster (2013) summarizing the bacterial competition sensing framework. Bacteria can detect the presence of competitors by quorum sensing (Qo increasing) and increasing cell damage (D increasing). The increase of cell density can be detected by nutritional stress (N) increasing.

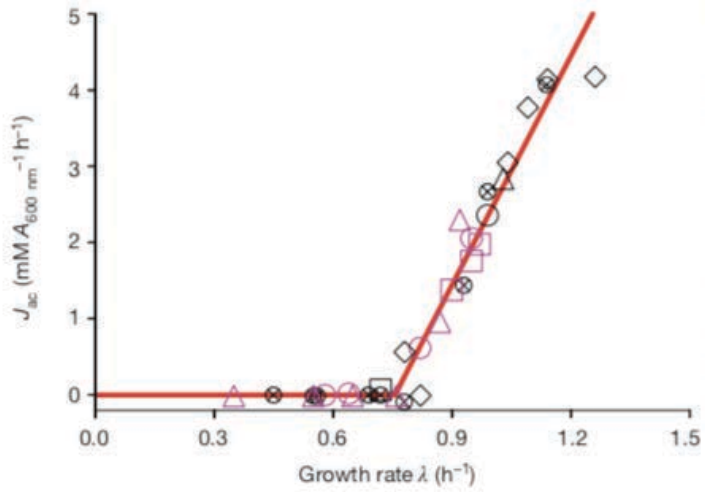
Despite this limitation, their model is able to explain nutritional competition in an elegant and very predictive way. This model could potentially be used to predict an increase in competition in an environment experiencing a nutrient pulse, with an increase in cell density while genetic diversity is maintained, as illustrated in the green square on Figure 4.

When considering cooperation and its selection, no model has shown how this interaction could be established nor how nutrient dynamics would impact it, with the exception of the "Black Queen's hypothesis". If we look at the competition sensing model, we could hypothesize that in low nutrient environments, bacterial density would be lower, thus

decreasing the cell damage sensed by bacteria in the community. However, this still does not explain the cooperation selective process. Nutritional stress adaptations of bacteria limited by a nutrient in their environment might provide some information regarding this.

4.3 Metabolic overflow: a metabolic response of cell under nutrient stress

As defined in the article by Basan et al. (2015), "Overflow metabolism refers to the seemingly wasteful strategy in which cells use fermentation instead of the more efficient respiration to generate energy, despite the availability of oxygen". This metabolic overflow leads to the release of fermentation products, such as intermediates of glucose degradation in the glycolysis or sometimes the Krebbs pathway (e.g. lactate, ethanol, acetate or oxalate) from cells. This has been observed in fast growing eukaryotic and bacterial cells, but an explanation for this type of metabolic regulation has been lacking until recently. Basan et al. (2015) observed that above a certain threshold growth rate, *E.coli* batch cultures started to carry out metabolic overflow by secreting acetate as a byproduct. The acetate secretion per biomass observed above the growth rate threshold was linearly correlated to the growth rate. Basan et al. (2015) defined this phenomenon as the acetate line (*Figure 5*).



Strain	Description	Carbon	Symbol
NCM3722	WT	Glucose	○
NCM3722	WT	Lactose	△
NCM3722	WT	Glycerol	□
NCM3722	WT	Other	⊗
NQ1243	Uptake titration	Glucose	○
NQ381	Uptake titration	Lactose	△
NQ636 NQ638 NQ640	*glpK* mutant	Glycerol	□
NQ3722	+ 7 AA	Various	◇

Figure 5: Figure and legend from Basan et al.(2015). Acetate excretion rate (Jac) is linearly correlated with the growth rate (λ) for wild-type (WT) cells grown in minimal medium with various glycolytic carbon sources (black symbols), and for cells with titratable or mutant uptake systems (purple symbols) (Extended Data Table 1). Black diamonds indicate various carbon sources supplemented with seven non-degradable amino acids (AA). The red line shows the best-fit of all the data to equation in Basan et al. (2015).

This was observed for different carbon sources. Basan et al. (2015) emitted the hypothesis that a metabolic shunt could occur due to the high bioenergetic cost of using proteomes for respiration (lower energy ratio for building the pathway proteome/ energy generated by such proteome) relative to the fermentation pathway. At high carbon uptake and high potential growth rate, cells can obtain the highest growth rate by using the more efficient fermentative pathway. On the other hand, if carbon uptake is low and growth rate is also low, it is more useful to rely on the more carbon efficient respiratory pathway to maximize the carbon flux to support growth (Basan et al. 2015). This is an example of a selective pressure to optimize growth yield and minimizing the protein pool needed to reach it.

Other selective pressures can also lead to metabolic overflow when cells grow under limiting nutrient conditions. The review by Carlson et al. (2018) describes how metabolic shunts can be observed when *E. coli* batch cultures are cultivated under different limiting nutrient conditions in chemostat studies. The metabolic shunts are triggered by culture on low carbon or low iron and lead to secretion of acetate or formate, while under more severe nutritional stress, formate and acetate are predicted to decrease and lactate starts to accumulate in the batch culture medium (Carlson et al. 2018). This time, the selective pressure occurs on the limitation of a key nutrient being used to produce enzymes involved into oxidative metabolisms (e.g. iron) and leads to a shunt selecting the metabolic pathway being less limited by this low nutrient availability (fermentation).

4.3.1 Releasing costless metabolites can lead to cross-feeding cooperation

The secretion of metabolic byproducts could trigger the beginning of cross-feeding interactions among the different members of the community and thus promote an increased cooperation as well as a decrease in competition (Carlson et al. 2018). This was first suggested by Pfeiffer and Bonhoeffer (2004) who showed in their modelling experiments that cross-feeding could arise from a set of energetic and metabolic optimization principles: “the rate of ATP production is maximized, the concentration of enzymes of the pathway is minimized, and the concentration of intermediates of the pathway is minimized”.

In addition to this work, a more recent modelling experiment based on flux balance analysis (FBA) simulations of 24 microbial species co-cultured under various carbon source combinations was recently published (Pacheco, Moel, and Segrè 2019). Their simulations showed that costless metabolites secreted by one of the two co-cultured members could stabilize cross-feeding interactions without being detrimental for the secreting bacteria (the growth rate of the secreting bacteria stayed the same as the growth rate computed without the byproduct secretion).

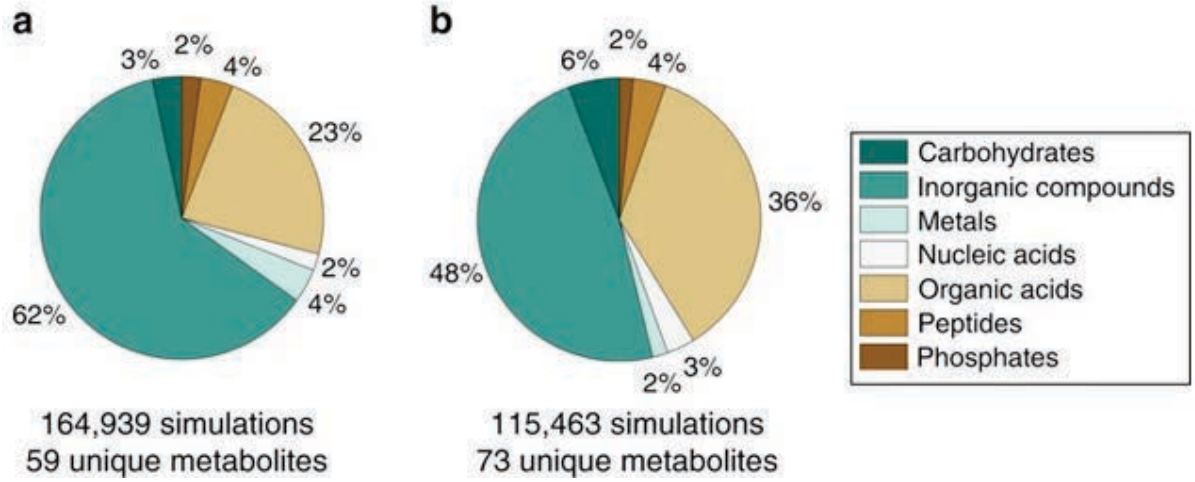


Figure 6:Figure from Pacheco, Moel, and Segrè (2019) summarizing the different sorts of costless metabolites secreted during all their simulations in (a) Oxic co-cultures and (b) Anoxic co-cultures.

In addition, they also identified the main costless metabolites secreted by the strains during co-culture (Figure 6). The dominant metabolite excreted is inorganic (e.g. water, CO_2, …), followed by organic acids, representing more than 20% of the costless metabolites in both oxic as well as anoxic co-culture simulations (Figure 6). We have introduced the possible mechanisms of bacterial competition as well as bacterial cooperation. These interactions were shown to be dynamic. Now we will review the different tools available to study bacterial interactions and then summarize the current research on bacterial interactions and the effect of nutrients dynamics.

5 Studying bacterial interaction in the environments, available tools

5.1 Culture based methods

As we have seen, the vast majority of studies carried out on bacterial interactions are culture-based experiments. Such methods present a bias for investigate microbial interactions at the community level as they are not representative of the natural environment (Figure 7). The key advantages of such systems are that they are easy to follow over time and the reduced complexity in terms of microbial diversity makes it possible to track secreted metabolites to specific strains (Chignell et al. 2018; Herschend et al. 2017) or to visualize physical interactions by microscopy or more complex techniques such as nanoSIM (Musat et al. 2016) .

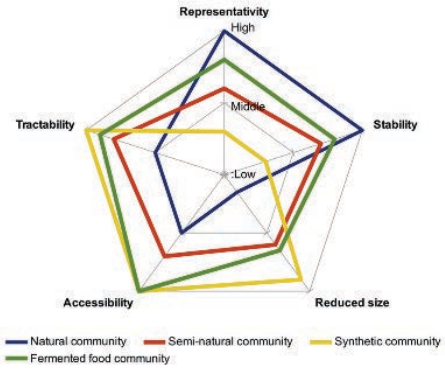


Figure 7: Characterization of the different model communities according to: (1) reduced size, (2) representativity, (3) stability, (4) accessibility and (5) tractability. Figure and legend are from Blasche et al. (2017)

Culture methods include co-cultures where differential growth rates compared to the pure culture can be used to deduce whether the interaction is positive or negative. Such methods can also be used on co-cultures of model organisms where their respective genomes are already known in order to assess how a positive or negative interaction can affect gene expression patterns by generating differential transcriptomic profiles (Hansen et al. 2017; Khan et al. 2018; Molina-Santiago et al. 2017; McClure et al. 2018). Finally, modelling can also be applied in order to predict, based on previous culture experiments and genome analyses, metabolic networks and how an interaction or a perturbation of the system (e.g. increase of nutrient) could affect them (e.g. Zeng and Yang 2019). Such analyses, called Flux Balance Analysis (FBA), have the potential to predict how the metabolic fluxes would be affected by a

perturbation. This analysis relies on the fact that each metabolic reaction is known and can be estimated by equation systems that can be solved. Based on such an approach, Zelezniak et al. (2015), developed a tool to estimate what metabolites could be exchanged and predicted if the interactions between the considered species could be positive or negative. However, this system is currently limited by the number of species that can be computed by the program (<100) which makes it unsuitable for screening complex communities (pers. com. from the author of the tool).

5.2 Culture free methods

The use of culture free methods is recent, but is rapidly developing as it can overcome the representativity bias of the previously described methods. The two main culture free methods used to study bacterial interactions include co-variance/co-occurrence networks and metagenome or metatranscriptome analyses. Both techniques have limitations.

The network approach relies mainly on the assumption that taxa which covary positively across time cooperate and the ones which covary negatively compete (Figure 8). This approach has been used for microbial communities from oceans (Lima-Mendez et al., 2015; Ruan, 2006), soils (Barberán et al., 2012; Ding et al., 2015), human microbiomes (Faust et al., 2012) and heavy-metal-polluted sediments (Yin et al., 2015). These networks often use co-variance to infer positive (cooperative) and negative (competitive) bacterial interactions (e.g. Ruan, 2006), but co-variance might also indicate that the populations are responding to other stimuli simultaneously.

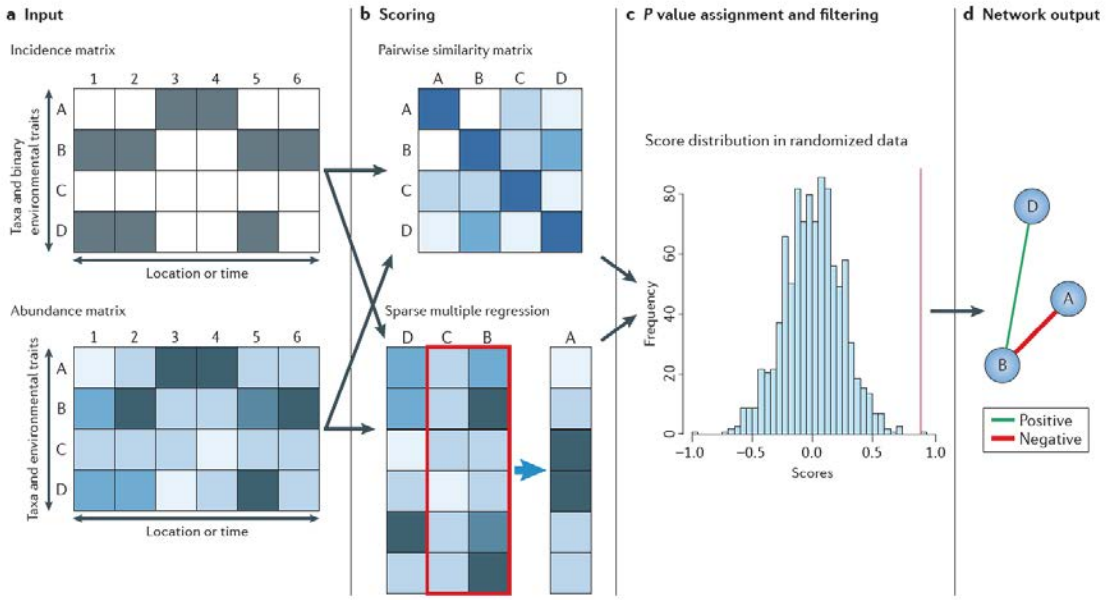


Figure 8: The goal of network inference is to identify combinations of microorganisms that show significant co-presence or mutual exclusion patterns across samples and to combine them into a network. a | Network inference starts from an incidence or an abundance matrix, both of which store observations across different samples, locations or time points. b | Pairwise scores between taxa are then computed using a suitable similarity or distance measure. A range of such measures are used in the literature (for example, Pearson, Spearman, hypergeometric distribution and the Jaccard index). In contrast to similarity-based approaches, multiple regression can detect relationships that involve more than two taxa. To reduce over-fitting, sparse multiple regression is usually carried out — that is, the source taxa subset that best predicts the target taxon's abundance is selected. In addition, the regression model is cross-validated: that is, after regression coefficients have been identified with a training data set, the model's prediction accuracy is quantified on a test data set. c | In the next step, a random score distribution is generated by repeating the scoring step a large number of times (often 1,000 times or more). The random score distribution computes the P value (that is, the probability of obtaining a score by chance that is equal to or better than the observed score) to measure the significance of the predicted relationship. The P value is usually adjusted for multiple testing with procedures such as Bonferroni or Benjamini–Hochberg. d | Taxon pairs with P values below the threshold are visualized as a network, where nodes represent taxa and edges represent the significant relationships between them. The edge thickness can reflect the strength of the relationship. Figure and legend are from (Faust and Raes 2012).

A second strategy to study bacterial interactions is by tracking genes identified as proxies and determine whether they increase or decrease across samples. What is currently limiting is that there is currently no clear consensus on genes that are good proxies for cooperation and competition. If we look at the summary table (Table 3), many of the identified genes have been shown to be implicated in both competition and cooperation. For example, the secretion of exopolysaccharides (EPS) is often seen as a cooperative trait as biofilms have many synergies within them. Nonetheless, Oliveira et al. (2015) observed that EPS secretion could also be triggered by exposure to sublethal antibiotic concentrations, showing a link to competition. Competition was hypothesized to be mediated by antibiotic release in a number of studies (Cornforth and Foster 2013b; Oliveira et al. 2015; Ponce-Soto et al. 2015; Song et al. 2017), but Cordero et al. (2012) observed that antibiotics secretion can also be a public good. This is why it is also important to track the increase in diversity of antibiotic resistance genes across time. If the diversity of genes increases, then we can exclude the hypothesis that antibiotic secretion is a public good, since the community responds to an increase in the number of toxic compounds secreted. This is more compatible with the hypothesis of increased competition among the different members of the bacterial community. A newly identified proxy of cooperation are plasmid backbone genes. Indeed, several articles showed that collaboration could be maintained by genetic exchanges and observed that genes coding

for public goods were preferentially located on mobile elements and hotspots of recombination in bacterial genomes (Dimitriu et al. 2014; 2015; 2016; Nogueira et al. 2009).

Table 3: Table summarizing the different types of genes considered as being clues of cooperation or competition between bacteria. As we can see, most of them don't show a clear separation between competition of cooperation interactions. In addition, when the supporting article was a lab experiment, the experimental design is given by the following code between brackets and specifies if the observation was done on a single species = multiple/single strains culture (MS/SS), a co-culture of two or more species (CC) or an enriched environmental bacterial community (E).

Genes used as surrogate of bacterial interactions	Clue of	Type of study	Articles
Antibiotics	Competition (MS) Cooperation (E)	Lab experiment	(Vasse et al. 2017; Cordero et al. 2012)
Genetic information transfer (plasmids)	Cooperation (MS)	Lab experiment Bioinformatics	(Dimitriu et al. 2014; 2015; 2016; Nogueira et al. 2009)
Type VI secretion system (T6SS)	Competition (MS)	Lab experiment	(Basler, Ho, and Mekalanos 2013; Brunet et al. 2013)
Contact dependent inhibition (CDI)	Competition Cooperation (SS/CC)	Lab experiment Modelling	(Jones, Low, and Hayes 2017; Blanchard, Celik, and Lu 2014)
Exopolysaccharides (EPS)	Competition (MS) Cooperation (CC/E)	Lab experiments	(Oliveira et al. 2015; Frost et al. 2018; Nadell, Drescher, and Foster 2016; Song et al. 2017)
Quorum sensing	Competition (CC) Cooperation (SS)	Lab experiments	(Oshri et al. 2018; Darch et al. 2012; Czárán and Hoekstra 2009; Diggle et al. 2007; Miller and Bassler 2001; Goo et al. 2015)

6 Nutrients and bacterial interactions, a short review

Most of the studies on the impact of nutrients on bacterial interactions have been carried out in lab culture experiments (Mitri and Foster 2013). Co-culture studies give the advantage of being able to define whether the interaction is positive or negative for the cultured species. If the growth rate from the co-cultured species is higher than its growth rate in pure culture, the interaction is termed as positive (cooperation), and if the growth rate is reduced in the co-culture, it's considered negative (competition).

Dynamic changes in nutrient concentrations have been shown to influence bacterial interactions with ramifications for microbial community structure and function (Friedman and Gore, 2017; Khan et al., 2018). In these pure culture studies, either cooperation or competition were the dominant interaction strategy depending on the nutrients considered and their concentrations (Brockhurst et al., 2008, 2010, Lambert et al., 2011, 2014; Ravindran, 2017). Interference competition was hypothesized to be mediated by antibiotic release (Cornforth and Foster, 2013; Oliveira et al., 2015; Ponce-Soto et al., 2015; Song et al., 2017) and was shown to be affected by the nutrient supply (Hol et al., 2014). For example, a sensitive *E.coli* strain co-existed with a colicin-secreting *E.coli* strain when co-cultivated on a poor growth medium (sugars), but not on a rich medium (amino acids and peptides), where the colicin-secreting E.coli strain released antibiotics (Hol et al., 2014). While these studies have provided information on different nutrient effects on bacterial interactions under controlled conditions, they might not predict microbial interactions in the environment.

Microcosm or mesocosm approaches have been used more recently to study microbial communities and the results have varied (Ali et al., 2016; Ponce-Soto et al., 2015; Song et al., 2017). Although no studies on the effect of carbon content on microbial interactions have been published to date, one study measured an increase in antibiotic resistance genes in strains of *Enterococcus faecalis* cultivated in eutrophic sediment mesocosms amended with nitrogen and phosphorus (Ali et al. 2016). Other studies observed a decline in antibiotic resistance in cultivable bacterial populations from an oligotrophic lake in mesocosms amended with nitrogen and phosphorus and from soil bacteria cultivated on agar plate amended with increasing nutrient medium concentrations (Ponce-Soto et al. 2015; Song et al. 2017). The main difference between these two sets of studies is that one used a PCR based method to track antibiotic resistance (Ali et al. 2016), while the others used culture-based methods (Ponce-Soto et al. 2015; Song et al. 2017). Culture based techniques could have a higher bias since they alter the bacterial community by selecting members able to grow on media.

7 Arctic snow: a model habitat for studying bacterial interaction dynamics caused by seasonal changes in nutrient concentrations

Arctic snow could be referred to as an extreme environment. Indeed, temperatures are below 0°C, water availability is low, and during the spring season, UV radiation can be very high at its surface (Maccario et al. 2015). To survive in such an environment, bacteria have developed a range of strategies and adaptations. To survive to the photo-oxidative stresses induced by

high UV irradiation, bacteria can, for example, produce anti-oxidative enzymes able to cross react with the ROS (Reactive Oxygen Species) generated by UV and repair their damaged DNA (Sinha and Häder 2002; Ziegelhoffer and Donohue 2009). Microorganisms are also exposed to variable nutrient concentrations and osmotic stress (Maccario et al. 2015).

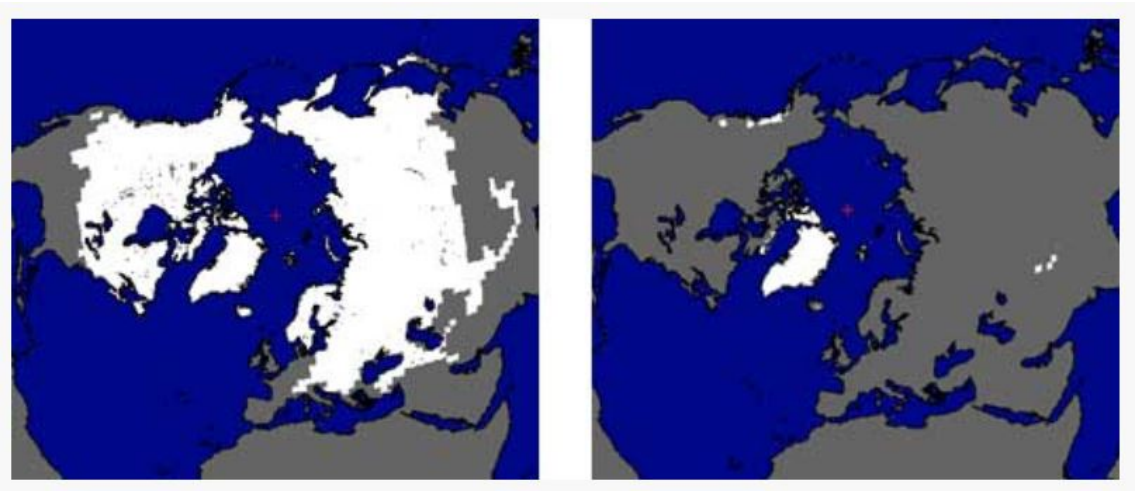


Figure 9: Average snow extent across the Northern Hemisphere reaches its maximum in January (left), and its minimum in August (right). White indicates snow, blue shows oceans and water, and gray indicates land. Figure and legend from the website of the NSIDC (National Snow and Ice Data Center.)

An interesting feature of this environment is that, during the spring season, a rapid increase in nutrients can be observed in the snow and the environment can become quite rich as compared to winter snow. This environment is thus dynamic and displays a range of environmental variations. This property is interesting since the snow can be used as a model environment to compare bacterial interactions in an environment poor in nutrients and the same environment enriched in nutrients during the end of the spring season.

Research on microbial communities of the Arctic snow have shown that there was a dynamic community capable of responding to environmental changes with the potential to carry out a vast range of metabolic activities (Hell et al. 2013; Catherine Larose, Dommergue, and Vogel 2013; Catherine Larose et al. 2013; Catherine Larose et al. 2010; Maccario, Vogel, and Larose 2014). However, none of those studies measured the activity of the tracked bacterial community. Some controversy remains as to whether there is a sustainable microbial community in the snow, given the low bacterial density (10^2-10^5 cells) compared to other cryospheric environments (Boetius et al. 2015), but more recent studies suggest that there are seasonal changes in the community and that these changes are driven by environmental factors (Catherine Larose et al. 2013; Maccario, Vogel, and Larose 2014; Lutz, Anesio, Raiswell, et al. 2016).

Bacterial communities are predominant in the snow and it has been shown that they can impact geochemical cycles in the Arctic and Antarctica (Catherine Larose et al. 2013; Antony et al. 2016). It has been shown that some bacterial communities in other environments have the potential to couple biogeochemical cycles via metabolic coupling or through bacterial interactions (e.g. syntrophy) (Beal, House, and Orphan 2009; Thamdrup et al. 1993). For the moment, nothing is known about the potential interactions of biogeochemical cycles in the snow and if they are coupled through bacterial interactions. Studying snow microbial

communities and how they interact could help understand the impact of arctic snow environments on global geochemical cycles.

Microbial snow communities have been characterized in several studies and reviewed in Boetius et al. (2015) and Maccario et al. (2015). The snow environments generally dominated by Proteobacteria (alpha- and beta- proteobacteria being present in most of the published communities) and also Bacteroidetes (Boetius et al. 2015; Maccario et al. 2015). This is unsurprising if we consider that Arctic snow environments are affected by large seasonal variations in term of light exposition (affecting primary productivity) and temperature fluctuations. Betaproteobacteria are considered r-strategists, able to exploit a wide variety of nutritional sources, and are thus likely to survive in these very dynamic (Hell et al. 2013). Hell et al. (2013) were unable to determine spatial variations in microbial communities during their study on an Arctic melting snowpack, but Maccario, Vogel, and Larose (2014) detected depth variation in addition to seasonal variation, showing that snow could be a highly stratified environment. The observed spatial variability was linked to distinct environmental conditions such as UV exposition for example. Lutz et al. (2016) also observed that bacterial community diversity could be explained by geography and biochemical properties of the snow. However, the vast majority of these studies have focused on spring or summer snow (Maccario et al. 2015)and little is known about bacterial communities and their metabolic and biogeochemical activity in the winter. Hamilton et al. (2013) recently observed subglacial bacterial activity and showed that the amount of phylogenetically related species in the subglacial environment was lower than in the ice. This was attributed to the isolation of the communities from atmospheric nutrient inputs, leading to an oligotrophic environment (Hamilton et al. 2013). In their discussion, they hypothesized that these highly diversified (phylogenetically) subglacial communities carried out mutualistic interactions to increase resources (nutrients) availability. If we make a parallel with the snow, we would expect snow to be more oligotrophic in the winter than in the spring (since there is no photosynthesis).

Although microbial interactions have never been studied in the snow, Gokul et al. (2016) used a network approach to study bacterial communities in cryoconite. Core taxa (key species) were predominantly Actinobacteria (related to isolates from soil humus) and interacted positively with the rest of the bacterial community. The authors also showed that these core OTUs had a stronger influence on the community structure than the studied environmental conditions (Gokul et al. 2016). This finding could be an illustration of the concept of Konopka et al. (2015), that states that the bacterial community can be modified by interactions that will impose a dynamic change even in the absence of any modification in the environmental conditions.

8 Hypotheses of the work

Arctic snow microbial communities were selected because arctic snow carbon content varies by several orders of magnitude during the spring season (Twickler et al. 1986) and is generally considered a low carbon environment. Recently, using COG functions characteristic of oligotrophy or copiotrophy as proposed by Lauro et al. (2009), Maccario et al. (2019) showed that arctic snow bacterial communities were adapted to oligotrophic lifestyles. Although oligotrophic, carbon content in the snow increases over the spring season (Grannas et al. 2007; Haan et al. 2001; Twickler et al. 1986). In addition, Arctic snow has varying nutrient conditions that affect bacterial community structure and function (Larose et al. 2013).

Based on the previous theories outlined in this introduction, the overarching hypothesis of this PhD is that an increase in organic acids in the snow would increase competition and reduce collaboration. This hypothesis is supported by the observation that an increase in carbon increased interference competition (F. J. Hol et al. 2014). In opposition, cooperation could be higher in more limited nutrient environments, as shown by Benomar et al. (2015), who observed that metabolite exchanges could be triggered by nutritional stresses.

In Chapter 2, I used a multidisciplinary approach combining chemistry, metagenomics, metatranscriptomics and network analysis to identify shifts in microbial interactions in Arctic snowpacks during the spring. Snowpack communities were investigated over two months to assess changes in community structure, activity and function and relate these to shifts in organic acid concentrations.

Given the challenges related to analysing samples with low sequencing depths, I needed to develop new tools to improve annotation and analysis. In Chapter 3, I present a new bioinformic pipeline called EggVio that was validated using an existing dataset available in the laboratory.

In Chapter 4, I carried out a microcosm study to validate the experimental results obtained from field data. Arctic snow collected in Svalbard was incubated at -5°C and nutrients were added to test whether organic acids could shift the interactions of microbial communities. I applied the pipeline described in Chapter 3 and showed that organic acid concentrations modify interactions.

Based on our results, we suggest that the seasonal increase or organic acids in the arctic snow is caused by specific members of the endogenous microbial community (mainly Fungi). In addition, this increase of organic acids in the microbial community causes a shift in microbial interactions. Bacterial competition is the main interaction affected by organic acids, as we showed that it could increase significantly (four-fold increase) when we amended arctic snow with acetate in our microcosms. In contrast, we suggest that an increase in organic acids had little or no effect on bacterial cooperation.

9 Bibliography

Antony, Runa, Aritri Sanyal, Neelam Kapse, Prashant K. Dhakephalkar, Meloth Thamban, and Shanta Nair. 2016. “Microbial Communities Associated with Antarctic Snow Pack and Their Biogeochemical Implications.” *Microbiological Research* 192 (November): 192–202. https://doi.org/10.1016/j.micres.2016.07.004.

Basan, Markus, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R. Williamson, and Terence Hwa. 2015. “Overflow Metabolism in *Escherichia Coli* Results from Efficient Proteome Allocation.” *Nature* 528 (7580): 99–104. https://doi.org/10.1038/nature15765.

Basler, Marek, Brian T. Ho, and John J. Mekalanos. 2013. “Tit-for-Tat: Type VI Secretion System Counterattack during Bacterial Cell-Cell Interactions.” *Cell* 152 (4): 884–94. https://doi.org/10.1016/j.cell.2013.01.042.

Beal, Emily J., Christopher H. House, and Victoria J. Orphan. 2009. “Manganese- and Iron-Dependent Marine Methane Oxidation.” *Science* 325 (5937): 184–87. https://doi.org/10.1126/science.1169984.

Benomar, Saida, David Ranava, María Luz Cárdenas, Eric Trably, Yan Rafrafi, Adrien Ducret, Jérôme Hamelin, Elisabeth Lojou, Jean-Philippe Steyer, and Marie-Thérèse Giudici-Orticoni. 2015. “Nutritional Stress Induces Exchange of Cell Material and Energetic Coupling between Bacterial Species.” *Nature Communications* 6 (February): 6283. https://doi.org/10.1038/ncomms7283.

Blanchard, Andrew E., Venhar Celik, and Ting Lu. 2014. “Extinction, Coexistence, and Localized Patterns of a Bacterial Population with Contact-Dependent Inhibition.” *BMC Systems Biology* 8 (February): 23. https://doi.org/10.1186/1752-0509-8-23.

Blasche, Sonja, Yongkyu Kim, Ana Paula Oliveira, and Kiran R. Patil. 2017. “Model Microbial Communities for Ecosystems Biology.” *Current Opinion in Systems Biology*, Systems biology of model organisms, 6 (December): 51–57. https://doi.org/10.1016/j.coisb.2017.09.002.

Boetius, Antje, Alexandre M. Anesio, Jody W. Deming, Jill A. Mikucki, and Josephine Z. Rapp. 2015. “Microbial Ecology of the Cryosphere: Sea Ice and Glacial Habitats.” *Nat Rev Micro* 13 (11): 677–90.

Brockhurst, Michael A., Michelle G. J. L. Habets, Ben Libberton, Angus Buckling, and Andy Gardner. 2010. “Ecological Drivers of the Evolution of Public-Goods Cooperation in Bacteria.” *Ecology* 91 (2): 334–40. https://doi.org/10.1890/09-0293.1.

Brunet, Yannick R., Leon Espinosa, Seddik Harchouni, Tâm Mignot, and Eric Cascales. 2013. “Imaging Type VI Secretion-Mediated Bacterial Killing.” *Cell Reports* 3 (1): 36–41. https://doi.org/10.1016/j.celrep.2012.11.027.

Burgin, Amy J., Wendy H. Yang, Stephen K. Hamilton, and Whendee L. Silver. 2011. “Beyond Carbon and Nitrogen: How the Microbial Energy Economy Couples Elemental Cycles in Diverse Ecosystems.” *Frontiers in Ecology and the Environment* 9 (1): 44–52. https://doi.org/10.1890/090227.

Carlson, Ross P., Ashley E. Beck, Poonam Phalak, Matthew W. Fields, Tomas Gedeon, Luke Hanley, William R. Harcombe, Michael A. Henson, and Jeffrey J. Heys. 2018. “Competitive Resource Allocation to Metabolic Pathways Contributes to Overflow Metabolisms and Emergent Properties in Cross-Feeding Microbial Consortia.” *Biochemical Society Transactions* 46 (2): 269–84. https://doi.org/10.1042/BST20170242.

Chignell, J. F., S. Park, C. M. R. Lacerda, S. K. De Long, and K. F. Reardon. 2018. “Label-Free Proteomics of a Defined, Binary Co-Culture Reveals Diversity of Competitive Responses Between Members of a Model Soil Microbial System.” *Microbial Ecology* 75 (3): 701–19. https://doi.org/10.1007/s00248-017-1072-1.

Cordero, Otto X., Hans Wildschutte, Benjamin Kirkup, Sarah Proehl, Lynn Ngo, Fatima Hussain, Frederique Le Roux, Tracy Mincer, and Martin F. Polz. 2012. “Ecological Populations of Bacteria Act as Socially Cohesive Units of Antibiotic Production and Resistance.” *Science* 337 (6099): 1228–31. https://doi.org/10.1126/science.1219385.

Cornforth, Daniel M., and Kevin R. Foster. 2013a. “Competition Sensing: The Social Side of Bacterial Stress Responses.” *Nature Reviews. Microbiology* 11 (4): 285–93. https://doi.org/10.1038/nrmicro2977.

———. 2013b. “Competition Sensing: The Social Side of Bacterial Stress Responses.” *Nature Reviews. Microbiology* 11 (4): 285–93. https://doi.org/10.1038/nrmicro2977.

Czárán, Tamás, and Rolf F. Hoekstra. 2009. “Microbial Communication, Cooperation and Cheating: Quorum Sensing Drives the Evolution of Cooperation in Bacteria.” *PLOS ONE* 4 (8): e6655. https://doi.org/10.1371/journal.pone.0006655.

Darch, Sophie E., Stuart A. West, Klaus Winzer, and Stephen P. Diggle. 2012. “Density-Dependent Fitness Benefits in Quorum-Sensing Bacterial Populations.” *Proceedings of the National Academy of Sciences* 109 (21): 8259–63. https://doi.org/10.1073/pnas.1118131109.

Darwin, Charles. 1859. *On The Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Deng, Yi-Jie, and Shiao Y. Wang. 2016. “Synergistic Growth in Bacteria Depends on Substrate Complexity.” *Journal of Microbiology* 54 (1): 23–30. https://doi.org/10.1007/s12275-016-5461-9.

Diggle, Stephen P., Ashleigh S. Griffin, Genevieve S. Campbell, and Stuart A. West. 2007. “Cooperation and Conflict in Quorum-Sensing Bacterial Populations.” *Nature* 450 (7168): 411–14. https://doi.org/10.1038/nature06279.

Dimitriu, Tatiana, Chantal Lotton, Julien Bénard-Capelle, Dusan Misevic, Sam P. Brown, Ariel B. Lindner, and François Taddei. 2014. “Genetic Information Transfer Promotes Cooperation in Bacteria.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (30): 11103–8. https://doi.org/10.1073/pnas.1406840111.

Dimitriu, Tatiana, Dusan Misevic, Ariel B Lindner, and François Taddei. 2015. “Mobile Genetic Elements Are Involved in Bacterial Sociality.” *Mobile Genetic Elements* 5 (1): 7–11. https://doi.org/10.1080/2159256X.2015.1006110.

Dimitriu, Tatiana, Dusan Misevic, Chantal Lotton, Sam P. Brown, Ariel B. Lindner, and François Taddei. 2016. “Indirect Fitness Benefits Enable the Spread of Host Genes Promoting Costly Transfer of Beneficial Plasmids.” *PLOS Biology* 14 (6): e1002478. https://doi.org/10.1371/journal.pbio.1002478.

D’Onofrio, Anthony, Jason M. Crawford, Eric J. Stewart, Kathrin Witt, Ekaterina Gavrish, Slava Epstein, Jon Clardy, and Kim Lewis. 2010. “Siderophores from Neighboring Organisms Promote the Growth of Uncultured Bacteria.” *Chemistry & Biology* 17 (3): 254–64. https://doi.org/10.1016/j.chembiol.2010.02.010.

D’Souza, Glen, and Christian Kost. 2016. “Experimental Evolution of Metabolic Dependency in Bacteria.” *PLOS Genetics* 12 (11): e1006364. https://doi.org/10.1371/journal.pgen.1006364.

Faust, Karoline, and Jeroen Raes. 2012. “Microbial Interactions: From Networks to Models.” *Nature Reviews Microbiology* 10 (8): 538–50. https://doi.org/10.1038/nrmicro2832.

Fierer, Noah, Mark A. Bradford, and Robert B. Jackson. 2007. “Toward an Ecological Classification of Soil Bacteria.” *Ecology* 88 (6): 1354–64.

Finkel, Omri M., Isai Salas-González, Gabriel Castrillo, Theresa F. Law, Jonathan M. Conway, Corbin D. Jones, and Jeffery L. Dangl. 2019. “Root Development Is Maintained by Specific Bacteria-Bacteria Interactions within a Complex Microbiome.” *BioRxiv*, May, 645655. https://doi.org/10.1101/645655.

Foster, Kevin R., and Thomas Bell. 2012. “Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species.” *Current Biology* 22 (19): 1845–50. https://doi.org/10.1016/j.cub.2012.08.005.

Frost, Isabel, William P. J. Smith, Sara Mitri, Alvaro San Millan, Yohan Davit, James M. Osborne, Joe M. Pitt-Francis, R. Craig MacLean, and Kevin R. Foster. 2018. “Cooperation, Competition and Antibiotic Resistance in Bacterial Colonies.” *The ISME Journal* 12 (6): 1582–93. https://doi.org/10.1038/s41396-018-0090-4.

Gillott, Cedric. 1995. “The Biotic Environment.” In *Entomology*, edited by Cedric Gillott, 659–90. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-4380-8_23.

Gokul, Jarishma K., Andrew J. Hodson, Eli R. Saetnan, Tristram D. L. Irvine-Fynn, Philippa J. Westall, Andrew P. Detheridge, Nozomu Takeuchi, Jennifer Bussell, Luis A. J. Mur, and Arwyn Edwards. 2016. “Taxon Interactions Control the Distributions of Cryoconite Bacteria Colonizing a High Arctic Ice Cap.” *Molecular Ecology* 25 (15): 3752–67. https://doi.org/10.1111/mec.13715.

Goo, Eunhye, Jae Hyung An, Yongsung Kang, and Ingyu Hwang. 2015. “Control of Bacterial Metabolism by Quorum Sensing.” *Trends in Microbiology* 23 (9): 567–76. https://doi.org/10.1016/j.tim.2015.05.007.

Grannas, A. M., A. E. Jones, J. Dibb, M. Ammann, C. Anastasio, H. J. Beine, M. Bergin, et al. 2007. “An Overview of Snow Photochemistry: Evidence, Mechanisms and Impacts.” *Atmos. Chem. Phys.* 7 (16): 4329–73. https://doi.org/10.5194/acp-7-4329-2007.

Haan, D., Y. Zuo, V. Gros, and C. a. M. Brenninkmeijer. 2001. “Photochemical Production of Carbon Monoxide in Snow.” *Journal of Atmospheric Chemistry* 40 (3): 217–30. https://doi.org/10.1023/A:1012216112683.

Hamilton, Trinity L., John W. Peters, Mark L. Skidmore, and Eric S. Boyd. 2013. “Molecular Evidence for an Active Endogenous Microbiome beneath Glacial Ice.” *The ISME Journal* 7 (7): 1402–12. https://doi.org/10.1038/ismej.2013.31.

Hansen, Lea Benedicte Skov, Dawei Ren, Mette Burmølle, and Søren J. Sørensen. 2017. “Distinct Gene Expression Profile of *Xanthomonas Retroflexus* Engaged in Synergistic Multispecies Biofilm Formation.” *The ISME Journal* 11 (1): 300–303. https://doi.org/10.1038/ismej.2016.107.

Harrison, Freya, and Angus Buckling. 2005. “Hypermutability Impedes Cooperation in Pathogenic Bacteria.” *Current Biology* 15 (21): 1968–71. https://doi.org/10.1016/j.cub.2005.09.048.

Hell, Katherina, Arwyn Edwards, Jakub Zarsky, Sabine M. Podmirseg, Susan Girdwood, Justin A. Pachebat, Heribert Insam, and Birgit Sattler. 2013. “The Dynamic Bacterial Communities of a Melting High Arctic Glacier Snowpack.” *The ISME Journal* 7 (9): 1814–26. https://doi.org/10.1038/ismej.2013.51.

Herschend, Jakob, Zacharias B. V. Damholt, Andrea M. Marquard, Birte Svensson, Søren J. Sørensen, Per Hägglund, and Mette Burmølle. 2017. “A Meta-Proteomics Approach to Study the Interspecies Interactions Affecting Microbial Biofilm Development in a Model Community.” *Scientific Reports* 7 (1): 16483. https://doi.org/10.1038/s41598-017-16633-6.

Ho, Adrian, Roey Angel, Annelies J. Veraart, Anne Daebeler, Zhongjun Jia, Sang Yoon Kim, Frederiek-Maarten Kerckhof, Nico Boon, and Paul L. E. Bodelier. 2016. “Biotic Interactions in Microbial Communities as Modulators of Biogeochemical Processes: Methanotrophy as a Model System.” *Frontiers in Microbiology* 7 (August). https://doi.org/10.3389/fmicb.2016.01285.

Ho, Adrian, Di Lonardo, D. Paolo, and Paul L. E. Bodelier. 2017. “Revisiting Life Strategy Concepts in Environmental Microbial Ecology.” *FEMS Microbiology Ecology* 93 (3). https://doi.org/10.1093/femsec/fix006.

Hol, Felix J. H., Peter Galajda, Krisztina Nagy, Rutger G. Woolthuis, Cees Dekker, and Juan E. Keymer. 2013. “Spatial Structure Facilitates Cooperation in a Social Dilemma: Empirical Evidence from a Bacterial Community.” *PloS One* 8 (10): e77042. https://doi.org/10.1371/journal.pone.0077042.

Hol, Felix J. H., Peter Galajda, Rutger G. Woolthuis, Cees Dekker, and Juan E. Keymer. 2015. “The Idiosyncrasy of Spatial Structure in Bacterial Competition.” *BMC Research Notes* 8 (June): 245. https://doi.org/10.1186/s13104-015-1169-x.

Hol, Felix JH, Mathias J. Voges, Cees Dekker, and Juan E. Keymer. 2014. “Nutrient-Responsive Regulation Determines Biodiversity in a Colicin-Mediated Bacterial Community.” *BMC Biology* 12 (August): 68. https://doi.org/10.1186/s12915-014-0068-2.

John, Martina, Antoine Prandota Trzcinski, Yan Zhou, and Wun Jern Ng. 2017. “Microbial Stress Mediated Intercellular Nanotubes in an Anaerobic Microbial Consortium Digesting Cellulose.” *Scientific Reports* 7 (1): 18006. https://doi.org/10.1038/s41598-017-18198-w.

Jones, Allison M., David A. Low, and Christopher S. Hayes. 2017. “Can’t You Hear Me Knocking: Contact-Dependent Competition and Cooperation in Bacteria.” *Emerging Topics in Life Sciences* 1 (1): 75–83. https://doi.org/10.1042/ETLS20160019.

Khan, Nymul, Yukari Maezato, Ryan S. McClure, Colin J. Brislawn, Jennifer M. Mobberley, Nancy Isern, William B. Chrisler, et al. 2018. “Phenotypic Responses to Interspecies Competition and Commensalism in a Naturally-Derived Microbial Co-Culture.” *Scientific Reports* 8 (1): 297. https://doi.org/10.1038/s41598-017-18630-1.

Klappenbach, Joel A., John M. Dunbar, and Thomas M. Schmidt. 2000. “RRNA Operon Copy Number Reflects Ecological Strategies of Bacteria.” *Applied and Environmental Microbiology* 66 (4): 1328–33. https://doi.org/10.1128/AEM.66.4.1328-1333.2000.

Konopka, Allan, Stephen Lindemann, and Jim Fredrickson. 2015. “Dynamics in Microbial Communities: Unraveling Mechanisms to Identify Principles.” *ISME J* 9 (7): 1488–95.

Kümmerli, Rolf, Ashleigh S. Griffin, Stuart A. West, Angus Buckling, and Freya Harrison. 2009. “Viscous Medium Promotes Cooperation in the Pathogenic Bacterium Pseudomonas Aeruginosa.” *Proceedings of the Royal Society of London B: Biological Sciences* 276 (1672): 3531–38. https://doi.org/10.1098/rspb.2009.0861.

Lambert, Guillaume, David Liao, Saurabh Vyawahare, and Robert H. Austin. 2011. “Anomalous Spatial Redistribution of Competing Bacteria under Starvation

Conditions ▿ .” *Journal of Bacteriology* 193 (8): 1878–83. https://doi.org/10.1128/JB.01430-10.
Lambert, Guillaume, Saurabh Vyawahare, and Robert H. Austin. 2014. “Bacteria and Game Theory: The Rise and Fall of Cooperation in Spatially Heterogeneous Environments.” *Interface Focus* 4 (4). https://doi.org/10.1098/rsfs.2014.0029.
Larose, Catherine, Aurélien Dommergue, and Timothy M. Vogel. 2013. “Microbial Nitrogen Cycling in Arctic Snowpacks.” *Environmental Research Letters* 8 (3): 5004. https://doi.org/10.1088/1748-9326/8/3/035004.
Larose, Catherine, Emmanuel Prestat, Sébastien Cecillon, Sibel Berger, Cédric Malandain, Delina Lyon, Christophe Ferrari, Dominique Schneider, Aurélien Dommergue, and Timothy M. Vogel. 2013. “Interactions between Snow Chemistry, Mercury Inputs and Microbial Population Dynamics in an Arctic Snowpack.” *PLOS ONE* 8 (11): e79972. https://doi.org/10.1371/journal.pone.0079972.
Larose, Catheririe, Aurelien Dommergue, Martine De Angelis, Daniel Cossa, Bernard Averty, Nicolas Marusczak, Nicolas Soumis, Dominique Schneider, and Christophe Ferrari. 2010. “Springtime Changes in Snow Chemistry Lead to New Insights into Mercury Methylation in the Arctic.” *Geochimica Et Cosmochimica Acta* 74 (22): 6263–75. https://doi.org/10.1016/j.gca.2010.08.043.
Lauro, Federico M., Diane McDougald, Torsten Thomas, Timothy J. Williams, Suhelen Egan, Scott Rice, Matthew Z. DeMaere, et al. 2009. “The Genomic Basis of Trophic Strategy in Marine Bacteria.” *Proceedings of the National Academy of Sciences* 106 (37): 15527–33. https://doi.org/10.1073/pnas.0903507106.
Lidicker, William Z. 1979. “A Clarification of Interactions in Ecological Systems.” *BioScience* 29 (8): 475–77. https://doi.org/10.2307/1307540.
Lutz, Stefanie, Alexandre M. Anesio, Arwyn Edwards, and Liane G. Benning. 2016. “Linking Microbial Diversity and Functionality of Arctic Glacial Surface Habitats.” *Environmental Microbiology*, n/a-n/a. https://doi.org/10.1111/1462-2920.13494.
Lutz, Stefanie, Alexandre M. Anesio, Rob Raiswell, Arwyn Edwards, Rob J. Newton, Fiona Gill, and Liane G. Benning. 2016. “The Biogeography of Red Snow Microbiomes and Their Role in Melting Arctic Glaciers.” *Nature Communications* 7: 11968.
Maccario, Lorrie, Shelly D. Carpenter, Jody W. Deming, Timothy M. Vogel, and Catherine Larose. 2019. “Sources and Selection of Snow-Specific Microbial Communities in a Greenlandic Sea Ice Snow Cover.” *Scientific Reports* 9 (1): 2290. https://doi.org/10.1038/s41598-019-38744-y.
Maccario, Lorrie, Laura Sanguino, Timothy M. Vogel, and Catherine Larose. 2015. “Snow and Ice Ecosystems: Not so Extreme.” *Research in Microbiology*, Special issue on Microbial diversity, adaptation and evolution, 166 (10): 782–95. https://doi.org/10.1016/j.resmic.2015.09.002.
Maccario, Lorrie, Timothy M. Vogel, and Catherine Larose. 2014. “Potential Drivers of Microbial Community Structure and Function in Arctic Spring Snow.” *Frontiers in Microbiology* 5 (August). https://doi.org/10.3389/fmicb.2014.00413.
Mas, Alix, Shahrad Jamshidi, Yvan Lagadeuc, Damien Eveillard, and Philippe Vandenkoornhuyse. 2016. “Beyond the Black Queen Hypothesis.” *ISME Journal* 10 (9): 2085–91. https://doi.org/10.1038/ismej.2016.22.
Mc Ginty, Sorcha E., Daniel J. Rankin, and Sam P. Brown. 2011. “Horizontal Gene Transfer and the Evolution of Bacterial Cooperation.” *Evolution* 65 (1): 21–32. https://doi.org/10.1111/j.1558-5646.2010.01121.x.

McClure, Ryan S., Christopher C. Overall, Eric A. Hill, Hyun-Seob Song, Moiz Charania, Hans C. Bernstein, Jason E. McDermott, and Alexander S. Beliaev. 2018. “Species-Specific Transcriptomic Network Inference of Interspecies Interactions.” *The ISME Journal* 12 (8): 2011. https://doi.org/10.1038/s41396-018-0145-6.

Mikucki, Jill A., Ann Pearson, David T. Johnston, Alexandra V. Turchyn, James Farquhar, Daniel P. Schrag, Ariel D. Anbar, John C. Priscu, and Peter A. Lee. 2009. “A Contemporary Microbially Maintained Subglacial Ferrous ‘Ocean.’” *Science* 324 (5925): 397–400. https://doi.org/10.1126/science.1167350.

Miller, Melissa B., and Bonnie L. Bassler. 2001. “QUORUM SENSING IN BACTERIA.” *Annual Review of Microbiology* 55 (1): 165–99. https://doi.org/10.1146/annurev.micro.55.1.165.

Mitri, Sara, and Kevin Richard Foster. 2013. “The Genotypic View of Social Interactions in Microbial Communities.” *Annual Review of Genetics* 47 (1): 247–73. https://doi.org/10.1146/annurev-genet-111212-133307.

Molina-Santiago, Carlos, Zulema Udaondo, Baldo F. Cordero, and Juan L. Ramos. 2017. “Interspecies Cross-Talk between Co-Cultured Pseudomonas Putida and Escherichia Coli.” *Environmental Microbiology Reports* 9 (4): 441–48. https://doi.org/10.1111/1758-2229.12553.

Morris, J. Jeffrey. 2015. “Black Queen Evolution: The Role of Leakiness in Structuring Microbial Communities.” *Trends in Genetics* 31 (8): 475–82. https://doi.org/10.1016/j.tig.2015.05.004.

Musat, Niculina, Florin Musat, Peter Kilian Weber, and Jennifer Pett-Ridge. 2016. “Tracking Microbial Interactions with NanoSIMS.” *Current Opinion in Biotechnology*, Analytical biotechnology, 41 (October): 114–21. https://doi.org/10.1016/j.copbio.2016.06.007.

Nadell, Carey D., Knut Drescher, and Kevin R. Foster. 2016. “Spatial Structure, Cooperation and Competition in Biofilms.” *Nat Rev Micro* 14 (9): 589–600.

Nogueira, Teresa, Daniel J. Rankin, Marie Touchon, François Taddei, Sam P. Brown, and Eduardo P. C. Rocha. 2009. “Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence.” *Current Biology* 19 (20): 1683–91. https://doi.org/10.1016/j.cub.2009.08.056.

Oliveira, Nuno M., Esteban Martinez-Garcia, Joao Xavier, William M. Durham, Roberto Kolter, Wook Kim, and Kevin R. Foster. 2015. “Biofilm Formation As a Response to Ecological Competition.” *PLOS Biology* 13 (7): e1002191. https://doi.org/10.1371/journal.pbio.1002191.

Orr, H. Allen. 2009. “Fitness and Its Role in Evolutionary Genetics.” *Nature Reviews. Genetics* 10 (8): 531–39. https://doi.org/10.1038/nrg2603.

Oshri, Ron D., Keren S. Zrihen, Itzhak Shner, Shira Omer Bendori, and Avigdor Eldar. 2018. “Selection for Increased Quorum-Sensing Cooperation in Pseudomonas Aeruginosa through the Shut-down of a Drug Resistance Pump.” *The ISME Journal* 12 (10): 2458. https://doi.org/10.1038/s41396-018-0205-y.

Overbeek, Leonard S. van, and Kari Saikkonen. 2016. “Impact of Bacterial–Fungal Interactions on the Colonization of the Endosphere.” *Trends in Plant Science*, Special Issue: Unravelling the Secrets of the Rhizosphere, 21 (3): 230–42. https://doi.org/10.1016/j.tplants.2016.01.003.

Pacheco, Alan R., Mauricio Moel, and Daniel Segrè. 2019. “Costless Metabolic Secretions as Drivers of Interspecies Interactions in Microbial Ecosystems.” *Nature Communications* 10 (1): 1–12. https://doi.org/10.1038/s41467-018-07946-9.

Pande, Samay, Shraddha Shitut, Lisa Freund, Martin Westermann, Felix Bertels, Claudia Colesie, Ilka B. Bischofs, and Christian Kost. 2015. “Metabolic Cross-Feeding via Intercellular Nanotubes among Bacteria.” *Nature Communications* 6 (February): 6238. https://doi.org/10.1038/ncomms7238.

Pfeiffer, Thomas, and Sebastian Bonhoeffer. 2004. “Evolution of Cross-Feeding in Microbial Populations.” *The American Naturalist* 163 (6): E126–35. https://doi.org/10.1086/383593.

Ponce-Soto, Gabriel Y., Eneas Aguirre-von-Wobeser, Luis E. Eguiarte, James J. Elser, Zarraz M.-P. Lee, and Valeria Souza. 2015. “Enrichment Experiment Changes Microbial Interactions in an Ultra-Oligotrophic Environment.” *Frontiers in Microbiology* 6: 246. https://doi.org/10.3389/fmicb.2015.00246.

Ravindran, Sandeep. 2017. “Inner Workings: Bacteria Work Together to Survive Earth’s Depths.” *Proceedings of the National Academy of Sciences* 114 (5): 788–90. https://doi.org/10.1073/pnas.1621079114.

Ren, Dawei, Jonas S Madsen, Soren J Sorensen, and Mette Burmolle. 2015. “High Prevalence of Biofilm Synergy among Bacterial Soil Isolates in Cocultures Indicates Bacterial Interspecific Cooperation.” *ISME J* 9 (1): 81–89.

Rønn, Regin, Mette Vestergård, and Flemming Ekelund. 2015. “Interactions Between Bacteria, Protozoa and Nematodes in Soil.” *Acta Protozoologica* 51 (3): 223–35.

Ross-Gillespie, Adin, Andy Gardner, Angus Buckling, Stuart A. West, and Ashleigh S. Griffin. 2009. “Density Dependence and Cooperation: Theory and a Test with Bacteria.” *Evolution; International Journal of Organic Evolution* 63 (9): 2315–25. https://doi.org/10.1111/j.1558-5646.2009.00723.x.

Ross-Gillespie, Adin, Andy Gardner, Stuart A. West, and Ashleigh S. Griffin. 2007. “Frequency Dependence and Cooperation: Theory and a Test with Bacteria.” *The American Naturalist* 170 (3): 331–42. https://doi.org/10.1086/519860.

Schlesinger, William H, Jonathan J Cole, Adrien C Finzi, and Elisabeth A Holland. 2011. “Introduction to Coupled Biogeochemical Cycles.” *Frontiers in Ecology and the Environment* 9 (1): 5–8. https://doi.org/10.1890/090235.

Seth, Erica C., and Michiko E. Taga. 2014. “Nutrient Cross-Feeding in the Microbial World.” *Frontiers in Microbiology* 5. https://doi.org/10.3389/fmicb.2014.00350.

Sinha, Rajeshwar P., and Donat-P. Häder. 2002. “UV-Induced DNA Damage and Repair: A Review.” *Photochemical & Photobiological Sciences* 1 (4): 225–36. https://doi.org/10.1039/B201230H.

Song, Ho-Kyung, Woojin Song, Mincheol Kim, Binu M. Tripathi, Hyoki Kim, Piotr Jablonski, and Jonathan M. Adams. 2017. “Bacterial Strategies along Nutrient and Time Gradients, Revealed by Metagenomic Analysis of Laboratory Microcosms.” *FEMS Microbiology Ecology* 93 (10). https://doi.org/10.1093/femsec/fix114.

(Stuart) Barker, James S.F. 2009. “Defining Fitness in Natural and Domesticated Populations.” In *Adaptation and Fitness in Animal Populations: Evolutionary and Breeding Perspectives on Genetic Resource Management*, edited by Julius van der Werf, Hans-Ulrich Graser, Richard Frankham, and Cedric Gondro, 3–14. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-9005-9_1.

Tao, Jiemeng, Delong Meng, Chong Qin, Xueduan Liu, Yili Liang, Yunhua Xiao, Zhenghua Liu, Yabing Gu, Juan Li, and Huaqun Yin. 2018. “Integrated Network Analysis Reveals the Importance of Microbial Interactions for Maize Growth.” *Applied Microbiology and Biotechnology* 102 (8): 3805–18. https://doi.org/10.1007/s00253-018-8837-4.

Tecon, Robin, Ali Ebrahimi, Hannah Kleyer, Shai Erev Levi, and Dani Or. 2018. "Cell-to-Cell Bacterial Interactions Promoted by Drier Conditions on Soil Surfaces." *Proceedings of the National Academy of Sciences* 115 (39): 9791–96. https://doi.org/10.1073/pnas.1808274115.

Tecon, Robin, and Dani Or. 2017. "Cooperation in Carbon Source Degradation Shapes Spatial Self-Organization of Microbial Consortia on Hydrated Surfaces." *Scientific Reports* 7 (March). https://doi.org/10.1038/srep43726.

Thamdrup, Bo, Kai Finster, Jens Würgler Hansen, and Friedhelm Bak. 1993. "Bacterial Disproportionation of Elemental Sulfur Coupled to Chemical Reduction of Iron or Manganese." *Applied and Environmental Microbiology* 59 (1): 101–8.

Torres-Monroy, Ingrid, and Matthias S. Ullrich. 2018. "Identification of Bacterial Genes Expressed During Diatom-Bacteria Interactions Using an in Vivo Expression Technology Approach." *Frontiers in Marine Science* 5. https://doi.org/10.3389/fmars.2018.00200.

Twickler, Mark S., Mary Jo Spencer, W. Berry Lyons, and Paul A. Mayewski. 1986. "Measurement of Organic Carbon in Polar Snow Samples." *Nature* 320 (6058): 156–58. https://doi.org/10.1038/320156a0.

Van Valen, Lee. 1973. "A New Evolutionary Law." *Evolutionary Theory* 1: 1–30.

Vartoukian, Sonia R., Aleksandra Adamowska, Megan Lawlor, Rebecca Moazzez, Floyd E. Dewhirst, and William G. Wade. 2016. "In Vitro Cultivation of 'Unculturable' Oral Bacteria, Facilitated by Community Culture and Media Supplementation with Siderophores." *PLOS ONE* 11 (1): e0146926. https://doi.org/10.1371/journal.pone.0146926.

Vasse, Marie, Robert J. Noble, Andrei R. Akhmetzhanov, Clara Torres-Barceló, James Gurney, Simon Benateau, Claire Gougat-Barbera, Oliver Kaltz, and Michael E. Hochberg. 2017. "Antibiotic Stress Selects against Cooperation in the Pathogenic Bacterium Pseudomonas Aeruginosa." *Proceedings of the National Academy of Sciences* 114 (3): 546–51. https://doi.org/10.1073/pnas.1612522114.

Velez, Patricia, Laura Espinosa-Asuar, Mario Figueroa, Jaime Gasca-Pineda, Eneas Aguirre-von-Wobeser, Luis E. Eguiarte, Abril Hernandez-Monroy, and Valeria Souza. 2018. "Nutrient Dependent Cross-Kingdom Interactions: Fungi and Bacteria From an Oligotrophic Desert Oasis." *Frontiers in Microbiology* 9 (August). https://doi.org/10.3389/fmicb.2018.01755.

Zelezniak, Aleksej, Sergej Andrejev, Olga Ponomarova, Daniel R. Mende, Peer Bork, and Kiran Raosaheb Patil. 2015. "Metabolic Dependencies Drive Species Co-Occurrence in Diverse Microbial Communities." *Proceedings of the National Academy of Sciences* 112 (20): 6449–54. https://doi.org/10.1073/pnas.1421834112.

Zeng, Hong, and Aidong Yang. 2019. "Modelling Overflow Metabolism in Escherichia Coli with Flux Balance Analysis Incorporating Differential Proteomic Efficiencies of Energy Pathways." *BMC Systems Biology* 13 (1): 3. https://doi.org/10.1186/s12918-018-0677-4.

Zengler, Karsten, and Livia S. Zaramela. 2018. "The Social Network of Microorganisms — How Auxotrophies Shape Complex Communities." *Nature Reviews Microbiology* 16 (6): 383–90. https://doi.org/10.1038/s41579-018-0004-5.

Ziegelhoffer, E.C., and T.J. Donohue. 2009. "Bacterial Responses to Photo-Oxidative Stress." *Nature Reviews Microbiology* 7 (12): 856–63. https://doi.org/10.1038/nrmicro2237.

Chapter II - Do Organic Substrates Drive Microbial Community Interactions in Arctic Snow?

1 Abstract

The effect of nutrients on microbial interactions, including competition and collaboration, has mainly been studied in laboratories, but their potential application to complex ecosystems is unknown. Here, we examined the effect of changes in organic acids among other parameters on snow microbial communities *in situ* over 2 months. We compared snow bacterial communities from a low organic acid content period to that from a higher organic acid period. We hypothesized that an increase in organic acids would shift the dominant microbial interaction from collaboration to competition. To evaluate microbial interactions, we built taxonomic co-variance networks from OTUs obtained from 16S rRNA gene sequencing. In addition, we tracked marker genes of microbial cooperation (plasmid backbone genes) and competition (antibiotic resistance genes) across both sampling periods in metagenomes and metatranscriptomes. Our results showed a decrease in the average connectivity of the network during late spring compared to the early spring that we interpreted as a decrease of cooperation. This observation was strengthened by the significantly more abundant plasmid backbone genes in the metagenomes from the early spring. The modularity of the network from the late spring was also found to be higher than the one from the early spring, which is another possible indicator of increased competition. Antibiotic resistance genes were significantly more abundant in the late spring metagenomes. In addition, antibiotic resistance genes were also positively correlated to the organic acid concentration of the snow across both seasons. Snow organic acid content might be responsible for this change in bacterial interactions in the Arctic snow community.

2 Introduction

Dynamic changes in nutrient concentrations have been shown to influence bacterial interactions with ramifications for microbial community structure and function (Friedman and Gore, 2017; Khan et al., 2018). In these pure culture studies, either cooperation or competition was the dominant interaction strategy depending on the nutrients considered and their concentrations (Brockhurst et al., 2008, 2010, Lambert et al., 2011, 2014; Ravindran, 2017). Interference competition was hypothesized to be mediated by antibiotic release (Cornforth and Foster, 2013; Oliveira et al., 2015; Ponce-Soto et al., 2015; Song et al., 2017) and was shown to be affected by the nutrient supply (Hol et al., 2014). For example, a sensitive *Escherichia coli* strain was observed to co-exist with a colicin-secreting *E. coli*strain when co-cultivated on a poor growth medium (sugars), but not on a rich medium (amino acids and peptides), where the colicin-secreting *E. coli* strain released antibiotics (Hol et al., 2014). Cooperation was also proposed to be mediated by either metabolic or genetic exchanges between different collaborative strains (Nogueira et al., 2009; Mc Ginty et al., 2011; Dimitriu et al., 2014, 2015; Benomar et al., 2015; Wall, 2016; Tecon and Or, 2017) and has also been shown to be affected by nutrient supply (Benomar et al., 2015). Several studies have examined the importance of horizontal gene

transfer in maintaining cooperation in synthetic bacterial communities (Czárán and Hoekstra, 2009; Nogueira et al., 2009; Dimitriu et al., 2014, 2015; Wall, 2016). Therefore, cooperation might be promoted by increasing assortment among cooperative alleles (Dimitriu et al., 2014) or by increasing kin selection (Nogueira et al., 2009; Wall, 2016). In addition, most of the genes coding for public goods appeared to be preferentially localized on mobile genetic elements (plasmids) and at hotspots of genome recombination (Nogueira et al., 2009).

The majority of research concerning nutrient-related effects on bacterial interactions has been generated with culture-based experiments (Mitri and Foster, 2013). While these studies have provided information on different nutrient effects on bacterial interactions under controlled conditions, they might not predict microbial interactions in the environment. Microcosm or mesocosm approaches have been used more recently to study microbial communities and the results have varied (Ponce-Soto et al., 2015; Ali et al., 2016; Song et al., 2017). Although no studies concerning the effect of carbon content on microbial interactions have been published to date, one study measured an increase in antibiotic resistance genes in strains of *Enterococcus faecalis* cultivated in eutrophic sediment mesocosms amended with nitrogen and phosphorus (Ali et al., 2016). Other studies observed a decline of antibiotic resistance in cultivable bacterial populations from an oligotrophic lake in mesocosms amended with nitrogen and phosphorus and from soil bacteria cultivated on agar plates amended with increasing nutrient medium concentrations (Ponce-Soto et al., 2015; Song et al., 2017). The main difference between these two sets of studies is that one used a PCR based method to track antibiotic resistance (Ali et al., 2016), while the others used culture based methods (Ponce-Soto et al., 2015; Song et al., 2017). Culture based techniques could have a higher bias since they alter the bacterial community by selecting members able to grow on media.

Nutrient dynamics also affect bacterial community structure (Campbell et al., 2010). For example, an increase in organic matter during soil fertilization was shown to decrease bacterial community evenness in Arctic tundra soil (Koyama et al., 2014). The observed effect of nutrients on bacterial community structure might be indirect and mediated in part by bacterial interactions. The low cultivability associated with environmental bacteria might be mainly due to the co-dependency of bacteria that are auxotrophic for some critical functions and, therefore, are obligate co-operators (Pande and Kost, 2017). Thus, bacterial communities might be viewed as networks of cooperating and competing individuals. Such a view has been explored by recent experiments that show a differential growth rate of environmental bacterial strains when co-cultured with other specific strains (Pande et al., 2014; Ren et al., 2015; Vartoukian et al., 2016). Bacterial interactions could provide a selective advantage to bacterial species as a function of nutrient concentrations and subsequently influence bacterial community structure.

Tracking bacterial interactions *in situ* can be performed through networks, such as co-variance networks based on taxonomic data (Faust and Raes, 2012). This approach has been used for microbial communities from oceans (Ruan, 2006; Lima-Mendez et al., 2015), soils (Barberán et al., 2012; Ding et al., 2015), human microbiomes (Faust et al., 2012) and heavy-metal-polluted sediments (Yin et al., 2015). These networks often use co-variance to infer positive (cooperative) and negative (competitive) bacterial interactions (e.g., Ruan, 2006), but co-variance might also indicate that the populations are responding to other stimuli, simultaneously. An approach combining gene markers for bacterial interactions based on pure culture studies and taxonomy-based co-variance networks described above should strengthen the results obtained. Here, we applied this combined approach using antibiotic

resistance as the surrogate for competition and plasmid structural genes for collaboration, and taxonomy-based co-variance networks on microbial communities sampled from an Arctic snowpack over the spring season. Arctic snow microbial communities were selected because arctic snow carbon content varies by several orders of magnitude during the spring season (Twickler et al., 1986) and is generally considered a low carbon environment. Recently, using COG functions characteristic of oligotrophy or copiotrophy as proposed by Lauro et al. (2009), Maccario et al. (2019)showed that arctic snow bacterial communities were adapted to oligotrophic lifestyles. However, oligotrophic the arctic snow environment is, carbon content increases over the spring season (Hacking et al., 1983; Twickler et al., 1986; Haan et al., 2001; Grannas et al., 2007). In addition, Arctic snow has varying nutrient conditions that affect bacterial community structure and function (Larose et al., 2013). We hypothesized that increases in organic acids (as a soluble subset of potential organic substrates) in the warming spring snow would increase competition (and reduce collaboration).

3 Materials and Methods

3.1 Field Sampling

Snow samples were collected during a 2011 springtime field campaign in Ny-Ålesund (Svalbard, Norway, 78°56'N, 11°52'E). Surface snow layers (upper 3 cm) (2L meltwater equivalent) were collected into sterile bags using a sterilized shovel as described previously (Larose et al., 2010a). A total of 31 samples were collected between mid-April to beginning of June 2011. The spring research campaign was held between April, 2011 and June, 2011 at Ny Ålesund in the Spitsbergen Island of Svalbard, Norway (78°56'N, 11°52'E). The field site, a 50 m^2 perimeter with restricted access (to reduce contamination from human sources), is located along the south coast of the Kongsfjorden, which is oriented SE-NW and open to the sea on the west side (Supplementary Figure S1). We added a map in supporting information. The Kongsfjorden was free of sea ice throughout the campaign. Specific sampling dates can be found in the chemistry table (see dataset at Supplementary Table S1). In addition, different weather and snow conditions were monitored over the sampling period (Supplementary Figure S2). Samples for snow chemistry were collected, stored frozen, sent back to the laboratory in France for analysis as described in Larose et al. (2010a, b). Snow samples collected for microbiology were processed immediately after collection in the field laboratory. Samples were left to melt at room temperature prior to filtering onto sterile 0.22 µM 47 mm filters (Millipore) using a sterile filtration unit (Nalge Nunc International Corporation) and filters were stored in Eppendorf tubes filled with the extraction buffer from the PowerWater extraction kit (MoBio) at −20°C for further analysis. Samples for major ions and particles were collected in sterile polycarbonate Accuvettes© sealed with polyethylene caps. All samples were stored frozen (−20°C) and in the dark until analysis.

3.2 Chemical Analysis

Samples were melted in a class 100 clean room at LGGE-CNRS laboratory (Grenoble, France). They were then transferred into Dionex glass vials previously rinsed with ultra-pure Millipore water (conductivity > 18.2 mΩ, TOC < 10 ng/g) and analyzed less than 24 h after melting. Analyses were performed by conductivity-suppressed ion chromatography using a Dionex ICS 3000© apparatus and a Dionex AS40© autosampler placed in the clean room facilities. Different chemical parameters were measured during this study (e.g., major/minor ions, organic acids, and pH). Soluble anions (methyl sulfonic acid (MSA), SO_4, NO_3, Cl) and cations (Na, NH_4, K, Mg, Ca) and organic acids were analyzed by ionic chromatography (IC, Dionex ICS3000). AS/AG 11HC and CS/CG 12A columns were used for anions and cations analyses,

respectively. All chemical analyses were carried out at on the airOsol platform of the IGE laboratory in Grenoble, France. This data set can be found in Supplementary Table S1. The following parameters were used for statistical analyses [Organic acids (oxalate, lactate, glutarate, propionate, succinate, formate, acetate), NO_3^-, NH_4^+, SO_4^{2-}, mercury, fluoride, calcium, magnesium, bromide, strontium, lithium, sodium, chloride, potassium, number of particles, methyl sulfonic acid (MSA)] and pH. For values below the detection limit, we used the detection limit divided by 2.

3.3 DNA Extraction and Sequencing

The DNA from 20 surface snow samples collected between April and May 2011 (CH3N-1 to CH3N-37 or early spring ES) and 16 surface snow samples collected from May to June 2011 (CH3N-40 to CH3N-76 or late spring LS) were extracted for taxonomic analysis. Snow was melted at 4°C before filtering on 0.2 µm filters. DNA was extracted from filters using the DNeasy PowerWater Kit (Qiagen) following the manufacturer's instructions. Then, the DNA was quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific) and the V3–V4 regions of the 16S rRNA genes were amplified by a PCR of 35 cycles at 92°C 30 s, 55°C 30 s and 72°C. air Forward primer is composed of the Illumina adapter 5'TC GTCGGCAGCGTCAGATGTGTATAAGAGACAG coupled to the 16s rRNA gene primer part CCTACGGGNGGCWGCAG and Reverse primer is composed of the Illumina adapter 5'GT CTCGTGGGCTCGGAGATGTGTATAAGAGACAG coupled to the 16s rRNA gene primer part GACTACHVGGGTAT CTAATCC. The 16S rRNA gene primers are from Klindworth et al. (2013). Simultaneous adapter insertion and amplification was performed using the Platinum PCR SuperMix (Invitrogen). Libraries for 16S rRNA gene sequencing were prepared using the 16S rRNA gene Library Preparation Workflow recommended by Illumina. Paired end sequencing was then carried out on a MiSeq sequencer (Illumina) at the laboratory in Lyon. Size of samples before and after clustering is provided in Supplementary Table S2.

Eight samples (CH3N−1 to CH3N−10) collected between April and May and twelve samples (CH3N-40 to CH3N-66) collected between May and June underwent metagenomic and metatranscriptomic sequencing. Not all samples were analyzed for 16S rRNA genes as some of the metagenomic samples did not have any DNA remaining for the 16S rRNA analysis. In addition, we selected extra samples for the 16S rRNA gene based network analysis. For the metatranscriptomic/metagenomic analyses, total nucleic acids were extracted using PowerWater RNA isolation kit (MoBio) following the manufacturer's instructions, except that the DNAse treatment step was omitted. The RNA fraction of nucleic acids was then further purified using RNeasy kit from Qiagen following the manufacturer's instructions. cDNA libraries were prepared from RNA using Tetro cDNA synthesis kit (Bioline). DNA and cDNA samples were then amplified using multiple displacement amplification with the illustra™ GenomiPhi™ HS DNA Amplification Kit (GE Healthcare) since concentrations were too low for library preparation and sequenced using a Roche 454 Titanium pyrosequencer to generate longer reads than illumina MiSeq. Not all samples had sufficient amounts of DNA for sequencing, resulting unbalanced groups (i.e., 8 for ES and 12 for LS). The reads produced from the 454 were 350 bp ± 100 bp average fragment length following quality filtering (Supplementary Figure S3). The depth of sequencing for each sample is reported in Supplementary Table S3. Sequences are publically available at ftp://ftp-adn.ec-lyon.fr/Snow_organic_acids_bacterial_interactions.

3.4 Bioinformatic Pipeline for Quality Filtering, *de novo* Clustering, and 16S rRNA Gene Annotation

We used USEARCH (v 9.2) and the UPARSE pipeline (Edgar, 2013) for quality filtering and clustering of our 16S rRNA gene datasets (for details on parameters used see Supplementary Material and also the provided script). We annotated the representative sequence of each cluster using RDP classifier (Wang et al., 2007) with a bootstrap threshold of 80%. We normalized the OTU counts by using the R package MetagenomeSeq (Paulson et al., 2013).

3.5 Metagenomic and Metatranscriptomic Annotation and Dataset Generation

The raw files from 454 pyrosequencing were processed using Mothur (Schloss et al., 2009) for quality filtering with the settings recommended in Schloss et al. (2011). FastQC (Andrews, 2010) was also used to control for base overrepresentation. Some remains of adapters were found and Usearch (Edgar, 2010) was used to trim our sequences. The resulting.fastq files were functionally annotated using EggNOG-Mapper (Huerta-Cepas et al., 2017), based on eggNOG orthology data (Huerta-Cepas et al., 2016), using the default parameters. The sequence searches were performed using diamond (Buchfink et al., 2015). Resulting annotations were imported into R (R Development Core Team, 2011) to build gene count tables. Reads annotated as eukaryotic sequences were filtered out based on the tax id associated to each sequence annotation using the R package taxize to obtain a bacterial and archaeal dataset (Chamberlain and Szöcs, 2013). The "Retrieve/ID mapping" function[1] from uniprot was used to convert the string ids (EggNOG) into uniprot protein names to generate functional gene tables for each metagenomic and metatranscriptomic dataset. The GO annotation associated to these protein names was used for subsequent analyses.

3.6 Chemical/Molecular Biology Data Analysis

The chemical data were evaluated for differences between sample groups. Data were log transformed (except pH) and a PCA was calculated using the ade4 package (Dray and Dufour, 2007) in R. Co-inertia analysis (Dolédec and Chessel, 1994) was used to test the impact of snow chemistry on bacterial communities using the R package ade4 (Dray and Dufour, 2007). Chemical data sets were compared to microbial taxonomy (OTU table 16S rRNA gene at the genus level), metagenomes (gene annotation level and EggNOG-Mapper annotations) and metatranscriptomes (gene annotation level and EggNOG-Mapper annotations). The significance of each co-inertia was tested using a permutation test (10000 permutations).

3.7 ANOSIM Analysis

The OTU tables were processed with the ADONIS function from the vegan (Dixon, 2003) package in R to carry out ANOSIM (ANalysis Of SIMilarities) analysis. This is a non-parametric test to detect whether more similarities exist between samples inside a sampling group than with the rest of the dataset. We used this method with a randomization test (10000 permutations) to test for differences in similarity between the groups of samples from early spring (ES) and late spring (LS).

3.8 Network Analysis with the OTUs

Based on the OTU tables generated previously with USEARCH for ES and LS groups, a co-variance network was built. Prior to building the network, a filtering step was used to remove OTUs present in less than eight samples (50% of the samples used to build each network). FastLSA (Durno et al., 2013), an improved version of LSA (Local Similarity Analysis)

(Ruan, 2006) was used to compute the networks. LSA has been shown by Weiss et al. (2016) to detect significant co-variance on time series data. We used a lag of zero and filtered out the results that were not significant at the 95% confidence interval (p-val < 0.05). These data were then imported into R and the packages igraph (Csardi and Nepusz, 2006) and GGally, which is an extension from ggplot2 (Wickham, 2009), were used to visualize the co-variance networks obtained. After the network assembly, we compared their respective densities.

3.9 Functional Analysis of Microbial Communities

Metagenomes and metatranscriptomes were pooled into groups based on chemical analysis and co-inertia results. Four groups were determined: early spring (ES) metagenomes, early spring (ES) metatranscriptomes, late spring (LS) metagenomes and late spring (LS) metatranscriptomes. Annotation diversity and differences in profiles between the genes retrieved in the metagenomes and the metatranscriptomes of these groups were compared with Venn diagrams using R package limma (Ritchie et al., 2015). Differential protein gene abundance was compared between the metagenomic profiles of the ES and LS groups using the R package edgeR (Robinson et al., 2010). The p-value was set at 0.05.

3.10 Plasmid Marker and Antibiotic Gene Identification in Metagenomes and Metatranscriptomes

Plasmid structural related protein names were identified by retrieving the proteins annotated with the GO term id GO:0005727 (extrachromosomal circular DNA). In addition, a regular search of protein names using the keyword “plasmid” was carried out. Antibiotic response GO terms were extracted using a custom set of protein names retrieved from Uniprot (Supplementary Table S4 for complete list). Protein names annotated with the GO id GO:0017000 (antibiotic biosynthetic process) were also used. To mine for antibiotic resistance genes determinants (ARGDs) in both our metagenomic and metatranscriptomic datasets, reads were also annotated using Diamond blastx (Buchfink et al., 2015) against the CARD database (McArthur et al., 2013). All the hits that were returned with an e-value lower than 10^{-10}, a z-score higher than 50 and a sequence similarity higher than 60% were considered as significant. For all the annotations, the best hit method was adopted to retrieve one unique annotation per read. Annotations were normalized by the total read count from their respective sample (after the removal of eukaryotic sequences from the total read counts).

4 Results

4.1 Snow Chemistry

Changes in snow chemical composition were monitored during the spring sampling period (April to June 2011, Supplementary Table S1). The chemical composition in early spring samples (ES) was different (PERMANOVA p-value = 0.0015) than late spring samples (LS) as shown by principal component analysis (PCA) (Figure 1). The difference in the early and late spring samples was due to the increase in most organic acids (acetate, oxalate, succinate and formate) and a decrease in lactate concentrations in late spring as well as changes in pH.

Many inorganic salts (e.g., sulfate, bromide) were at higher concentrations in the early spring samples.

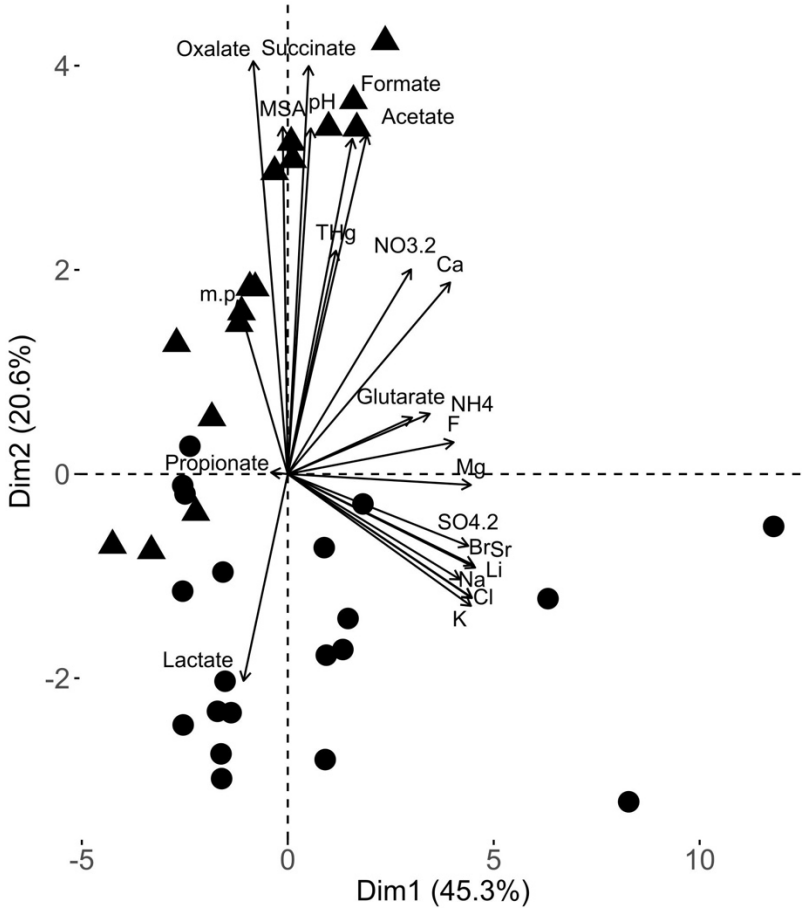


Figure 1: Principal component analysis biplot from the snow chemical analyses of the samples used in this study. The different chemical variables considered in this PCA are represented by vectors. The samples [black dots (early spring samples) and triangles (late spring samples)] are represented based on their respective projections.

4.2 Relationship Between Snow Chemistry and Microbial Data During Early and Late Spring

The co-variance of the chemistry and taxonomic datasets was determined (co-inertia coefficient RV = 0.48, *p*-value = 0.01). The co-inertia analysis did not highlight any clear relationship between taxonomy and chemistry. The metagenomic and metatranscriptomic relative abundances in different functional classes also co-varied with snow chemistry (Table 1). The level of annotation (i.e., proteins vs. gene onthology (GO) categories) influenced their relative co-variance. The co-inertia coefficient (RV) was the highest for metagenomic (vs. metatranscriptomic) datasets when using the GO terms. The co-inertia plot was similar to the PCA carried out using the chemistry data (Supplementary Figure S4). We observed a separation between the samples from the early and late spring along the first axis of the co-inertia plot (Supplementary Figure S4). The chemical variables with the highest influence on first axis of the co-structure were organic acids (acetate, succinate, oxalate and formate and lactate), pH and some major ions (fluoride, calcium). Similar to the OTU analysis, no specific proteins were found to have a significantly higher contribution to the co-inertia.

Table 1: Comparison of the different co-inertia calculated with the snow chemistry of the different snow samples and the different datasets such as 16S rRNA sequence clusters and the metagenomes/metatranscriptomes annotations determined with the Eggnog mapper.

Dataset	Annotation	Co-inertia RV	*p*-value
Metagenomes	Genes id	0.44	0.033
	GO terms	0.45	0.003
	Kegg pathways	0.59	0.0002
Metatranscriptomes	Genes id	0.43	0.072
	GO terms	0.44	0.023
	Kegg pathways	0.37	0.064
16S rRNA sequencing	OTU 97% id	0.48	0.01

4.3 Bacterial Community Structure

After filtering of the 16S rRNA gene reads, the samples had an average of 16 757 reads and a median of 8944 reads. Based on the annotation of cluster seeds using RDP classifier, the observed genera were mainly affiliated to Proteobacteria, Cyanobacteria, Bacteroidetes, Acidobacteria, Firmicutes and Actinobacteria. Linear correlation between individual variables was low (R = 0.14) and the analysis of similarity (ANOSIM) of the 16S rRNA gene derived OTUs from the early and late spring samples had a p-value = 0.03 (perm = 10 000). SIMPER analysis showed that the contribution from any individual OTU to the observed between-groups dissimilarity never exceeded 0.4%. The core community (defined as the OTUs appearing in more than 50% of the samples from one sampling period) from the early spring appeared to be bigger than the one from the late spring (59 vs. 29 OTUs with 17 shared OTUs between the two periods) (Supplementary Table S5). This threshold of 50% was based on the guidelines suggested by Weiss et al. (2016), although different levels up to 80% were examined and these higher values did not change the shared OTUs significantly. These two core communities (59 and 29 OTUs) were then used to build co-variance networks. The variations and annotation of the OTUs varied between samples and time during the spring season (Supplementary Figure S5).

4.4 Exploring Cooperation Using Interaction Networks

More OTUs co-varied positively in the early spring (ES) network than in the late spring (LS) network. The networks from early spring and late spring shared three interactions (Figure 2; red circles). The ES network displayed higher average node connectivity, but had a lower modularity than the network from the LS period (Table 2). The graph density and its transitivity were higher in the LS network, while the average edge betweenness and closeness were found to be higher in the ES network. We also investigated to which extend the size of the networks could be considered as different since the core communities from which the networks were derived were different in size (59 OTUs for ES vs. 29 OTUs for LS core community) (Supplementary Table S5). To do so, we considered the amount of interactions

retrieved as positive in the respective networks (59 vs. 10) and standardized it by the total amount of possible interactions that were possible to build with their respective input sets of OTUs (i.e., which corresponds to a binomial coefficient computed for *n* = number of core OTUs and *k* = 2). This comparison confirmed our initial findings since the ratio of significant positive co-variances observed in each network was higher for ES (0.034) as compared to LS (0.025).

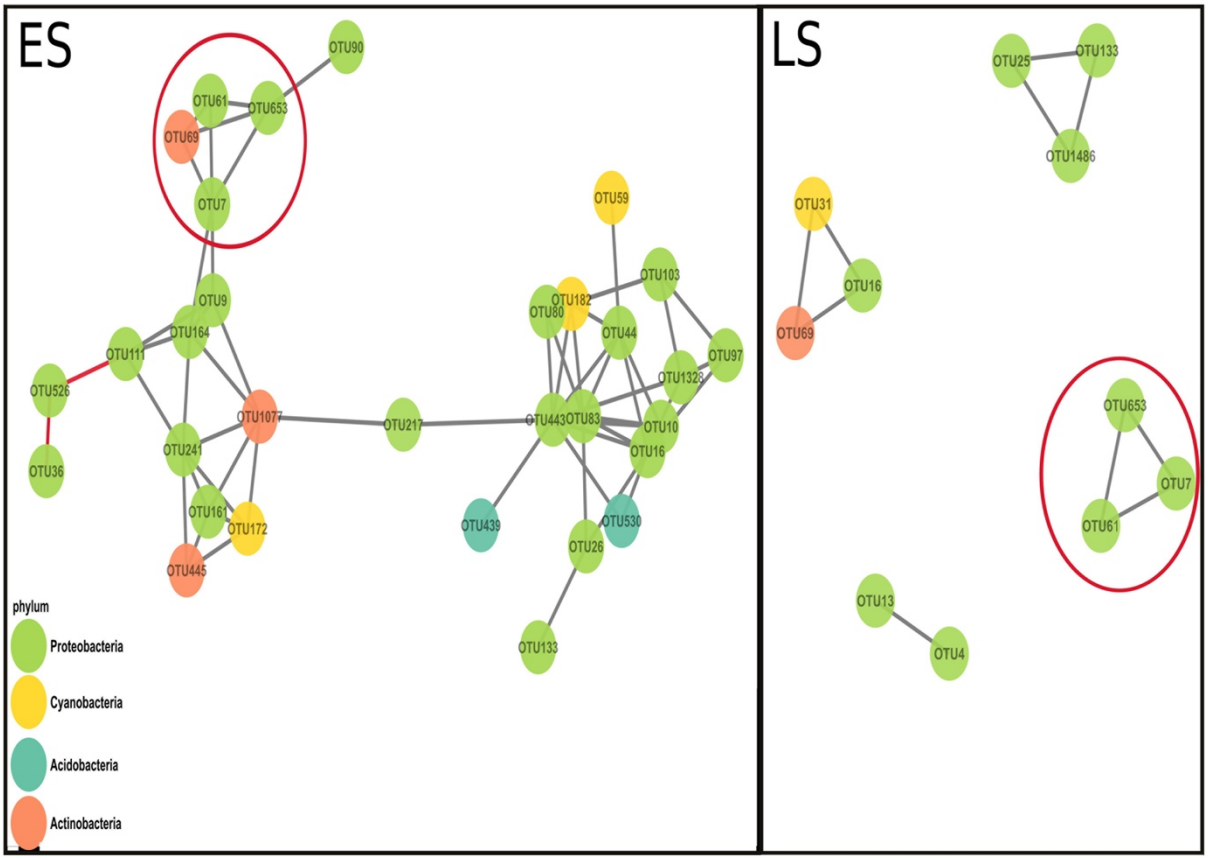


Figure 2: Co-variance networks built from the OTU normalized counts from early spring (ES) and late spring (LS). Each dot represents an OTU (the colors represent different phyla) and each black line represents a positive co-variance (considered as a surrogate of cooperation) and the two red lines in the ES networks represent two negative co-variances (interpreted as a possible competitive interaction). The red circles highlight the interactions that both networks shared. The average connectivity (average amount of positive co-variance a node possesses in a network) is higher in the ES network (=4) compared to the LS network (=1.82). The modularity was higher in the LS network (0.72) than in the ES network (0.532).

Table 2: The main network properties observed in the two co-variance networks build from OTU clusters of 16S rRNA gene sequencing data.

Network property	Early spring	Late spring
Average node connectivity	4	1.82
Modularity	0.532	0.72
Graph density (group adhesion)	0.14	0.18
Networks connectivity (group cohesion)	1	0
Transitivity	0.48	1
Average node closeness (normalized)	0.28	0.11
Average edge betweenness	36.62	0

4.5 Bacterial Community Function

We used the KEGG metabolic pathways obtained from the EGGNOG annotations to determine the main metabolic pathways in the snow metagenomes and metatranscriptomes. The dominant pathways were similar for both metagenomes and metatranscriptomes (Supplementary Tables S6, S7) and were related to amino acid (i.e., arginine and proline metabolism), nucleic acid (i.e., purine/pyrimidine metabolism) and carbohydrate (butanoate, propionate and pyruvate) metabolism/catabolism. Nitrogen metabolism, bacterial chemotaxis, and ABC transporters were also present among the most abundant pathways. Pathways related to vitamin biosynthesis (i.e., folate *biosynthesis*), antibiotic metabolism (i.e., streptomycin and vancomycin biosynthesis pathways), methane metabolism, photosynthesis, cell motility (flagellar assembly), DNA repair, polyunsaturated fatty acid metabolisms as xenobiotic degradation (i.e., naphthalene, ketone) were also identified in the metagenomes and metatranscriptomes. Heatmaps with the 50 most abundant KEGG pathways in our metagenomes and metatranscriptomes are shown in Supplementary Figures S6, S7, respectively.

4.6 Bacterial Community Functional Changes From Early (ES) to Late (LS) Spring

Venn diagrams were constructed at the protein level (gene product) and at the GO term level from the annotated metagenomic and metatranscriptomic datasets. At both the protein level and the GO level, a more diverse group of genes was annotated for LS samples than for ES samples (Figure 3). The metagenomes and metatranscriptomes in late spring shared more genes between them than they did in early spring. In addition, the overlap between early spring metatranscriptomes and late spring metagenomes was larger than the overlap between early spring metagenomes and early spring metatranscriptomes. The overlap between early spring and late spring metatranscriptomes was larger than the overlap between early spring metatranscriptomes and metagenomes.

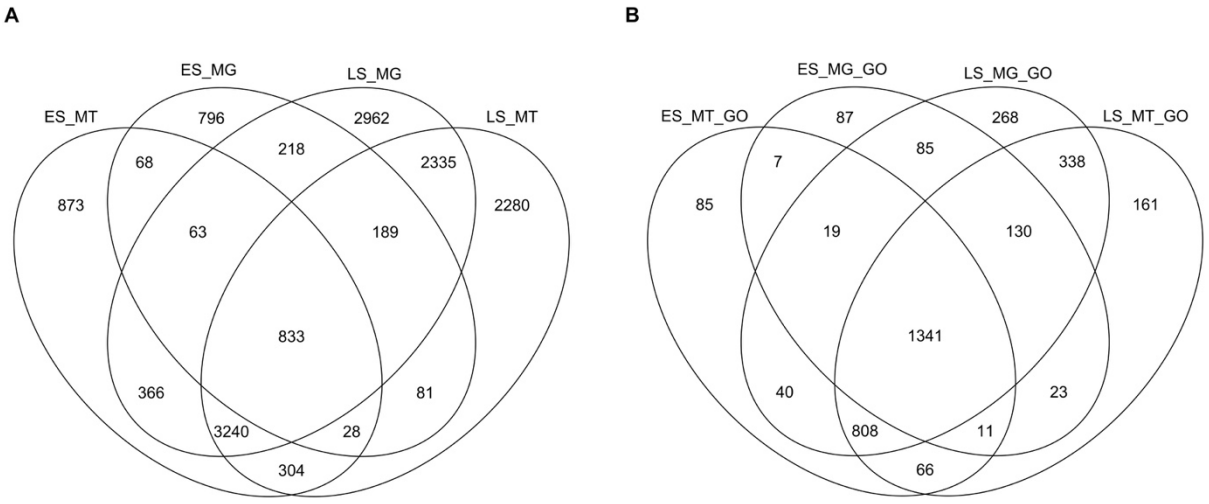


*Figure 3: Venn diagrams displaying the functional overlap from the metagenomes (MG) and the metatranscriptomes (MT) from the early spring (ES) and the late spring (LS) periods based on two different levels of annotations (using EGGNOG-mapper) retrieved using UNIPROT: **(A)** protein name level and **(B)** GO (gene onthology) categories.*

The GO categories that were more abundant in ES metagenomes and metatranscriptomes were related to resistance to chloramphenicol, plasmid maintenance, and cellular stress like ribophagy and autophagy (see Supplementary Table S8 for details). Among the GO categories that were more abundant in LS metagenomes and metatranscriptomes, several were related to lactate/oxalate catabolism and acetate and formate metabolism as well as phosphate starvation (see Supplementary Table S9 for details). Some examples for acetate include cation/acetate symporter (log FC 4.4, *p*-value 0.004) and acetyl-coenzyme A synthetase (logFC 4.1, *p*-value 0.001). The proteins names retrieved as being in relation with organic acid catabolism were formyl-CoA:oxalate CoA-transferase (FCOCT) and formate dehydrogenase (FDH). The tax ids from those genes were from bacterial species from the Comamonadaceae and the Ralstoniaceae, two families from the order of Burkholderiales. Virus related terms (i.e., viral process and capsule organization) were also more abundant in the late spring samples.

In total, 1463 proteins were shown to be significantly more abundant in the metagenomic dataset from either of the two sampling periods by EdgeR (see Supplementary Table S10 for more abundant in ES and Supplementary Table S11 for more abundant in LS for details) of which 125 were more abundant in ES metagenomes (logfold < 0), while 1338 were more abundant in LS metagenomes (logfold > 0) (Figure 4). The annotated proteins that were most enriched in the ES metagenomes with the largest logfold changes between early and late spring were linked to chloramphenicol resistance (logFold = −11.5), plasmid structure genes (logFold = −7.6) (Supplementary Table S10 for details). Annotated proteins involved in plasmid maintenance and plasmid partition were more abundant in the early spring (Table 3). Annotated proteins that had the largest logfold changes between late and early spring were linked to environmental sensing (logFold = 7.6–8.3), a membrane-transport protein (logFold = 8.1) and a putative exported protein (logFold = 9.4). Antibiotic resistance proteins (tetR, penicillin binding protein, bleomycin resistance, and macrolide resistance) and proteins involved in antibiotic biosynthesis (amidase) were more abundant in the late spring (Table 4). Sequences related to viruses and chemotaxis were also observed at higher abundances in LS metagenomes (Supplementary Table S11 for details).

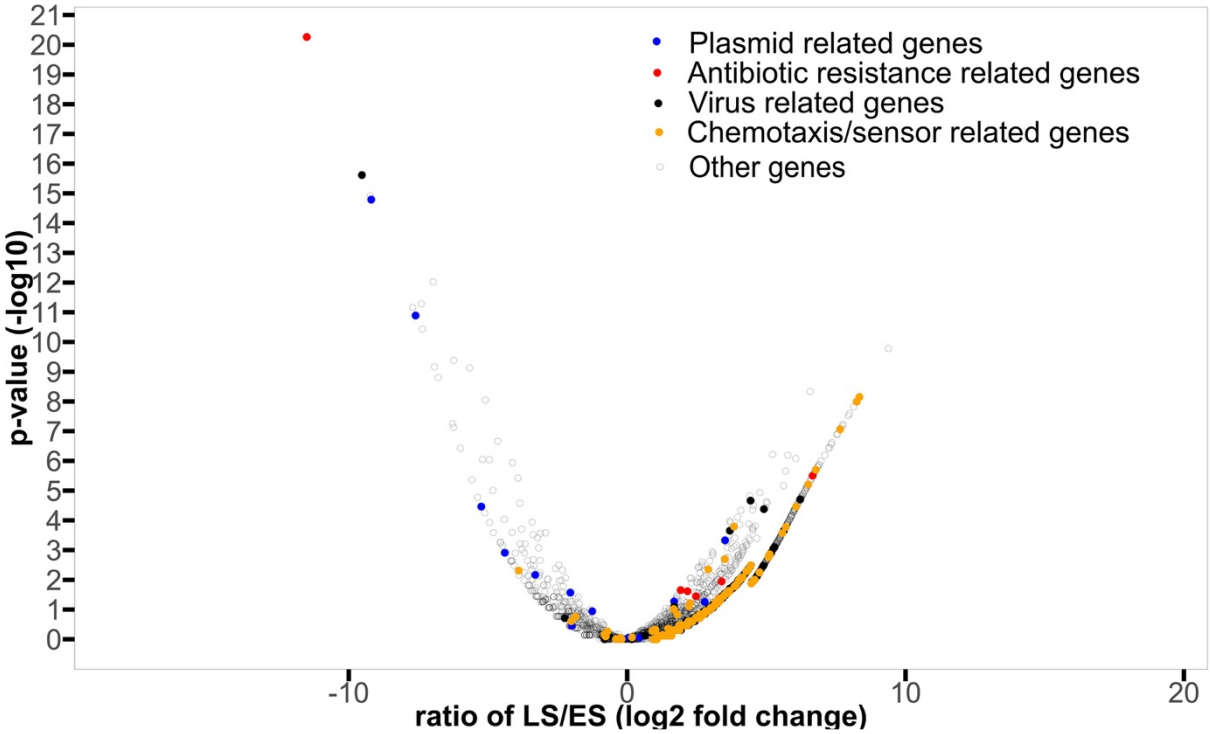


Figure 4: Volcano plot displaying the protein names significantly enriched in early or late spring metagenomes compared to the other period. The log10 of the p-value significance of the differential abundance study retrieved from edgeR is plotted as a function of the logFold change observed for the respective protein names used in the study (filtered out for occurrences lower than two samples). The cutoff of p-val > 0.05 (log10(0.05) = 1.3) has been used in this study. The plasmid structural protein names (replication proteins and toxin anti-toxin complex, considered as surrogate of bacterial cooperation) identified are plotted as blue dots, the antibiotic resistance/synthesis protein names (surrogate of bacterial competition) are plotted as red dot. We plotted protein names related to viruses in black and protein names related to chemotaxis and sensors as orange dots.

Table 3: Protein names related to plasmid structure genes determined by edgeR as being significantly enriched in metagenomes from early spring (logFC < 0) or late spring (logFC > 0).

Protein	logFC	logCPM	*P*-value
Replication initiation protein (Protein E) (Protein rep)	−9.193	12.645	1.62×10^{-15}
Rep protein (Fragment)	−7.600	11.182	1.288×10^{-11}
Putative plasmid maintenance system antidote protein, XRE family	−5.240	9.607	3×10^{-5}
XRE family plasmid maintenance system antidote protein	−4.397	9.255	0.001
Plasmid maintenance system killer	−3.300	9.107	0.007
Plasmid recombination protein	−2.041	10.565	0.027
Plasmid recombination protein.1	−2.041	10.565	0.027
Replication protein	3.517	11.551	0.0005

The logCPM represents the average abundance of the protein name across the whole dataset and is an indicator of how much signal was present in the dataset to test the enrichment with edgeR.

Table 4: Protein names related to antibiotic resistance or synthesis genes returned by edgeR as being significantly enriched in metagenomes from early spring (logFC < 0) or late spring (logFC > 0).

Protein	logFC	logCPM	*P*-value
Chloramphenicol acetyltransferase (EC 2.3.1.28)	−11.509	14.915	5.50×10^{-21}
Transcriptional regulator, TetR family	1.922	11.399	2.2×10^{-2}
Beta-lactamase	2.168	10.564	2.4×10^{-2}
Penicillin-binding protein 1B (PBP-1b) (PBP1b) (Murein polymerase)	2.471	9.616	0.036
Penicillin-binding protein 2	3.301	9.189	0.043
Glyoxalase/bleomycin resistance protein/dioxygenase	3.393	9.801	0.011
Macrolide export ATP-binding/permease protein MacB (EC 3.6.3.-)	3.807	9.312	0.017
Penicillin-binding protein	3.873	9.338	0.017
Putative amidase	6.663	10.997	3.15e − 06

The logCPM represents the log2 average abundance of the protein name across the whole dataset and is an indicator of how much signal was present in the dataset to test the enrichment with edgeR.

4.7 Changes in Antibiotic Resistance Gene Determinants in the Snow

Using the CARD antibiotic resistance gene database (McArthur et al., 2013), metagenomic and metatranscriptomic sequences were annotated for antibiotic resistance genes. The number of the different antibiotic resistance gene determinants (ARGDs) was greater for the late spring samples and the overlap between metagenomes and metatranscriptome ARGDs was higher for the late spring samples (Supplementary Figure S8). Both the number of metatranscriptomic sequences annotated as ARGDs and the diversity of these genes correlated to organic acid concentrations (Figure 5). The annotated early spring taxonomy of the chloramphenicol acetyl-transferase had two tax ids from the database (*Clostridium scindens and Pseudoflavonifractor capillosus*). For the late spring samples, the sequences annotated as the putative amidase were assigned eight different taxa ids (two strains of *Pseudomonas fluorescens, Nocardia farcinica, Gemmatimonas aurantiaca, Sinorhizobium fredii, Rubrivivax benzoatilyticus*, and *Gordonia alkanivorans*). The sequences annotated as the protein MacB involved in macrolide resistance was assigned to four different tax ids (*P. fluorescens, Stenotrophomonas maltophilia, Nostoc* sp., and *Achromobacter insuavis*). Interestingly, *Pseudomonas* was found in the early spring interaction network and implicated in a negative interaction.

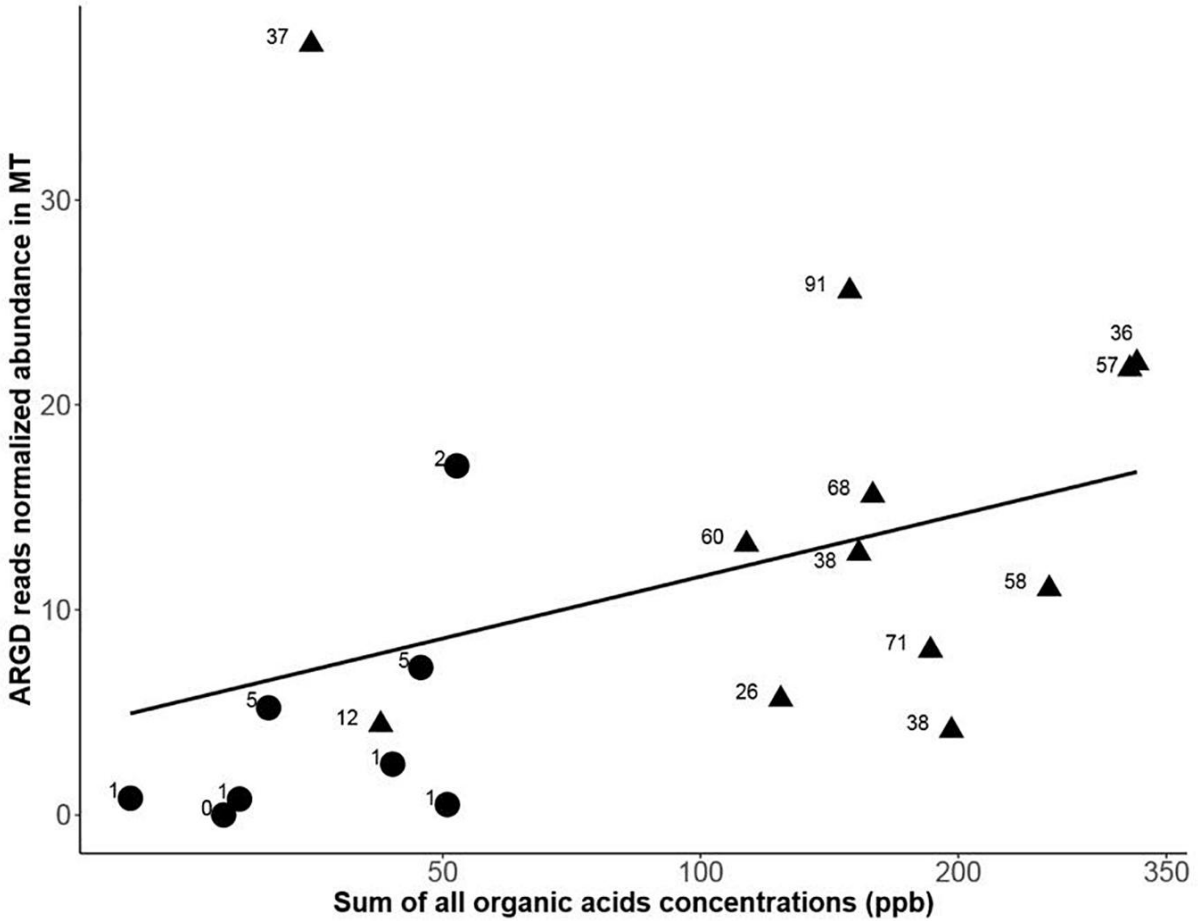


Figure 5: Antibiotic resistance genes (ARGD) transcription annotated from the metatranscriptome datasets (MT) vs. the total sum of organic acids amounts measured in the snow samples (black dot, early spring samples and black triangle, late spring samples). The numbers display the amount of different ARGD genes annotated in each sample. A Spearman correlation between ARGD transcription and total organic acids concentration had a rho = 0.57 and a p-value = 0.010.

5 Discussion

5.1 Interactions Between Organic Acids and Bacterial Communities in Snow

Among the different snow chemical parameters that were tightly coupled to changes in microbial functions (metagenomic) (Table 1), total measured organic acid concentration ranged from around 3 ppb to over 2000 ppb (see Supplementary Table S1). For samples that had metagenomic sequencing performed, the total organic acids ranged from 6ppb to 350 ppb (see Figure 5). Increases in organic acid concentrations were previously observed in Svalbard (Larose et al., 2013) and Greenland snow (Twickler et al., 1986). We also saw an increase in genes related to organic acid metabolism (e.g., acetate catabolism) in LS metatranscriptomes, which could reflect an active response of the snow community. Metatranscriptomes might provide a sensitive and rapid indicator of environmental signals while metagenomes might be more representative of changes over longer periods of time in relation to their chemical environment. The late spring protein-coding genes (both from metagenomic and metatranscriptomic data sets) overlapped more with the metatranscriptomes from the early spring than with the metagenomes from the early spring (Figure 3). These trends were also observed at the GO term annotation level. This pattern might indicate that some of the low abundance active taxa from the early spring season (not observed in the metagenomes but observed in the metatranscriptomes from early spring) became dominant during the late spring season (observed in the late spring metagenomes) and stayed active during this period (also present in late spring metatranscriptomes). This was consistent with the associated taxonomy based on the functional gene annotation where taxa observed only in the early metatranscriptomes but not in the metagenomes that were also retrieved in the late spring metagenomes and metatranscriptomes (Supplementary Figure S9).

5.2 Bacterial Communities of the Snow Shift From Cooperation Toward Competition as Organic Acid Levels Increased

Plasmids might be involved in cooperative interactions and could serve as a marker for microbial collaboration. For example, genes coding for public goods were preferentially located on mobile elements or close to integrases when incorporated into genomes (Nogueira et al., 2009; Mc Ginty et al., 2011). In addition, conjugation and gene transfer through plasmids were associated with bacterial cooperation (Dimitriu et al., 2014). Gene transfer might drive cooperation among bacteria by increasing their genetic similarity that would select cooperative behavior via kin selection (Nogueira et al., 2009). The sequences related to plasmid structural proteins were more abundant in early spring metagenomes than in late spring when the organic acid concentrations were higher. While this does not show causality, it is consistent with the hypothesis that organic acids might impact microbial interactions.
Antibiotics might be proxies for bacterial competition and their related marker genes (production and resistance) have been used to track bacterial interference competition (Ponce-Soto et al., 2015; Goordial et al., 2017). In our study, sequences annotated as antibiotic resistance and secretion proteins were more abundant in late spring metatranscriptomes and metagenomes (Figure 4). Sequences annotated as putative amidase, penicillin amidase, and penicillin amylase were only observed in the late spring metagenomes. These proteins are known to be involved in some derivatives of penicillin and

lactone biosynthesis; this last molecule is one of the main constituents of macrolide antibiotics (Omura, 2002). We correlated an increase in the number and diversity of antibiotic resistant gene determinants to an increase of organic acid content in the snow (Figure 5). Competition might increase as the environment becomes richer in organic acids and result in bacterial communities actively transcribing genes for an increasingly diverse set of ARGDs. While antibiotic resistance is also sometimes associated with cooperative traits (Cordero et al., 2012), the diversity of antibiotic genes would be low as the entire community shares the public good. In our data sets, only early spring samples had low antibiotic gene diversity (see Table 4), which might be compatible with the hypothesis of antibiotics secreted as a public good to protect the whole cooperative community.

Physical changes of the snowpack might also induce a shift from cooperation to competition. As the season progressed, the snowpack became gradually warmer and wetter. This likely increased motility of the bacterial population within the snow as indicated by an increase in the relative abundance of proteins related to chemotaxis and motility (i.e., receptors, flagella) in late spring samples (Supplementary Tables S9, S10). A decrease in the environmental stratification of the snow ecosystem with observed changes in snow crystal morphology (from faceted crystals to rounded ones) and a loss of snow layers was also apparent throughout the entire spring period. Several studies have shown that bacterial cooperation was counter-selected when the stratification of the environment decreased to the benefit of competitive bacterial strains (F. J. H. Kümmerli et al., 2009; Hol et al., 2013, 2015). The transition from a cold dry snowpack to a warmer wetter one might have led to increased habitat mixing among micro-organisms. Increased mixing could increase the viral-microbial contact, which would lead to increased infection rate (Ashby et al., 2014; Simmons et al., 2018). This possible increased infection rate was consistent with the increased viral related sequences and GO terms in late spring metagenomes (Supplementary Tables S9, S11 for details).

5.3 Microbial Networks Respond to the Shift of Cooperation Toward Competition

Co-variance networks have been used recently to study bacterial interactions and two network characteristics, connectivity and modularity, were considered as proxies for cooperation and competition, respectively. The early spring (ES) network had a higher average connectivity than the late spring network (Figure 2 and Table 2). This was further confirmed by the higher ratio observed between the positive interactions retrieved in the ES network and all the possible interactions than the same ratio for the LS network. We also compared the intensity of the respective co-variances observed in these two networks by looking at their respective local spatial autocorrelation (LSA) coefficients (similar to a correlation coefficient with values between 0 and 1 for positive co-variances) and did not observe any significant differences in their distribution [between 0.81 and 0.91 (Supplementary Figure S10)]. Higher average connectivity can be interpreted as a marker of cooperation within the early spring bacterial community. This property is also related to an increased resistance to change (local resilience) since the presence of several organisms within the network can contribute to resisting to local perturbations (Scheffer et al., 2012). In the context of positive bacterial interactions, metabolic exchanges between the different members of the community could enhance the resilience of the cooperative strains when the nutrient composition changes. As shown by Benomar et al. (2015), nutrient stress can induce metabolic exchanges between two bacterial strains. However, once perturbations are too great, the whole network structure can be transformed (Scheffer et al., 2012). The overlap

between the covariance networks of early and late spring communities was low (only two interactions, see Figure 2), even though their core communities overlapped by more than 50% of the OTUs (Supplementary Table S5). The changes in nutrients over a short period of time and the decrease in environmental stratification might have led to the differences in the positive interaction networks for the bacterial communities from early and late spring snow (Figure 2 and Table 2).

The late spring network displayed a higher modularity than the early spring network. High modularity is linked to a higher adaptive capacity, since the network is more heterogeneous (Scheffer et al., 2012). This network configuration could be more advantageous in a dynamic environment where perturbations are more intense. An increase in environmental perturbations has also been associated with a decline in cooperation (Wilson et al., 2017). This effect was explained by a trade-off between access to nutrients (enhanced by spatial perturbations) and access to an auto-inducer to initiate cooperation (decreased by spatial perturbations) (Wilson et al., 2017). In our data, we observed more GO terms related to stress (mainly due to antibiotics and viruses but also to oxidative and osmotic stress) in the late spring metagenomes relative to the early spring metagenomes.

6 Conclusion

Increase in organic acid concentrations in the snow might have influenced bacterial interactions and led to a shift from cooperation to competition. Several other correlations were observed between community response and environmental chemical parameters. Physical changes of the snow structure leading to decreased stratification and increased mixing might have also contributed. Using a combined method of marker genes and network analysis, we evaluated bacterial interactions in the complex snow microbial communities. Future work should include controlled laboratory studies with snow enriched with organic acids to confirm the trends observed in this field study. In addition, we need to increase our knowledge of genetic markers of microbial interactions since the number of genes currently used to track cooperation and competition is still small and controversial.

7 Bibliography

Ali, L., Wang, Y.-Q., Zhang, J., Ajmal, M., Xiao, Z., Wu, J., Chen, J.-L., and Yu, D. (2016). Nutrient-induced antibiotic resistance in Enterococcus faecalis in the eutrophic environment. J. Glob. Antimicrob. Resist. 7, 78–83.

Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data.

Ashby, B., Gupta, S., and Buckling, A. (2014). Spatial Structure Mitigates Fitness Costs in Host-Paras`ite Coevolution. Am. Nat. 183, E64–E74.

Barberán, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J 6, 343–351.

Benomar, S., Ranava, D., Cárdenas, M.L., Trably, E., Rafrafi, Y., Ducret, A., Hamelin, J., Lojou, E., Steyer, J.-P., and Giudici-Orticoni, M.-T. (2015). Nutritional stress induces exchange of cell material and energetic coupling between bacterial species. Nat. Commun. 6, 6283.

Brockhurst, M.A., Buckling, A., Racey, D., and Gardner, A. (2008). Resource supply and the evolution of public-goods cooperation in bacteria. BMC Biol. 6, 20.

Brockhurst, M.A., Habets, M.G.J.L., Libberton, B., Buckling, A., and Gardner, A. (2010). Ecological drivers of the evolution of public-goods cooperation in bacteria. Ecology 91, 334–340.
Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.
Campbell, B.J., Polson, S.W., Hanson, T.E., Mack, M.C., and Schuur, E.A.G. (2010). The effect of nutrient deposition on bacterial communities in Arctic tundra soil. Environ. Microbiol. 12, 1842–1854.
Chamberlain, S.A., and Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. F1000Research 2.
Cordero, O.X., Wildschutte, H., Kirkup, B., Proehl, S., Ngo, L., Hussain, F., Roux, F.L., Mincer, T., and Polz, M.F. (2012). Ecological Populations of Bacteria Act as Socially Cohesive Units of Antibiotic Production and Resistance. Science 337, 1228–1231.

Cornforth, D.M., and Foster, K.R. (2013). Competition sensing: the social side of bacterial stress responses. Nat. Rev. Microbiol. 11, 285–293.
Czárán, T., and Hoekstra, R.F. (2009). Microbial Communication, Cooperation and Cheating: Quorum Sensing Drives the Evolution of Cooperation in Bacteria. PLOS ONE 4, e6655.
Dimitriu, T., Lotton, C., Bénard-Capelle, J., Misevic, D., Brown, S.P., Lindner, A.B., and Taddei, F. (2014). Genetic information transfer promotes cooperation in bacteria. Proc. Natl. Acad. Sci. U. S. A. 111, 11103–11108.
Dimitriu, T., Misevic, D., Lindner, A.B., and Taddei, F. (2015). Mobile genetic elements are involved in bacterial sociality. Mob. Genet. Elem. 5, 7–11.
Ding, J., Zhang, Y., Deng, Y., Cong, J., Lu, H., Sun, X., Yang, C., Yuan, T., Nostrand, J.D.V., Li, D., et al. (2015). Integrated metagenomics and network analysis of soil microbial community of the forest timberline. Sci. Rep. 5, 7994.
Dixon, P. (2003). VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14, 927–930.
Dolédec, S., and Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species–environment relationships. Freshw. Biol. 31, 277–294.
Dray, S., and Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. J. Stat. Softw. 22, 1–20.
Durno, W.E., Hanson, N.W., Konwar, K.M., and Hallam, S.J. (2013). Expanding the boundaries of local similarity analysis. BMC Genomics 14, S3.
Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.
Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998.
Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. Nat. Rev. Microbiol. 10, 538–550.
Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J., and Huttenhower, C. (2012). Microbial Co-occurrence Relationships in the Human Microbiome. PLOS Comput. Biol. 8, e1002606.
Friedman, J., and Gore, J. (2017). Ecological systems biology: The dynamics of interacting populations. Curr. Opin. Syst. Biol. 1, 114–121.
Gabor Csardi, and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal Complex Systems, 1695.

Goordial, J., Davila, A., Greer, C.W., Cannam, R., DiRuggiero, J., McKay, C.P., and Whyte, L.G. (2017). Comparative activity and functional ecology of permafrost soils and lithic niches in a hyper-arid polar desert. Environ. Microbiol. 19, 443–458.
Grannas, A.M., Jones, A.E., Dibb, J., Ammann, M., Anastasio, C., Beine, H.J., Bergin, M., Bottenheim, J., Boxe, C.S., Carver, G., et al. (2007). An overview of snow photochemistry: evidence, mechanisms and impacts. Atmos Chem Phys 7, 4329–4373.
Haan, D., Zuo, Y., Gros, V., and Brenninkmeijer, C. a. M. (2001). Photochemical Production of Carbon Monoxide in Snow. J. Atmospheric Chem. 40, 217–230.
Hacking, A.J., Taylor, I.W.F., Jarman, T.R., and Govan, J.R.W. (1983). Alginate Biosynthesis by Pseudomonas mendocina. J. Gen. Microbiol. 129, 3473–3480.
Hol, F.J., Voges, M.J., Dekker, C., and Keymer, J.E. (2014). Nutrient-responsive regulation determines biodiversity in a colicin-mediated bacterial community. BMC Biol. 12, 68.
Hol, F.J.H., Galajda, P., Nagy, K., Woolthuis, R.G., Dekker, C., and Keymer, J.E. (2013). Spatial structure facilitates cooperation in a social dilemma: empirical evidence from a bacterial community. PloS One 8, e77042.
Hol, F.J.H., Galajda, P., Woolthuis, R.G., Dekker, C., and Keymer, J.E. (2015). The idiosyncrasy of spatial structure in bacterial competition. BMC Res. Notes 8, 245.
Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44, D286-293.
Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol. Biol. Evol. 34, 2115–2122.
Khan, N., Maezato, Y., McClure, R.S., Brislawn, C.J., Mobberley, J.M., Isern, N., Chrisler, W.B., Markillie, L.M., Barney, B.M., Song, H.-S., et al. (2018). Phenotypic responses to interspecies competition and commensalism in a naturally-derived microbial co-culture. Sci. Rep. 8, 297.
Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glöckner, F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1.
Koyama, A., Wallenstein, M.D., Simpson, R.T., and Moore, J.C. (2014). Soil bacterial community composition altered by increased nutrient availability in Arctic tundra soils. Front. Microbiol. 5.
Kümmerli, R., Griffin, A.S., West, S.A., Buckling, A., and Harrison, F. (2009). Viscous medium promotes cooperation in the pathogenic bacterium Pseudomonas aeruginosa. Proc. R. Soc. Lond. B Biol. Sci. 276, 3531–3538.
Lambert, G., Liao, D., Vyawahare, S., and Austin, R.H. (2011). Anomalous Spatial Redistribution of Competing Bacteria under Starvation Conditions ▿. J. Bacteriol. 193, 1878–1883.
Lambert, G., Vyawahare, S., and Austin, R.H. (2014). Bacteria and game theory: the rise and fall of cooperation in spatially heterogeneous environments. Interface Focus 4.
Larose, C., Berger, S., Ferrari, C., Navarro, E., Dommergue, A., Schneider, D., and Vogel, T.M. (2010a). Microbial sequences retrieved from environmental samples from seasonal Arctic snow and meltwater from Svalbard, Norway. Extremophiles 14, 205–212.
Larose, C., Dommergue, A., De Angelis, M., Cossa, D., Averty, B., Marusczak, N., Soumis, N., Schneider, D., and Ferrari, C. (2010b). Springtime changes in snow chemistry lead to new insights into mercury methylation in the Arctic. Geochim. Cosmochim. Acta 74, 6263–6275.

Larose, C., Prestat, E., Cecillon, S., Berger, S., Malandain, C., Lyon, D., Ferrari, C., Schneider, D., Dommergue, A., and Vogel, T.M. (2013). Interactions between Snow Chemistry, Mercury Inputs and Microbial Population Dynamics in an Arctic Snowpack. PLOS ONE 8, e79972.
Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., DeMaere, M.Z., Ting, L., Ertan, H., Johnson, J., et al. (2009). The genomic basis of trophic strategy in marine bacteria. Proc. Natl. Acad. Sci. 106, 15527–15533.
Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., et al. (2015). Determinants of community structure in the global plankton interactome. Science 348, 1262073.
Maccario, L., Carpenter, S.D., Deming, J.W., Vogel, T.M., and Larose, C. (2019). Sources and selection of snow-specific microbial communities in a Greenlandic sea ice snow cover. Sci. Rep. 9, 2290.
Mc Ginty, S.E., Rankin, D.J., and Brown, S.P. (2011). Horizontal Gene Transfer and the Evolution of Bacterial Cooperation. Evolution 65, 21–32.
McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., et al. (2013). The comprehensive antibiotic resistance database. Antimicrob. Agents Chemother. 57, 3348–3357.
Mitri, S., and Foster, K.R. (2013). The Genotypic View of Social Interactions in Microbial Communities. Annu. Rev. Genet. 47, 247–273.
Nogueira, T., Rankin, D.J., Touchon, M., Taddei, F., Brown, S.P., and Rocha, E.P.C. (2009). Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. Curr. Biol. 19, 1683–1691.
Oliveira, N.M., Martinez-Garcia, E., Xavier, J., Durham, W.M., Kolter, R., Kim, W., and Foster, K.R. (2015). Biofilm Formation As a Response to Ecological Competition. PLOS Biol. 13, e1002191.
Omura, S. (2002). Macrolide Antibiotics: Chemistry, Biology, and Practice (Elsevier).
Pande, S., and Kost, C. (2017). Bacterial Unculturability and the Formation of Intercellular Metabolic Networks. Trends Microbiol. 25, 349–361.
Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., de Figueiredo, L.F., Kaleta, C., and Kost, C. (2014). Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. ISME J 8, 953–962.
Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. Nat. Methods 10, 1200–1202.
Ponce-Soto, G.Y., Aguirre-von-Wobeser, E., Eguiarte, L.E., Elser, J.J., Lee, Z.M.-P., and Souza, V. (2015). Enrichment experiment changes microbial interactions in an ultra-oligotrophic environment. Front. Microbiol. 6, 246.
R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing (Vienna, Austria).
Ravindran, S. (2017). Inner Workings: Bacteria work together to survive Earth's depths. Proc. Natl. Acad. Sci. 114, 788–790.
Ren, D., Madsen, J.S., Sorensen, S.J., and Burmolle, M. (2015). High prevalence of biofilm synergy among bacterial soil isolates in cocultures indicates bacterial interspecific cooperation. ISME J 9, 81–89.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47–e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.
Ruan, Q. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 22, 2532–2538.
Scheffer, M., Carpenter, S.R., Lenton, T.M., Bascompte, J., Brock, W., Dakos, V., Koppel, J. van de, Leemput, I.A. van de, Levin, S.A., Nes, E.H. van, et al. (2012). Anticipating Critical Transitions. Science 338, 344–348.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., et al. (2009). Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.
Schloss, P.D., Gevers, D., and Westcott, S.L. (2011). Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. PLOS ONE 6, e27310.
Simmons, M., Drescher, K., Nadell, C.D., and Bucci, V. (2018). Phage mobility is a core determinant of phage–bacteria coexistence in biofilms. ISME J. 12, 531–543.
Song, H.-K., Song, W., Kim, M., Tripathi, B.M., Kim, H., Jablonski, P., and Adams, J.M. (2017). Bacterial strategies along nutrient and time gradients, revealed by metagenomic analysis of laboratory microcosms. FEMS Microbiol. Ecol. 93.
Tecon, R., and Or, D. (2017). Cooperation in carbon source degradation shapes spatial self-organization of microbial consortia on hydrated surfaces. Sci. Rep. 7.
Twickler, M.S., Spencer, M.J., Lyons, W.B., and Mayewski, P.A. (1986). Measurement of organic carbon in polar snow samples. Nature 320, 156–158.
Vartoukian, S.R., Adamowska, A., Lawlor, M., Moazzez, R., Dewhirst, F.E., and Wade, W.G. (2016). In Vitro Cultivation of 'Unculturable' Oral Bacteria, Facilitated by Community Culture and Media Supplementation with Siderophores. PLOS ONE 11, e0146926.
Wall, D. (2016). Kin Recognition in Bacteria. Annu. Rev. Microbiol. 70, 143–160.
Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Env. Microbiol 73, 5261–5267.
Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., et al. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J 10, 1669–1681.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).
Wilson, C.E., Lopatkin, A.J., Craddock, T.J.A., Driscoll, W.W., Eldakar, O.T., Lopez, J.V., and Smith, R.P. (2017). Cooperation and competition shape ecological resistance during periodic spatial disturbance of engineered bacteria. Sci. Rep. 7.
Yin, H., Niu, J., Ren, Y., Cong, J., Zhang, X., Fan, F., Xiao, Y., Zhang, X., Deng, J., Xie, M., et al. (2015). An integrated insight into the response of sedimentary microbial communities to heavy metal contamination. Sci. Rep. 5, srep14266.

Chapter III - EggVio: a user friendly and versatile pipeline for assembly and functional annotation of shallow depth sequenced samples

1 Introduction

Metagenomic approaches are useful for investigating both the diversity and functioning of environmental microbial communities. Over the last decade, several tools and workflows have been released to assemble and analyse these datasets (*e.g.* Li et al. 2016; Bankevich *et al.* 2012; Wood and Salzberg 2014; Buchfink, Xie, and Huson 2015; Menzel, Ng, and Krogh 2016). Metagenomic data (derived from shotgun sequencing of total extracted DNA) has recently been used to assemble putative genomes, called metagenomic assembled genomes (MAGs), to determine taxonomy and to identify metabolic functions of microbial communities. Sequence reads are assembled into contigs, which improves the accuracy of annotation, and binned into MAGs. Both the contigs and the bins (putative genomes - MAGs) can provide more accurate taxonomical annotations. The accuracy of the processes suggested for assembling the reads in contigs and for binning them are dependent on sequencing depth; if it is too shallow, the coverage of the contigs will be low and only a small fraction of reads will be recruited into the assembly. As a consequence, the results might be less reliable and other strategies must be used.

There are several methods for improving the assembly into contigs and for the subsequent binning into MAGs. Co-assembly is one strategy that has been developed to improve the assembly quality when sequencing depth is limited. If the metagenomic dataset is composed of several samples that are known to contain similar bacterial communities, they can be pooled during the assembly step to increase the sequencing depth and improve the quality of the assembly. In this case, reads from different samples can be used to create a contig. The subsequent binning of contigs has often in the past relied on nucleotide frequency discrimination, although the binning could be improved by using differential coverage of the reads in the contigs in the different samples. Co-assembly of the contigs from multiple samples could, therefore, improve the quality of MAGs retrieved by increasing the discrimination among taxonomically related species or strains. For example, Delmont and Eren (2018) retrieved more MAGs from the Tara ocean dataset, including several strains of *Prochlorococcus*, than found in the original publication (Sunagawa et al. 2015) by using both co-assembly of the contigs and differential coverage for the binning. Co-assembly can also be used to track changes in taxonomy and functions in time-series datasets. If the genes of interest are abundant, shallower sequencing technologies other than hiseq (*e.g.* miSeq) can be used and co-assembly can increase the assembly quality. However, pipelines to deal with these specificities (time series, shallow depth sequencing, low assembly) are currently missing.

Code de champ modifié

Code de champ modifié

In this paper, we introduce a new workflow ("Eggvio") designed to annotate and analyse metagenomic datasets of any size (but optimized for shallow sequencing datasets). Co-assembly was coupled to a read annotation strategy to rescue reads that were not easily mapped back to the assembled contigs. In addition, an algorithm was built to define an e-value threshold based on the noise of read annotation derived from the reads used in the assembly. Several recent tools to carry out the assembly, mapping, binning and annotation were selected to build a user-friendly pipeline. Pipelines for metagenomics generally involve the installation of many tools or dependencies and require some previous informatics knowledge. To improve the reproducibility, accessibility and transparency, we designed a script that allows every user to install all the tools needed for the pipeline.

2 Material and methods

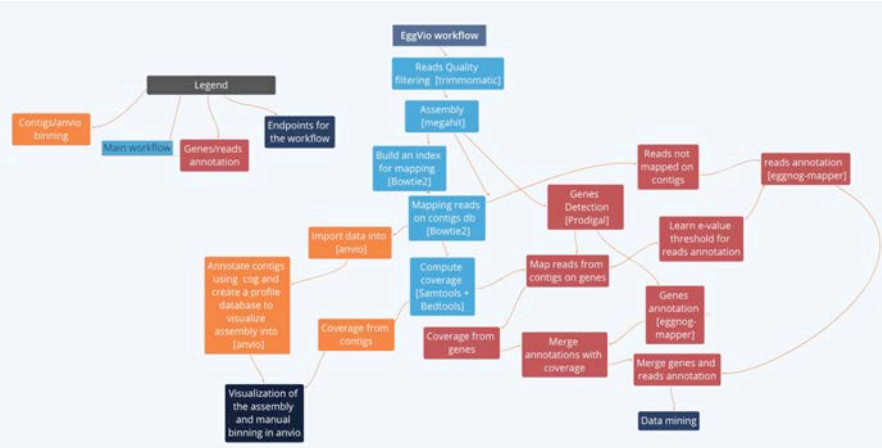


Figure 1: Summary of the workflow. The workflow is divided in three subparts illustrated by different colors. The main workflow (light blue) carries out steps which are mandatory to both MAG binning with anvio (orange) and functional annotation (red) using eggnog-mapper. Tools used to carry out every step of the pipeline are written between brackets inside the box. If no tool is given, then it means that the whole step is carried out using custom R or bash scripts without the need of any other tool.

EggVio is a flexible pipeline optimized for co-assembly, binning and annotation of low depth MiSeq samples (<10^6 reads per sample). Due to the computational needs for read annotation, it was designed to be used on a linux server with a batch slurm queue for job submission. It is coded mainly into bash scripts to process fastq data starting from quality filtering to functional annotations. R (R Development Core Team 2011) programming is used for merging annotations from contigs and reads that did not map back onto the assembly and also to determine the e-value threshold. The programs used to carry out every step of the pipeline are shortly summarized in the flowchart shown in Figure 1. In addition, an installation script (EggVio_install_tools.sh) is provided for most of the tools that do not require any root privileges for installation (this excludes the assembler, megahit). The user needs to download the annotation database for eggnog mapper and the taxonomic database for contig taxonomic annotation by kaiju separately. The databases were not included in the installation, because

multiple annotation databases are available and the user can choose which one to download themselves. As a guide, we included both steps in the wiki pipeline (https://gitlab.com/R_addict/eggvio/wikis/Download-and-installation-of-EggVio). The scripts for the pipeline are freely available on GitLab (https://gitlab.com/R_addict/eggvio) under the MIT open source license.

2.1 Description of the steps and citation of the tools used in the pipeline

The Eggvio pipeline can be applied to raw metagenomic illumina reads for MAG assembly and refinement in anvio or for functional annotation using a hybrid method (assembly and read annotation). The main steps of the pipeline are represented in Figure 1 in light blue. The first step is quality filtering using trimmomatic (Bolger, Lohse, and Usadel 2014) followed by assembly using megahit (D. Li et al. 2016). The assembled contigs are then renamed using anvio (Eren et al. 2015) and filtered by length (optional) to discard contigs that are too small for binning and gene prediction steps. Read mapping onto the assembly is computed using Bowtie2 (Langdon 2015). If the user is interested in assembling MAGs, the next steps of the pipeline consist in taxonomic annotation of the contigs using kaiju (Menzel, Ng, and Krogh 2016). These data can then be imported into anvio for binning of MAGs and assembly visualization. For functional annotation of the metagenomes, several other steps are needed and are described in the next section with a focus on the algorithm used to learn the e-value threshold for read annotation.

2.2 Description of the learning algorithm to estimate the e-value threshold for read annotation

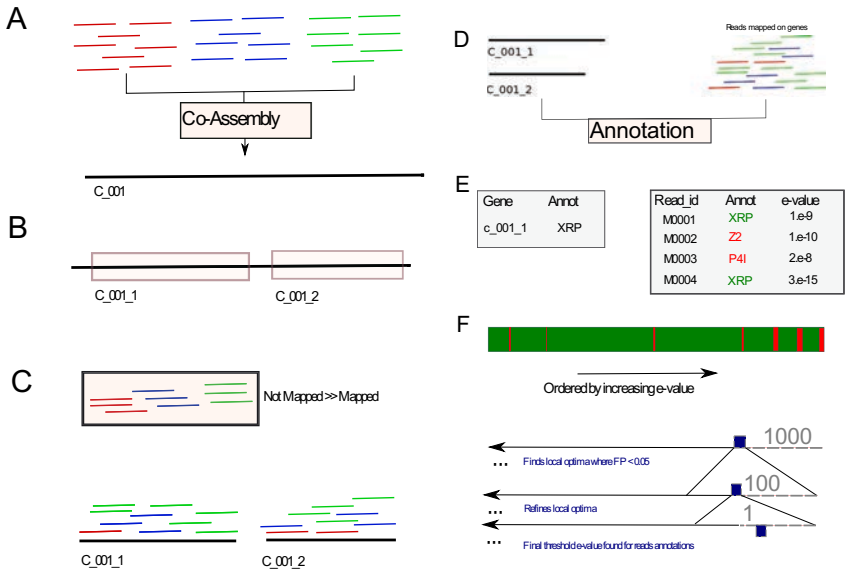


Figure 2 : Summary of the different steps of the EggVio pipeline needed for read annotation. ***A.*** *The reads from different samples from the dataset are co-assembled into contigs* ***B.*** *Genes are detected on the contigs and extracted in fasta.* ***C.*** *Reads mapped on the contigs (previous section) are mapped on the genes and their respective coverage across the dataset is computed.* ***D.*** *Reads which mapped successfully on the genes and the genes themselves are functionally annotated.* ***E.*** *The annotations of the genes (considered as a 'gold standard') are compared to the annotations of their respective mapped reads. If the annotations of reads and genes are different, the read annotation is considered as spurious (False Positive = FP) and if they are identical, then they are considered as true positives (TP)* ***F.*** *The algorithm to learn the threshold is a greedy algorithm based on successive refining steps. It works on a (decreasingly) ordered vector of e-values from read annotation with a corresponding vector returning the information whether the corresponding e-value returned a TP or a FP annotation. The algorithm will compute the percentage of FP starting from the whole dataset and then by removing iteratively the highest 1000 e-value annotations. When it finds the local optima, it will start computing the same statistics, but starting from the interval identified as optimal and refine it by removing iteratively 100 annotations and then for the last refinement step only 1. This e-value returned will then be used to annotate the reads that did not map onto genes with an expected probability of FP<0.05 .*

After co-assembly of the different samples of the dataset (Figure 2 A), gene detection on contigs (Figure 2 B) is carried out using prodigal (Hyatt et al. 2010). The coverage is then computed using Bowtie2 (Figure 2 C) and results are converted into counts using a custom bash script relying on functions from samtools (H. Li et al. 2009) and bedtools (Quinlan and Hall 2010). The genes and the reads which mapped onto them are then functionally annotated (Figure 2 D) by eggnog-mapper (emapper version: emapper-1.0.3-3-g3e22728 emapper) (Huerta-Cepas et al. 2017) using the diamond (Buchfink, Xie, and Huson 2015) mode with the eggnog orthology database (DB version: 4.5.1). These gene annotations and their respective mapped reads are compared to identify false positive (FP) annotations. If the read annotations differ from the genes they are mapped onto (considered as the 'gold standard' annotation), this is considered as a false positive (Figure 2 E). Based on these data, the e-value threshold (E.T.) learning algorithm is used to define a suitable threshold for read annotation where the

percentage of expected FP in the annotations considered as significant would be p-value <= 0.05.

This algorithm is written in R and is presented below.

```
#Variables:
#dataF= a dataframe with zero (FP) or 1(TP) to inform about how good is the
corresponding annotation
#stepN= a numeric value to define the size of the interval to use to compute the FP
#modeU= defines if the number of iterations has to be defined by the size of the
dataset (nrow(dataF)) or fixed to a specific number of steps (10).
#startPos= the starting position (a row number in dataF) from where to start
computing the FP for the first iteration

falsePosFind = function(dataF, stepN, modeU, startPos){
  if(modeU == 'raw'){
    roundNumb <- nrow(dataF) - (nrow(dataF) %% stepN)
    stepMax <- roundNumb / stepN
  }else{
    stepMax <- 10
  }
  res <- data.frame(rowN=rep(NA,stepMax+1), Fpos= rep(NA,stepMax+1))
  for(iter in c(1:stepMax)){
    if(iter == 1){
      res$rowN[1] <- startPos
      res$Fpos[1] <-
length(which(dataF[c(startPos:nrow(dataF)),1]==0))/(nrow(dataF)-startPos+1)
    }
    res$rowN[iter+1] <- (iter*stepN)+startPos
    fpos <-
length(which(dataF[c(((iter*stepN)+startPos):nrow(dataF)),1]==0))/(nrow(dataF)-
((iter*stepN)+startPos))
    if(length(fpos)==0){
      res$Fpos[iter+1] <- 0
    }else{
      res$Fpos[iter+1] <- fpos
    }
  }
  return(res)
}

#Core of the script
#steps defines the size of the intervals for each iteration of the learning
algorithm
steps <- c(1000,100,10,1)
#the dataframe from where to learn(learningData)is composed of two columns: V1=
zero (FP) or 1(TP) to inform about how good is the corresponding annotation and V2=
e-value associated to the annotation.
#The learning dataframe is first ordered by decreasing e-value (V2).
dataForLearn <- ([order(learningData$V2,decreasing = TRUE),]
#pValThresh=0.05 = fraction of FP tolerated
for(stepSize in steps){
  if(stepSize == 1000){
    threshRes <- falsePosFind(dataForLearn,stepSize,'raw',1)
  }else{
    threshRes <- falsePosFind(dataForLearn,stepSize,'default',whereToCut)
  }
   if(pValThresh < min(threshRes$Fpos)){
     errorM <- paste('The p-value threshold set is smaller than the best result
possible to reach for this dataset. Please set it above the following minimum
possible:', min(threshRes$Fpos))
     stop(errorM)
   }
    whereToCut <- threshRes$rowN[which(threshRes$Fpos  < pValThresh)[1]]
   if (stepSize > 1){
      whereToCut<-  whereToCut - stepSize
    }
    #the e-value threshold final result is here
    threshold_seed_ortholog_evalue <- dataForLearn[whereToCut,2]
    exactPval <- round(threshRes$Fpos[which(threshRes$Fpos  < pValThresh)[1]],5)

}
```

This function is designed to find an e-value threshold such that after filtering, the false discovery rate FDR = $\frac{FP}{FP+TP}$ <= 0.05 (where FP = False Positive and TP = True Positive)
At the same time, we would like to minimize the rejection of correct annotations. To meet both criteria, the function orders the dataset by decreasing e-value (the least significant e-value has the row name 1 in the ordered dataset). Then, the script calls the function "FalsePosFind" to compute the FDR with a threshold set every 1000 annotations. Since the data are ordered, the first threshold that meets the criteria FDR < 0.05 will be the best solution, as it will preserve the highest amount of TP. Once this position is found, the loop will iterate the function "FalsePosFind" a second time, but starting from the position located 1000 annotations higher than that initially found. The optimization serves to minimize the amount of TP rejected as false negative (FN) and to refine the threshold such that it is less restrictive by computing FDR every 100 annotations starting 1000 annotations away from the local optima found on the previous iteration. In total, 1000/100 = 10 FDR will be computed and then the best optima (first optima found) will be selected and refined further with intervals of 10 annotations and finally 1 annotation. This e-value threshold will then be used to annotate the reads not mapped on the contigs using eggnog-mapper.

2.3 Benchmarking the EggVio pipeline

Presentation of the dataset used in the study:

To test how our pipeline could enhance the annotation of shallow sequenced datasets, we used one of our in-house datasets from a polluted site bioremediation project (MISS). It consisted of a time series of 30 samples tracking chlorinated compounds in a polluted ground water site. After the injection of organic carbon to induce the biodegradation of chlorinated solvents in the groundwater (C1), three other samples were taken one month apart (C2, C3 and C4). Six replicates were collected for each time point. The biodegradation of chlorinated compounds was evaluated at each time point and qPCR analysis was carried out in order to evaluate the abundance of the pceA gene coding for an enzyme involved in reductive dechlorination of tetrachloroethene (PCE) to trichloroethene (TCE).

2.4 Evaluation of the threshold learning algorithm:

We first used this dataset to evaluate how our threshold learning for read annotation based on assembled data would perform. We computed the threshold for every annotation (gene id, gene name, Kegg orthologs = KO and Gene Onthology = GO) to compare the thresholds returned. We then focused on the KO annotation to test how learning on a subset of the dataset (since we could not assemble the whole dataset) would affect the threshold and FDR estimate. To test this, we randomly subsampled the learning dataset to predict the threshold and then observed how the FRD was affected when computed on the whole dataset.

2.5 Comparison of the sequencing results and the qPCR results on the tracking of the gene pceA

We also used this dataset to compare how the hybrid annotation (genes and reads annotations) would affect the results as compared to gene only annotation using the intermediate results from EggVio. Both annotations (hybrid and genes only) were then normalized by the RPKM method using the R package GenomEnvironR (github: https://gitlab.com/R_addict/genomenvironr). The percentage of sample being annotated as genes and as reads was then investigated for every sample. A NMDS of the samples was carried out for both types of annotation at the KO (Kegg orthologs) level. These sample representations were compared to an NMDS plot of the samples that included chemical data of several chlorinated compound concentrations measured in the water. Given that measurements were missing or below detection level for some of the chlorinated compounds, we only used PCE, TCE and cis-DCE in the analysis.

Finally, we determined whether read annotation using the hybrid EggVio approach would improve classic methods by comparing the abundances of pceA genes observed in the contig assembled data versus the hybrid annotation and the qPCR data. To determine whether this approach could improve the overall significance between biological and chemical data, we correlated gene abundances with environmental variables related to pceA activity (PCE and TCE).

3 Results

3.1 Benchmarking the EggVio pipeline

3.1.1 Summary of the assembly and reads annotation

In total, the assembly recruited only 11% of the reads (915788 reads from 8001127 reads in total) used in the co-assembly and generated over 319458 contigs. The N50 of the assembly was 593 bp of contig length. During the mapping, 100% of the assembly could recruit at least one read of coverage showing that no artefact contig had been generated during this step. In addition, most of the assembly displayed a coverage of 2 minimum (99% of bins of size 10 bp displayed a coverage of 2 or more).

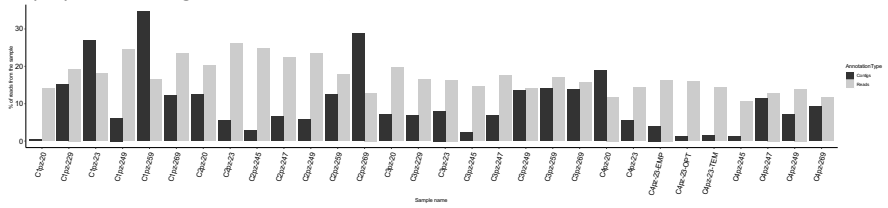


Figure 3: Summary of the percentage of reads annotated successfully (seed eggnog ortholog level) with the hybrid method from EggVio. The annotations derived from the genes predicted on contigs are represented as black bars and the annotations derived from direct read annotation are represented as grey bars.

The number of reads annotated by their annotation in their respective contig assembly was high for some samples (up to 34%), but heterogeneous, with some samples being annotated below 1% (Figure 3). Individual read annotation provided more homogeneous results, with all the samples having more than 10% and up to 26% of their reads annotated (Figure 3). For the hybrid method, the lowest percentage of annotated reads was around 15% and reached up to 65%. On the other hand, the range of possible annotations decreased dramatically when contigs were used. In other words, the annotation variability at the seed eggnog ortholog level (composed of a unique gene and its respective genome taxonomy id) increased significantly for the reads compared to the contigs. The gene annotation returned 18613 possible genes from the contig annotation compared to 408618 different genes (~ 22 times more) for the read annotation.

3.1.2 Evaluation of the read annotation threshold learning algorithm

In order to evaluate how our learning threshold performed, we first ran the algorithm on the fully assembled data. Then, we assessed how random subsampling of the training dataset would impact our learning threshold and its ability to keep the amount of false positive below 0.05. To do this, we subsampled the mapped reads to smaller fractions (respectively 80%, 50 % and 10% of the original dataset) 1000 times and ran the threshold analysis. The threshold results were then compared to the amount of True Positive (TP) and False positive (FP) in the original dataset at different annotation levels (gene ID, gene name, KO and GO, Table 1). An FDR = 0.05 was obtained for all the annotation levels except for Gene ID, where the error was too high with a threshold at FDR = 0.34. By default, the algorithm computes an e-value threshold where the False discovery rate (FDR) is below 5% of False positive (FP), i.e. FDR = 0.05. For each threshold, the number of true positives (TP) is also given. The fraction of correct annotations rejected with this threshold is also given (fraction of correct rejected) as well as the fraction of false annotations successfully rejected (fraction of false rejected). These two fractions are used to determine the sensitivity of the annotation.

*Table 1: Results of the e-value threshold learning algorithm for the MISS dataset. False discovery rate (**FDR**), false positive (**FP**), true positives **TP**, false negatives **FN**, and true negatives (**TN**)*

ANNOTATION	FDR	THRESHOLD E-VALUE	FN	TN	FP	TP	FRACTION OF CORRECT REJECTED	FRACTION OF FALSE REJECTED
GENE ID	0.34	1.8e-27	0.52084	0.47916	0.34001	0.65999	0.6348	0.75634
GENE NAME	0.05	2e-08	0.82506	0.17494	0.04995	0.95005	0.02547	0.09535
KEGG ORTHOLOGY	0.05	3.2e-15	0.88202	0.11798	0.04988	0.95012	0.26222	0.4752
GENE ONTHOLOGY	0.05	3.5e-17	0.83239	0.16761	0.04998	0.95002	0.30443	0.62618

Based on the annotation level, the threshold is variable for the same FDR (0.05). The highest threshold found for an FDR of 0.05 was for gene name annotation (2e-08) and the lowest was for GO terms annotation (Table 1). The vast majority of the annotations rejected are FN (above

80% of the rejected annotations during the learning) (Table 1). This feature is expected since the e-value distributions of the correct and incorrect annotations for KO (Figure 4) overlap. If we compare their medians (dashed lines on Figure 4), the e-value median of the correct annotations is much smaller (<1e-20) than that of the incorrect annotations (>1e-15). As a consequence, the percentage of correct annotations rejected is smaller (~26%) than that of the incorrect annotations (~47%) (Table 1).

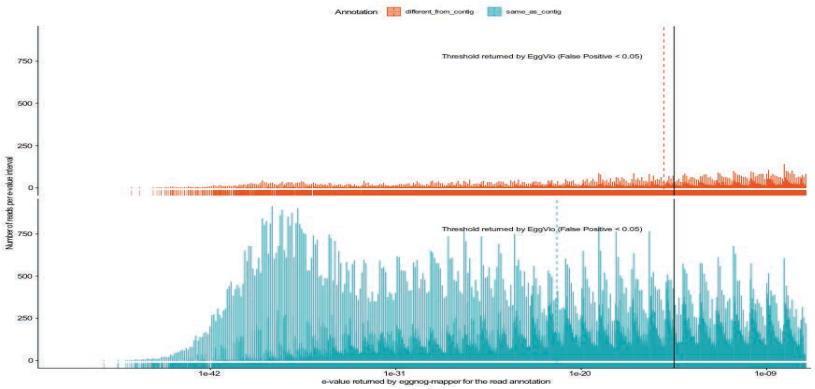


Figure 4 : Histograms of the distribution of e-value for the reads annotations from the reads mapped on the genes predicted from the contigs of the assembly. The vast majority of the read annotations are identical as the ones from the genes (blue green) at the opposite of the spurious annotations (red) different from the genes. The vertical black line represents the threshold returned by the learning algorithm from EggVio where the amount of spurious annotation represents less than 5% from the total number of annotations accepted with this threshold. The dashed vertical lines represent the median e-value of their respective distributions.

We then evaluated the impact of dataset subsampling on the e-value threshold estimate for read annotation and our FDR estimate for the whole dataset. After plotting the results of 1000 subsampling at different percentages of the dataset, we observed that the e-value estimate became more variable at smaller subsampling sizes (Figure 5) and the boxplots tended to deviate from the true estimate toward smaller e-values as they were moved toward the bottom of the y axis (Figure 5).

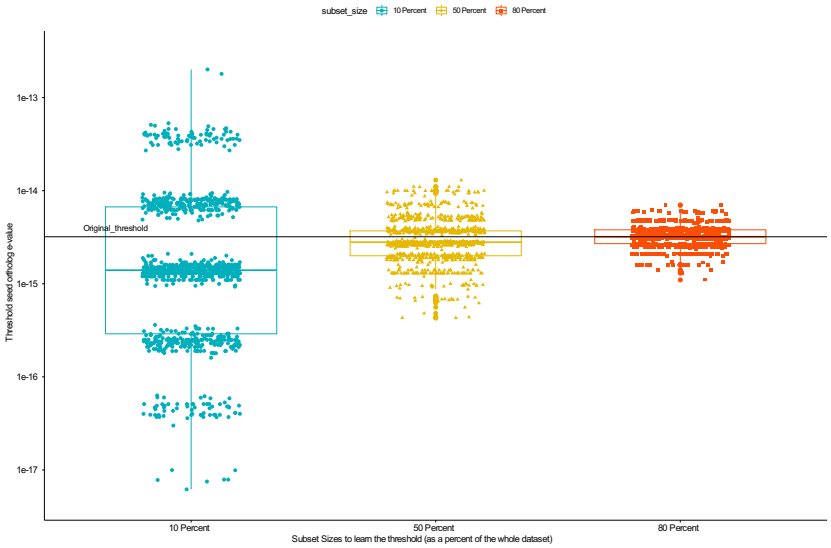


Figure 5: Effect of subsampling on the learning of the e-value for the read annotation. Each dot represents the threshold e-value returned by the algorithm based on its learning on a subset of the annotations (represented on the x axis as a percentage from the total dataset) from the reads mapped on the assembly genes.

This trend was confirmed by the amount of TP observed for each of the e-value thresholds returned for the different subsampling. The second quartile of the boxplots was generally above 0.95 (Figure 6) for the subsampling replicates, indicating that for more than 50% of the subsamples, the threshold returned was enriched with correct annotations and thus showed an FRD <0.05. In addition, the replicates with the lowest amount of TP were always above 0.945, and FRD <0.055 was observed in the worst case during the subsampling experiments (Figure 6).

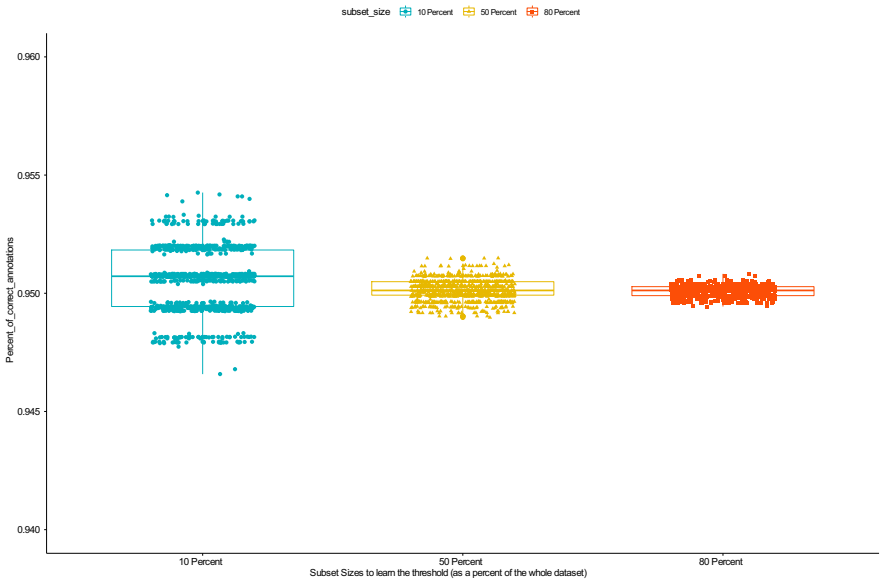


Figure 6 : Effect of subsampling on the amount of true positive (TP) after filtering. Each dot represents the amount of TP detected in the annotation of the reads mapped on the genes (whole dataset) when the learning was performed on a subset of the annotations (represented on the x axis as a percentage from the total dataset) from the reads mapped on the assembly genes.

3.2 The effect of the read annotation on the dataset representability and the detection of pceA genes.

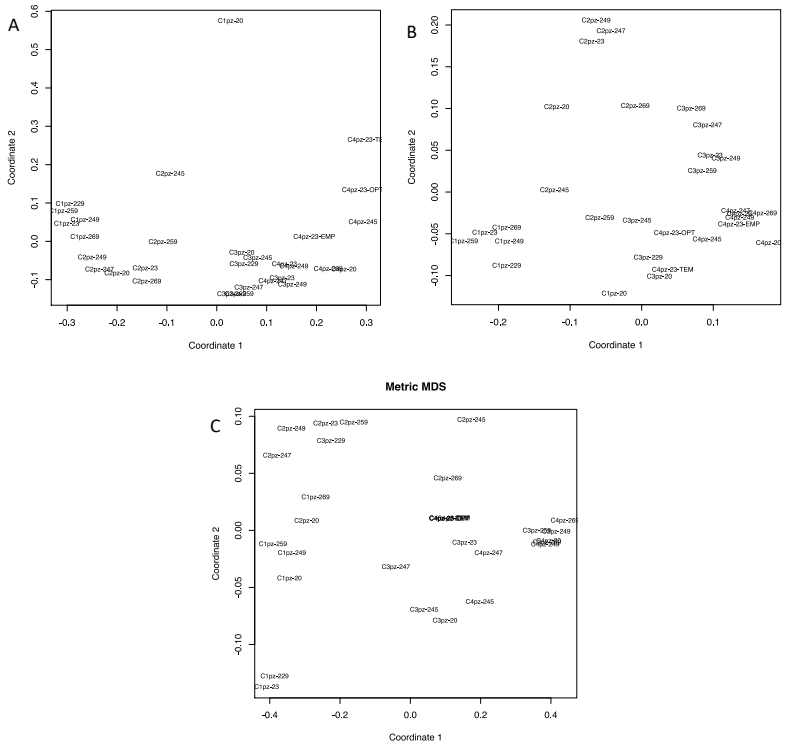


*Figure 7: NMDS of the different samples from the MISS dataset based on their respective normalized annotations (Kegg orthologs) established on the annotation of: **A** = genes predicted from the contigs assembly only **B** = genes predicted from the contigs assembly and annotations of reads not mapped on the assembly. **C.** = NMDS based on the chlorinated compounds measured (PCE, TCE and cis-TCE).*

We investigated how the NMDS representations of the MISS dataset based on the metagenomic data agreed with the NMDS representation based on the chemistry (Figure 7). The NMDS from both types of annotation (assembly and hybrid) clustered more together at a given sampling time (*e.g.*, C1 with samples C1) than between samples from different sampling times, but their grouping at a finer scale showed differences. For example, the sample C1pz-20 was far from all the other samples from the dataset when considering only the assembly annotation (Figure 7 A) but did not display any chemical features on the chemistry NMDS that could explain its position (Figure 7 C). This sample had the lowest annotated coverage (<1%) in the assembly (Figure 3). We also observed that like in the NMDS of the chemistry, the samples from sampling time C4 were more densely clustered together on the NMDS from the hybrid annotation (Figure 7 B) than for the assembly only (Figure 7 A). A last trend to observe is that at the opposite of the chemistry NMDS representation, the samples C3 are totally apart from the samples C2 and separate C2 and C4 samples in the assembly NMDS (Figure 7 A). The NMDS from the hybrid annotation agreed more with the continuum between C2, C3 and C4 and led to C2 and C3 samples to be partially intermixed (Figure 7 B).

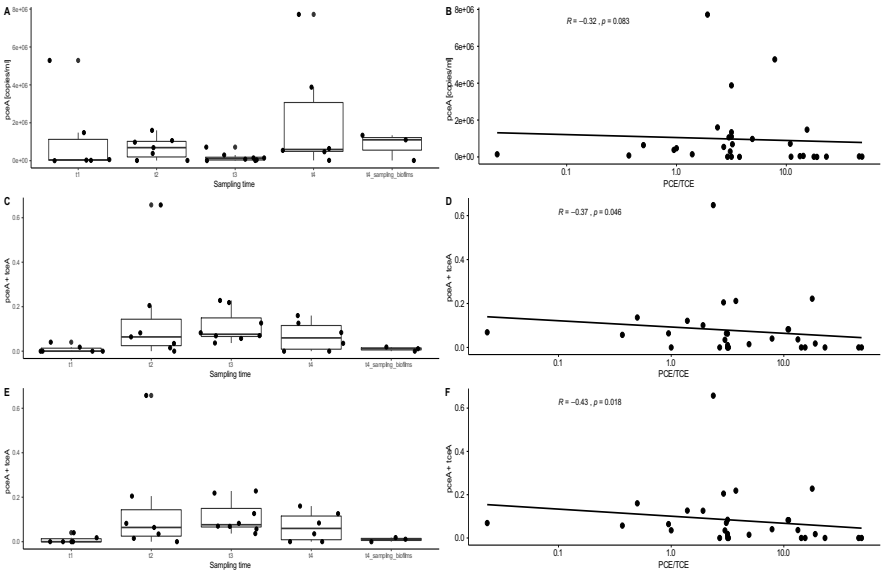


*Figure 8: Comparison of the abundances of the genes pceA detected using qPCR (**A**), sequencing with gene retrieved from the assembly alone (**C**) or with the reads annotation added (**E**). These abundances detected with those methods have then been correlated to the ratio of the target of the gene (PCE) and its end-product (TCE) using the qPCR data (B), the sequencing with gene retrieved from the assembly alone (**D**) or with the reads annotation added (**F**).*

The last result generated is the follow up of the relative abundance of the pceA genes. The abundance of these genes at different sampling times was determined for the different metagenomic annotations and quantified using qPCR (Figure 8 A, C and E). We observed that the sequencing data totally disagreed with the qPCR data (Figure 8 A versus C and E). The maximum amount of pceA quantified using qPCR was detected at time 1 and 4 at the opposite of the metagenomic data where the maximum observed quantities were detected at t2 and t3 (Figure 8 A versus C and E). Interestingly, we could observe that in the metagenomic data, the vast majority of the signal was coming from the assembly (>90%) by comparing the plots from the assembly to the hybrid annotation (Figure 8 C versus E). We could not detect any correlations between pceA gene quantification and PCE or TCE concentrations in the samples (data not shown) but we could detect a significant negative correlation between the quantification of pceA genes in metagenomic data and the ratio of the PCE over TCE concentration (Figure 8 D and F). interestingly, the detected correlation was more significant when computed using the hybrid data quantification (p-val=0.018) instead of the assembly annotations (p-val=0.046) alone (Figure 8 D and F).

4 Discussion

4.1 Assembly missed meaningful information in shallow depth datasets, but can be complemented by rescuing reads

The percentage of reads recruited in the assembly varied widely across samples (Figure 3), with an average of 11% of reads recruited and the amount of gene ids detected in the assembly were 22 times lower than in the read annotations. This is lower than what was observed in a study where full datasets were assembled and 10% to 30% more annotations were retrieved using an assembly free method (Anwar et al. 2019). The undersampling of the dataset was further supported by the NMDS representation that showed that sample C1pz-20 was the most unique sample based on the assembled data (Figure 7 A). However, this sample had the lowest coverage (<1%) in the assembly (Figure 3). After applying the hybrid method, the NMDS could be corrected (Figure 7 B) and showed a representation of the samples that was much more in accordance with the chemical dataset (Figure 7 C). In addition, the hybrid method increased the sensitivity of correlation detection between the relative abundance of pceA genes in the metagenomes and the concentrations of PCE over TCE (Figure 8 E versus F). This negative correlation was not detected in the qPCR data (p-val>0.05). We interpreted this correlation as PCE degradation occurring prior to sampling time rather than an instantaneous measure of the PCE degradation potential of the bacterial community (rather detected by metatranscritomics). Thus, at lower PCE/TCE ratios, a higher abundance of pceA genes could be interpreted as a selection for organisms able to degrade PCE into TCE in the microbial community.

4.2 EggVio E.T. algorithm can accurately quantify the noise added when rescuing reads for annotation

By randomly subsampling the learning dataset to evaluate how the e-value threshold estimate would affect the estimation of the FDR at the KO annotation level, we showed that in the worst case, the FDR would only be impacted by 5% leading to an accuracy (= TP/(FP+TP)) of 0.945 (Figure 6). We also observed that the recall (= TP/(TP+FN) = 1 - FRACTION OF CORRECT REJECTED) was above 69% if we exclude the gene id (seed eggnog ortholog). These estimated performances can be compared to other read annotation tools such as miFaser (Zhu et al. 2018), recently released and based on a custom high quality database using a custom score modeled after the HSSP metric for function transfer between full-length proteins (Schneider, de Daruvar, and Sander 1997). Based on reads generated from their database, they assessed that the accuracy of miFaser could reach 90% and the recall 50% (Zhu et al. 2018). However, since the databases are different and do not use the same criteria, comparisons are difficult to perform.

Concerning the lowest level of annotation accessible by eggnog-mapper (gene id = seed eggnog ortholog), the noise is high and we would, therefore, not recommend using this annotation level for any downstream analyses. The high noise at this annotation level compared to others is likely related to the inclusion of taxonomical information. Gene id annotation is composed of a tax id at the species level and the gene id from its original genome, making it challenging to accurately retrieve annotations from non-assembled reads.

Although not shown in this work, EggVio also integrates kaiju taxonomical annotation of the contigs, which is more accurate since it is carried out on longer reads, thus making the resolution much higher (as for functional annotation) and enables taxonomic analysis of the dominant strains of the population.

5 Conclusion

We showed that EggVio is a flexible pipeline for processing shallow depth sequencing datasets. In addition to assembling reads into contigs, the pipeline can rescue unrecruited reads while adding a predicted amount of noise (FDR = 5%) to the annotation. EggVio improved the correlation between metagenomic and environmental chemistry data. In addition, it removed the artifacts observed in the NMDS when calculated using the assembly alone and showed a more reliable community composition and allow a more robust sample comparison in the actual dataset. Nonetheless, further validation, using a mock community for example or benchmarking it against other possible tools such as miFaser is required to validate the full performance of the pipeline.

6 Bibliography

Anwar, Muhammad Zohaib, Anders Lanzen, Toke Bang-Andreasen, and Carsten Suhr Jacobsen. 2019. “To Assemble or Not to Resemble—A Validated Comparative Metatranscriptomics Workflow (CoMW).” *GigaScience* 8 (8). https://doi.org/10.1093/gigascience/giz096.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” *Journal of Computational Biology* 19 (5): 455–77. https://doi.org/10.1089/cmb.2012.0021.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. 2015. “Fast and Sensitive Protein Alignment Using DIAMOND.” *Nature Methods* 12 (1): 59–60. https://doi.org/10.1038/nmeth.3176.

Delmont, Tom O., and A. Murat Eren. 2018. “Linking Pangenomes and Metagenomes: The Prochlorococcus Metapangenome.” *PeerJ* 6 (January): e4320. https://doi.org/10.7717/peerj.4320.

Eren, A. Murat, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, and Tom O. Delmont. 2015. “Anvi’o: An Advanced Analysis and Visualization Platform for ‘omics Data.” *PeerJ* 3 (October): e1319. https://doi.org/10.7717/peerj.1319.

Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2017. “Fast Genome-Wide Functional Annotation through Orthology Assignment by EggNOG-Mapper.” *Molecular Biology and Evolution* 34 (8): 2115–22. https://doi.org/10.1093/molbev/msx148.

Hyatt, Doug, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. “Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification.” *BMC Bioinformatics* 11 (1): 119. https://doi.org/10.1186/1471-2105-11-119.

Langdon, W. B. 2015. “Performance of Genetic Programming Optimised Bowtie2 on Genome Comparison and Analytic Testing (GCAT) Benchmarks.” *BioData Mining* 8 (1): 1. https://doi.org/10.1186/s13040-014-0034-0.

Li, Dinghua, Ruibang Luo, Chi-Man Liu, Chi-Ming Leung, Hing-Fung Ting, Kunihiko Sadakane, Hiroshi Yamashita, and Tak-Wah Lam. 2016. “MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler Driven by Advanced Methodologies and Community Practices.” *Methods*, Pan-omics analysis of biological data, 102 (June): 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics (Oxford, England)* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Menzel, Peter, Kim Lee Ng, and Anders Krogh. 2016. “Fast and Sensitive Taxonomic Classification for Metagenomics with Kaiju.” *Nature Communications* 7 (April): 11257. https://doi.org/10.1038/ncomms11257.

Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna, Austria. URL http://www.R-project.org/.

Schneider, Reinhard, Antoine de Daruvar, and Chris Sander. 1997. “The HSSP Database of Protein Structure-Sequence Alignments.” *Nucleic Acids Research* 25 (1): 226–30. https://doi.org/10.1093/nar/25.1.226.

Sunagawa, Shinichi, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, et al. 2015. “Structure and Function of the Global Ocean Microbiome.” *Science* 348 (6237). https://doi.org/10.1126/science.1261359.

Wood, Derrick E., and Steven L. Salzberg. 2014. “Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments.” *Genome Biology* 15 (3): R46. https://doi.org/10.1186/gb-2014-15-3-r46.

Zhu, Chengsheng, Maximilian Miller, Srinayani Marpaka, Pavel Vaysberg, Malte C. Rühlemann, Guojun Wu, Femke-Anouska Heinsen, et al. 2018. “Functional Sequencing Read Annotation for High Precision Microbiome Analysis.” *Nucleic Acids Research* 46 (4): e23. https://doi.org/10.1093/nar/gkx1209.

Mis en forme : Bibliographie

Chapter IV - Effect of nutrient enrichment on bacterial interactions in a time series experiment on snow microbial communities

1 Abstract

Based on data collected in the field, we were able to show that organic acids have an effect on bacterial interactions and hypothesized that increases in carbon concentration could lead to a shift from cooperation to competition among microorganisms. In order to validate this hypothesis, we set up a microcosm experiment to study the effect of nutrient concentrations on snow microbial communities at temperatures below zero. Two series of microcosms (n= 4 replicates per treatment) were set up (-5°C in the dark) using Arctic snow: one that was enriched with sodium acetate and the other with water as a control, and destructively sampled over the course of a month (10 sampling times). The evolution of the microbial community over time and its response to the treatments was monitored by measuring cell abundance, taxonomy, functional potential and changes in chemistry. In order to follow changes in microbial interactions, we applied network analysis and studied changes in genetic markers of interactions. Based on our results, we were able to confirm that nutrient addition shifted the interactions from cooperation between fungi and bacteria to competition between bacteria. Co-variance networks showed that the percentage of negative interactions detected in the network of microcosms amended with sodium acetate was almost four times higher than that observed in the network of control microcosms. This work also highlighted the difficulty in identifying genes that participate in bacterial interactions. For example, the definitions of the genes involved in antibiotic metabolism differ between the Gene Onthology and the KEGG databases, which complicates data analysis and renders the interpretation of results more difficult.

2 Introduction

Microbial interactions are important in ecosystems and can affect community members in several ways, notably by impacting metabolism. Co-culture experiments have shown that the metatranscriptomes as well as the proteomes of bacteria differed as compared to monoculture (Molina-Santiago et al. 2017; Hansen et al. 2017; Chignell et al. 2018; Khan et al. 2018; Albers et al. 2018). In turn, such metabolic shifts can impact the bacterial community structure as a whole (Seth and Taga 2014). Bacterial interactions can also create emerging properties and influence the three dimensional organization of the community (Yannarell et al. 2019), enhance the growth rate of collaborating bacteria (Guillonneau et al. 2018) or provide protection against predators (Raghupathi et al. 2018). This is why the study of interactions and how they react to environmental changes is crucial for understanding the bacterial community as a whole. A lot of effort is still needed since the vast majority of the knowledge on bacterial interactions has been obtained from culture experiments (Mitri and Foster 2013). This is mainly due to the challenges related to tracking bacterial interactions in

natural ecosystems (Blasche et al. 2017) given the high level of possible confounding factors (Bergk Pinto et al. 2019).

In a previous field study, we investigated how organic acid concentrations impact bacterial interactions in Arctic snow microbial communities (Bergk Pinto et al. 2019). To study this, we used a dual approach to track bacterial interactions in snow sampled as a time series in Svalbard. We tracked genes that were considered as proxies of bacterial collaboration (plasmid backbone genes) or bacterial competition (antibiotic resistance genes) in snow metagenomes and metatranscriptomes. To strengthen this first analysis, we also used co-variance networks based on 16 rRNA gene sequencing data approaches to support our results. We observed that an increase in organic acids in snow was positively correlated to an increase in the diversity as well as to the total amount of antibiotic resistance genes in the snow metatranscriptomes. We also observed that a higher diversity of plasmid backbone genes was present in samples with lower organic acid concentrations. We observed a significantly higher density of positive co-variances in the low organic acid network than in the high organic acid one. Based on our results, we hypothesized that organic acid concentrations drove bacterial interactions in the arctic snow bacterial community, with a potentially significant influence on the snow ecosystem as a whole.

In order to validate our hypothesis, we tested the effect of organic acid concentrations on snow microbial communities in a microcosm experiment. Here, we present the results of the evolution of an Arctic snow microbial community amended with sodium acetate and compared the results to a control time series. In order to track microbial interactions, we applied the dual approach developed for the analysis of field data (Bergk-Pinto et al., 2019). A machine learning approach was also applied to identify the metabolic processes that were the most correlated to the nutrient levels in the snow in our metagenomes. We expected to retrieve metabolic pathways related to competition (e.g. antibiotic) or to cooperation (e.g gene transfer) if microbial interactions were important in the response to changes in nutrient concentration.

3 Material and methods

3.1.1 Snow

The snow used during the microcosm experiment was collected in Ny-Ålesund (Svalbard, Norway, 78°56'N, 11°52'E) during the month of March, 2012. A freshly fallen surface snow layer was collected as described in Larose et al. (2010) using sterile sampling bags and protective equipment. The snow was then shipped back to France and stored at -15C.

3.1.2 Microcosm set up and sampling

The microcosms (n=88) were prepared in a cold room tempered at -15°C. Before transferring 300 g (+-10g) of snow into washed and autoclave-sterilized 2L microcosm jars, snow was disaggregated in sterile Whirl-pak™ bags using a hammer and homogenized in two large polystyrene boxes coated with sterile sampling bags. Half of the microcosms were amended with 3000 ppb of a sodium acetate solution (1 ml of a solution prepared by diluting 78 mg of sodium acetate (Merck) into 60 ml of milliQ water that was filtered on 0.22 µm) and the other

half consisted of wet controls (CH) to which 1 ml of miliQ water was added. Control and spiked microcosms were homogenized by mixing with a sterilized spatula for roughly 10 seconds each. The homogenization method was tested by adding 1 ml of crystal violet stain to an extra microcosm. The jars were stored in the dark at -5 °C until sampling.

Samples were destructively sampled over a 3-week period at two to three days intervals. At each sampling time, 4 amended microcosms and their respective paired controls were randomly chosen by using the R command sample (). The selected microcosms were left to melt at room temperature prior to filtering. For biological samples, filtering was performed onto sterile 0.22 µM 47 mm filters (Millipore) using a sterile filtration unit (Nalge Nunc International Corporation) and filters were stored in Eppendorf tubes at -20°C for further analysis. For chemical analyses, 10 ml of melted snow was filtered using a Nalgene sterile syringe and a 0.22 µM filter (Millipore) and bottles containing the filtered water were stored at 4°C prior to measurements.

3.1.3 Chemical analysis

Chemistry samples were analyzed as detailed in Bergk Pinto et al. (2019). Briefly, organic acids (acetate, oxalate, succinate, lactate and formate) and ions (sodium, ammonium, nitrate, potassium, magnesium, calcium, chlore and sulfate) were analyzed using conductivity-suppressed ion chromatography (a Dionex ICS 3000© apparatus and a Dionex AS40©) at IGE laboratory (Grenoble, France). Some organic acids (succinate, lactate and formate) were always below the detection limit (1ppb) and were discarded from the analyses for this reason. The chemical measurements of this study can be found in supplementary material (see annex pp. 122-125).

3.1.4 Molecular analysis

DNA was extracted from filters using the DNeasy PowerWater Kit (Qiagen) following the manufacturer's instructions. DNA was quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific). 16S rRNA gene copies were quantified in the snow using qPCR. The samples were amplified in duplicates using a set of primers designed to amplify 314F until 534R. For each run, a standard curve was established using the following standards dilutions: starting from 10^8 gene copies until 10^2 gene copies by doing several successive dilutions by a factor of 10X.

3.1.5 DNA sequencing

DNA samples extracted from the 88 samples were processed to build 16S rRNA gene libraries for sequencing as described in Bergk Pinto et al. 2019. Briefly, the V3-V4 regions of the 16S rRNA genes were amplified by PCR and Libraries for 16S rRNA gene sequencing were prepared using the 16S rRNA gene Library Preparation Workflow recommended by Illumina. Libraries for ITS sequencing were prepared by amplifying the region of the fungal ITS2 gene using the following primer set from Taylor et al. (2016): ILL_5.8S_Fun 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAACTTTYRRCAAYGGATCWCT 3' as the forward primer sequence, and ILL_ITS4_Fun 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCCTCCGCTTATTGATATGCTTAART 3' as

the reverse primer sequence. The resulting libraries were then prepared by following the Library preparation workflow recommended by Illumina. Paired end sequencing was then carried out on a MiSeq sequencer (Illumina) at the laboratory in Lyon. Metagenomic libraries were prepared for the 88 samples from both times series using the Nextera kit (Illumina) following the manufacturer's instructions. Paired end sequencing was then carried out on a MiSeq sequencer (Illumina) at the laboratory in Lyon.

3.2 Bioinformatics for quality filtering and data processing

3.2.1 Quality filtering, amplicon sequence retrieval and taxonomy annotation for the 16S rRNA gene and ITS sequencing

The primers used for 16S rRNA genes and ITS amplification were removed from the reads using cutadapt (Martin 2011). The dada2 pipeline (v 1.12) (Callahan et al. 2016, 2) was used for quality filtering, trimming and identification of the amplicon sequence variants (ASV). The parameters were adapted to fit specificities of each sequencing type (ITS versus 16S rRNA genes). The two R scripts are included in the supplementary materials and provide details on the parameter values used for processing. The abundance tables were normalized by using the cumulative sum scaling (CSS) from the R package MetagenomeSeq (Paulson et al. 2013).

3.2.2 Metagenomes annotation

In order to annotate the metagenomes, samples were co-assembled and annotated using a custom pipeline called EggVio. This pipeline carries out all the steps in one single shot using several tools in a fully automated way. The quality filtering was done by using trimmomatic (Bolger, Lohse, and Usadel 2014) to remove the adapters from Nextera's kit and quality filter the remaining reads. Co-assembly was then carried out using megahit (Li et al. 2016). To compute the coverage of the assembly, reads from the samples were mapped back using bowtie2 (Langdon 2015). These results were then imported into anvi'o (Eren et al. 2015) to visualize the assembled genomes as well as their taxonomy. Gene identification of the assembled contigs was carried out using prodigal (Hyatt et al. 2010). Genes were functionally annotated using EggNOG-Mapper (Huerta-Cepas et al., 2017), based on eggNOG orthology data (Huerta-Cepas et al., 2016), using the default parameters in diamond mode. The sequence searches were performed using diamond (Buchfink et al., 2015). In addition, the pipeline used a read annotation strategy on the reads that could not be mapped onto contigs in order to improve the exhaustiveness of annotation. A custom R script merged the count coverage returned for contigs and read annotations to generate a final abundance table. These data were then normalized using the RPKM normalization function included in our package GenomEnvironR.

3.3 Bioinformatics for data analyses

3.3.1 Boost regression on KOs

In order to test whether our hypothesis about bacterial interactions was relevant in the bacterial communities of our microcosms, we used gradient boost machine learning to retrieved the KEGG orthologs (KOs) that were the most correlated to the nutrient levels in the snow. If bacterial interactions were significant in the response of microbial communities to

organic enrichment of the snow, metabolic pathways related to competition (e.g. antibiotic) or cooperation (e.g gene transfer) should be identified using this technique. Machine learning was carried out using a ratio between the snow acetate concentrations (main carbon source in our system) and the ammonium concentrations (main nitrogen source in our system). Briefly, the dataset was randomly split into a training and testing set and one thousand permutations of the model was cross-validated to determine parameters that achieved the lowest mean squared error (MSE). The optimized model (n trees, interaction depth of n) was run on the entire dataset to calculate the relative influence of each KO on ratios between the acetate and ammonium concentrations in the snow, and the mean and standard deviation for relative influence per KO over one thousand permutations was determined. To be considered as a KO of interest, the average influence threshold was set to 1%.

3.3.1 Data visualization and differential abundance analysis

Data visualization of normalized metagenomic data or normalized 16S rRNA gene and ITS gene sequencing data was carried out using the R package vegan (Dixon 2003). Raw count data tables were first transformed into gene name tables by summing up all the genes related to a certain gene name. EdgeR (Robinson, McCarthy, and Smyth 2010) was used to test which sets of KOs were more abundant in one of the two time series (p-value < 0.05 (95% confidence interval)).

3.3.1.1 Mining of genes related to microbial interactions using GO terms and KEGG orthology

We mined the gene names retrieved as significantly more abundant in one of the two time series to assess whether they were related to antibiotics, plasmids or acetate metabolism using specific sets of GO terms (see supp. mat. in the annexes pp. 126-127) associated with the retrieved annotation from eggnog mapper. These sets of genes were then annotated using KEGG orthology. To determine which pathways were significantly enriched in the data sets, the original pool of genes in each time series were randomly sampled (1000 times) to build random gene sets that could be compared to those retrieved by edgeR. The KEGG annotation for all the genes was recorded. For each pathway, the relative abundance of the genes retrieved in the original dataset was compared to their distribution in the random gene sets generated. The pathway was then considered significantly enriched if the observed relative abundance was bigger than 95% of the random gene sets. The resulting p-value was computed as: p-val = $1 - \frac{N_{obs\,random}}{1000}$ where $N_{obs\,random}$ is the number of random genes sets where the relative abundance of the genes related to the tested pathway is smaller than the one observed for the original gene set.

3.3.2 Network construction from the 16S rRNA gene and ITS data

Based on the ASVs of Bacteria and Fungi, networks were constructed for each time series separately. Prior to computing the networks, a filtering step was carried out using the R package GenomEnvironR to remove all ASVs present in less than 22 samples (50% of the samples) in each time series. The two groups of ASVs (129 in the water control series and 123 in amended acetate series) were compared and only the ASVs found in both groups were

used for the network calculations, resulting in 112 ASVs. In this way, the networks from both time series were built with the same ASVs. Then, the networks were computed using eLSA (Xia et al. 2013), a tool to compute local similarity scores on time series with replicates. The results were then filtered by p-value threshold set to 0.00001. The remaining significant local similarity scores were then imported into R to build networks and compare them for each data type. The extent to which the networks represented the core community was assessed by summing up the sequences affiliated to the core ASVs and dividing them by the total depth of sequencing from their respective samples (ITS and 16SrRNA sequencing) to generate relative abundances. The change in relative abundance of the core microbial community throughout the microcosm experiment was then monitored.

The percentage of positive and negative co-variance (LSA) was then retrieved for each of the networks. The co-variances were then subdivided based on the taxonomy of the two ASVs linked by co-variance (e.g. a bacterial ASV interacting with a fungal ASV). We interpreted positive co-variance as being a surrogate of positive interactions (cooperation) and negative ones as surrogates of negative interactions (competition). To test whether the observed trends in terms of dominant type of co-variance (interactions) were significant, we generated subnetworks for each time series by randomly subsampling a fixed number N of samples at each sampling time and then computing the networks using the same method as for the original ones. To test different subsampling sizes, we used N= 3, then 2 and finally only one sample per sampling time. We set the p-value to 0.001 as a cutoff for significance, since we measured the co-variances using a smaller number of replicates per time point. We generated 100 networks per time series and for each subsampling size to determine a confidence interval for the observed trends to confirm their significance.

4 Results

4.1 Chemistry of the microcosm time series

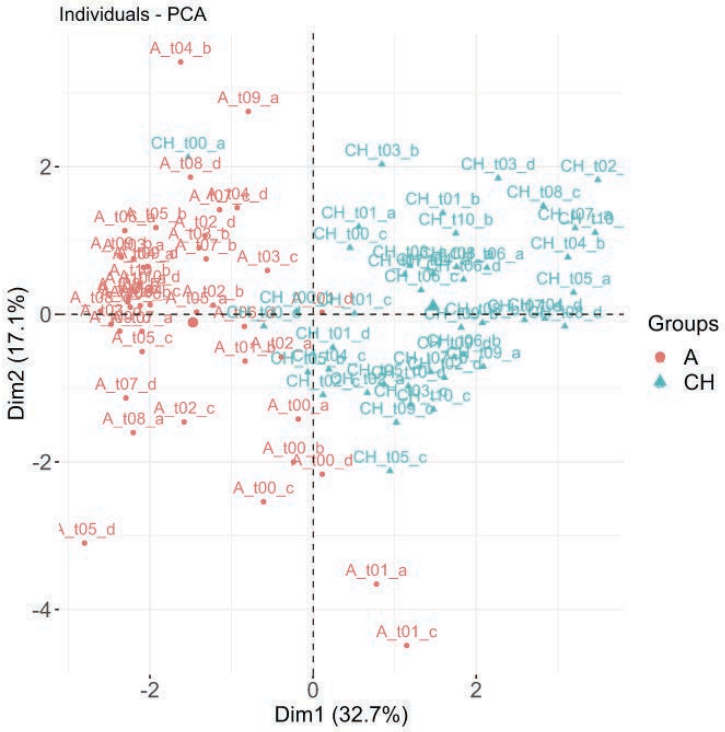
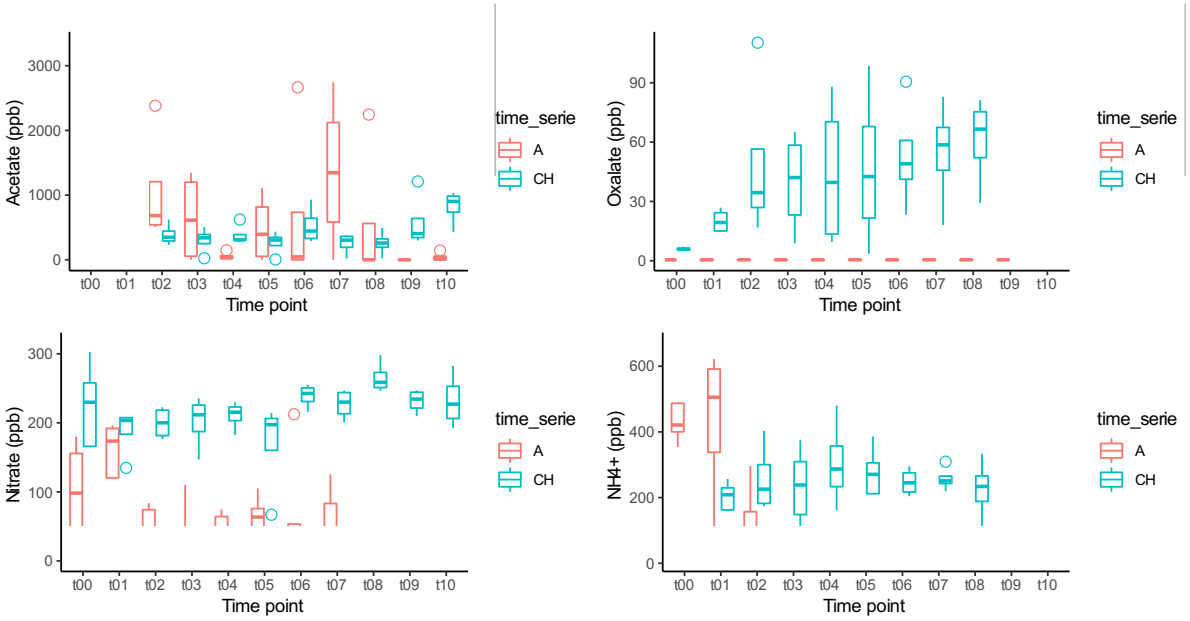


Figure 1: Principal component analysis biplot from the chemistry of the samples used in this study. Green triangles (water control samples) and red dots (acetate amended samples) are represented based on their respective projections. The sample codes next to the symbols are: time series id (A = acetate, CH = water) followed by the time of sampling (t0 to t10) and finishes by a letter to identify each replicate individually (a to d).

Principle component analysis (PCoA) of the chemical data showed that almost 50% of the total variability of the dataset was represented by the first two axes (Figure 1). We observed a separation between the time series along the first axis (Dim 1) of the PCA. The variables that contributed the most to this axis were nitrate and oxalate (above 20%), followed by sodium, ammonium and magnesium (above 10%). Acetate contribution to this separation was low (around 1%). The chemical variables that contributed the most to the second axis (Dim2) were chloride and sulfate (above 20%), followed by acetate, calcium and ammonium (above 10%). The t0 samples from both time series were represented at the center of the plot by their first axis coordinates but already separated along the second axis.



ic

acids (**A**: Acetate, **B**: Oxalate) and nitrogen (**C**: Nitrate, **D**: Ammonium) in the snow samples from both time series (CH: water control, A: acetate amended) of the study. For each sampling time, the boxplot represents the variation of measures reported for the four replicates sampled at each time and for each time series. Circles above boxplots represent outliers.

The evolution of organic carbon and nitrogen species over the course of the experiment are presented in figure 2. Each time series followed distinct patterns. At t_0, the acetate amendment was detected in the acetate time series (average 2000 ppb), while acetate concentrations were close to the detection limit in the water controls (Figure 2 A). The acetate concentrations in the acetate microcosms decreased throughout the experiment, with the exception of t_{07} where it increases. In the water control time series, the acetate concentrations increased with some slight fluctuations until the end of the experiment (with concentration value ranging from 20 ppb up to 1000 ppb). Oxalate concentrations also increased significantly in the water control, from 5 ppb to 50 ppb, but were undetected in the acetate time series (Figure 2 B).

Nitrate and ammonium followed similar trends for their respective time series. In the acetate series, the concentration of nitrate and ammonium decreased significantly (starting at 100 ppb for nitrate and around 500 ppb for ammonium to reach concentration below 1 ppb, the detection limit) (Figure 2 C and D). This trend was the strongest for ammonium, where measures were below detection limits after t_{02} (Figure 2 C). Nitrate was also under the detection limit after t_{07}. In the water control, nitrate concentrations did not vary significantly and were around 200 ppb. However, the ammonium concentration increased significantly during the experiment, from 20 ppb to a final concentration of about 200 ppb (Figure 2 C and D). Ammonium concentrations differed between the two time series at t_0, with higher concentrations in the acetate time series, despite no nitrogen additions (Figure 2 D).

4.2 Metabolic changes observed in the snow microbial community

4.2.1 Boost regression

The boost regression analysis retrieved KOs (KEGG Orthologs) identified as being the best predictors of the C:N ratio in our microcosms. We chose the C:N ratio based on the PcoA results that showed that carbon (acetate and oxalate) and nitrogen (ammonium and nitrate) contributed the most to explaining the variability in our control versus acetate amended microcosms (Figure 1).
KO genes with the highest influence during training (mean influence above 1%) were mainly classified in three categories: transporters, secondary metabolite biosynthesis and sugar metabolism (annex pp. 128-129). KOs linked to the biosynthesis of antibiotics (K01568 and K14681) as well as one KO related to secretion system (K02674) were also detected. In addition, 3 out of 18 KOs (16%) were related to microbial competition (Table 1). No single KO had a consistent, high impact on the boost regression validation. KOs influenced the prediction by 12% in the best validations and 2% in the worse (see annex p. 129).

Table 1: Definitions and pathways of the KOs with the highest influence and related to bacterial interactions retrieved by the boost regression (see figure 7 and table 1 in the annex pp. 128-129 for detailed influence and exhaustive results of the boost regression).

Kegg ortholog id	Name	Definition	Pathway
K01568	PDC, pdc	pyruvate decarboxylase [EC:4.1.1.1]	ko00010 Glycolysis / Gluconeogenesis ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites ko01130 Biosynthesis of antibiotics
K14681	argHA	argininosuccinate lyase / amino-acid N-acetyltransferase [EC:4.3.2.1 2.3.1.1]	ko00220 Arginine biosynthesis ko00250 Alanine, aspartate and glutamate metabolism ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites ko01130 Biosynthesis of antibiotics ko01210 2-Oxocarboxylic acid metabolism ko01230 Biosynthesis of amino acids
K02674	pilY1	type IV pilus assembly protein PilY1	Secretion system Bacterial motility proteins

4.2.2 Differential abundance of genes between both time series

Boost regression was used to identify variables that were predictors for the observed changes in our microcosms, but this approach cannot be used to track dynamic shifts in gene abundance during the experiment. The metagenomes of both time series shared a high number of genes (70% of the genes could be observed in at least 1 sample from each time series). We used edgeR to retrieve the genes that were statistically significantly more abundant in one of the two time series to follow metabolic changes between the acetate amended microcosms and the water control microcosms. 596 genes were returned, with 422 gene names more abundant in the water control (referred to as the water gene set) and 174 more abundant in the acetate amended time series (referred to as the acetate gene set). The absolute values of the observed Log2 fold changes (LFC) were all comprised between 0,2 and 2,89. However, after p-value correction, none of these were found to be significant.

a. Gene onthology (GO terms)

The functional annotations of the 596 gene names that were significantly more abundant in one of the two time series were analyzed using GO terms. No gene names related to plasmids (proxy for collaboration) were significantly more abundant in either of the two time series. Two genes related to antibiotics (AMRA and AMRB) were returned as significantly (p-value < 0,05) more abundant in the water time series (logFC close to 1,3 for both genes) before p-value correction. Three other genes related to antibiotics (using the GO term "response to antibiotics") were also retrieved as being significantly more abundant in the acetate time series. One gene name related to acetate metabolism (FG00176.1), coding for isocitrate lyase, was returned as being more abundant in the acetate time series (logFC = - 0,84).

b. KEGG

The functional annotations of the 596 gene names that were significantly more abundant in one of the two time series were also analyzed using KEGG. A total of 319 different pathways were significantly more abundant in the water time series, while 161 pathways were more abundant in the acetate treatment and 140 pathways were shared between both gene sets. Some of the pathways retrieved during this step, such as the biosynthesis of secondary metabolites, were similar to the ones that were returned by the boost regression. Among the shared pathways, antibiotic biosynthesis (map01130) was detected in both time series, with 33 annotated genes for the water set and 26 genes for the acetate set. This represented 7.8% of the water gene set (33/422) and 12.1% of the genes from the acetate gene set (21/174). However, the likelihood that this pathway was significantly more abundant was only confirmed for the acetate gene set (likelihood = 0.035) but not for the water gene set (likelihood = 0.647).

In total, 17 KEGG pathways were returned as being enriched in the gene set from the water control time series and 23 KEGG pathways for the acetate gene set. The complete list of pathways retrieved for both gene set is available in the annex (pp. 130-143) of this thesis. The pathways enriched in the water gene set were related to amino acid metabolism (tryptophan and histidine), pyruvate metabolism, platinum drug resistance and two-component system. This last pathway was also detected in the acetate gene set but its likelihood of being enriched

was close to significant (likelihood = 0.053). Concerning bacterial interactions, pathways related to siderophore metabolism were only identified in the water gene set, but its likelihood of being enriched compared to a random distribution was not significant. We also detected pathways related to fungi (yeast meiosis and yeast autophagy) that were close to significance in terms of likelihood (respectively 0.073 and 0.059).

Table 2: pathways returned as significantly enriched in the water control gene set. The first column describes the pathway id, its name and, between brackets, the number of KOs retrieved for this particular pathway. The second column gives the total number of genes identified in the water gene set with one or more KOs related to this particular pathway. The third column gives, as a fraction, the number of genes retrieved as being part of this pathway among all the genes from the water gene set (422 genes). The last column shows the likelihood computed by random distribution method (see material and method) to detect if the number of genes related to this pathway was enriched in the gene set compared to its relative abundance among the 1000 random gene sets generated.

PATHWAY DEFINITION	PATHWAY ID	NGENES PER PATHWAY	FRACTION GENES ENRICHED	LIKELIHOOD
BIOSYNTHESIS OF SECONDARY METABOLITES (48)	map01110	45.00	0.107	0.785
BIOSYNTHESIS OF ANTIBIOTICS (29)	map01130	33.00	0.078	0.647
TWO-COMPONENT SYSTEM (40)	map02020	27.00	0.064	0.022
PYRUVATE METABOLISM (14)	map00620	15.00	0.036	0.021
ABC TRANSPORTERS (45)	map02010	14.00	0.033	0.664
ABC TRANSPORTERS (45)	map02010	14.00	0.033	0.664
QUORUM SENSING (17)	map02024	14.00	0.033	0.186
TRYPTOPHAN METABOLISM (7)	map00380	11.00	0.026	0.042
GLYCEROLIPID METABOLISM (4)	map00561	8.00	0.019	0.023
CHLOROALKANE AND CHLOROALKENE DEGRADATION (5)	map00625	8.00	0.019	0.013
HISTIDINE METABOLISM (3)	map00340	7.00	0.017	0.016
PLATINUM DRUG RESISTANCE (4)	map01524	7.00	0.017	0.011
CITRATE CYCLE (TCA CYCLE) (8)	map00020	7.00	0.017	0.103
MEIOSIS - YEAST (7)	map04113	7.00	0.017	0.059
BETA-LACTAM RESISTANCE (10)	map01501	7.00	0.017	0.129
AUTOPHAGY - YEAST (4)	map04138	5.00	0.012	0.073
BACTERIAL SECRETION SYSTEM (4)	map03070	3.00	0.007	0.793
BIOSYNTHESIS OF SIDEROPHORE GROUP NONRIBOSOMAL PEPTIDES (5)	map01053	1.00	0.002	0.575

The pathways enriched in the acetate gene set were linked to antibiotics (including prodigiosin and ansamycin biosynthesis), fatty acids, carbon fixation and other metabolisms, including geraniol degradation. We also observed several other pathways related to bacterial interactions (quorum sensing, bacterial chemotaxis and biofilm formation). These pathways were also present in the water gene set, but their respective likelihood of being enriched compared to a random distribution was not significant (for detailed likelihood see the annex pp. 130-143). Some pathways of interest were present in both gene sets but with a non-significant likelihood such as the secretion system, the ABC transporter or the ATC cycle pathways.

Table 3: Pathways determined as significantly enriched in the acetate gene set. The first column describes the pathway id, its name and, between brackets, the number of KO retrieved for this particular pathway. The second column gives the total number of genes identified in the acetate gene set with one or more KO related to this particular pathway. The third column gives, as a fraction, the number of genes retrieved as being part of this pathway among all the genes from the acetate gene set (174 genes). The last column shows the likelihood computed by random distribution method (see material and method) to detect if the number of genes related to this pathway was enriched in the gene set compared to its relative abundance among the 1000 random gene sets generated.

PATHWAY DEFINITION	PATHWAY ID	NGENES PER PATHWAY	FRACTION GENES ENRICHED	LIKELIHOOD
BIOSYNTHESIS OF SECONDARY METABOLITES (38)	map01110	32	0.184	0.001
BIOSYNTHESIS OF ANTIBIOTICS (26)	map01130	21	0.121	0.035
FATTY ACID METABOLISM (7)	map01212	12	0.069	0.001
TWO-COMPONENT SYSTEM (15)	map02020	12	0.069	0.053
CARBON METABOLISM (15)	map01200	11	0.063	0.134
ABC TRANSPORTERS (28)	map02010	11	0.063	0.066
QUORUM SENSING (11)	map02024	10	0.057	0.025
FATTY ACID DEGRADATION (7)	map00071	9	0.052	0.006
BUTANOATE METABOLISM (10)	map00650	8	0.046	0.026
FATTY ACID BIOSYNTHESIS (3)	map00061	8	0.046	0.003
BENZOATE DEGRADATION (7)	map00362	7	0.04	0.017
CARBON FIXATION PATHWAYS IN PROKARYOTES (7)	map00720	5	0.029	0.042
PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS (10)	map00400	5	0.029	0.001
BACTERIAL CHEMOTAXIS (7)	map02030	5	0.029	0.011
ASCORBATE AND ALDARATE METABOLISM (4)	map00053	4	0.023	0.036
BIOTIN METABOLISM (1)	map00780	4	0.023	0.049
GERANIOL DEGRADATION (3)	map00281	4	0.023	0.018
PRODIGIOSIN BIOSYNTHESIS (1)	map00333	4	0.023	0.022
BIOFILM FORMATION - VIBRIO CHOLERAE (4)	map05111	4	0.023	0.028
INOSITOL PHOSPHATE METABOLISM (4)	map00562	3	0.017	0.047
CARBON FIXATION IN PHOTOSYNTHETIC ORGANISMS (2)	map00710	3	0.017	0.042
BIOSYNTHESIS OF ANSAMYCINS (1)	map01051	2	0.011	0.031
BIOSYNTHESIS OF VARIOUS SECONDARY METABOLITES - PART 2 (2)	map00998	2	0.011	0.029
CITRATE CYCLE (TCA CYCLE) (3)	map00020	2	0.011	0.451
BETA-LACTAM RESISTANCE (2)	map01501	2	0.011	0.51
BACTERIAL SECRETION SYSTEM (2)	map03070	2	0.011	0.503
FLAGELLAR ASSEMBLY (1)	map02040	1	0.006	0.479
CELL CYCLE - YEAST (1)	map04111	1	0.006	0.754

4.3 Snow microbial dynamics, composition and interactions using co-variance networks comparison

PCoA analysis of the bacterial community compositions did not show any clear separation among the samples and explained less than 10% of the variability along the first two axes. Similar results were obtained for the fungal community based on PCoA analysis of the normalized ITS data (for figures, see annex p. 121).

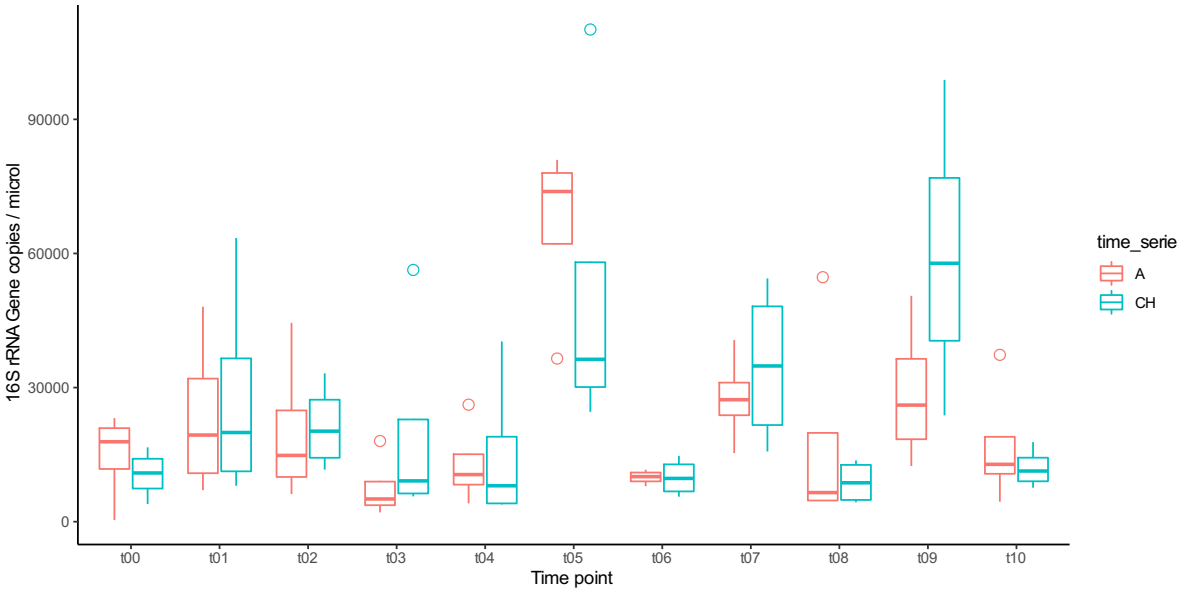


Figure 3: This figure represents boxplots of the absolute quantification of 16S rRNA genes (in copies/microliter) using qPCR for the replicates of the snow microcosms at each sampling time (2-3 days between each sampling). The boxplots in red represent the estimate of dispersion of the replicates from the acetate amended time series and the blue one represents the replicates of the water (control) time series. The circles represent possible outliers.

Based on qPCR analysis of 16S rRNA genes, control and acetate amended microcosms showed similar trends, with an increase in copy number after t1 and reached their highest densities at t05 and t09 (Figure 3). At t05, 16S rRNA gene copy numbers were significantly higher than those in the water control.

4.3.1 Bacterial taxonomy assessed by 16S rRNA gene sequencing and fungal taxonomy assessed by ITS sequencing

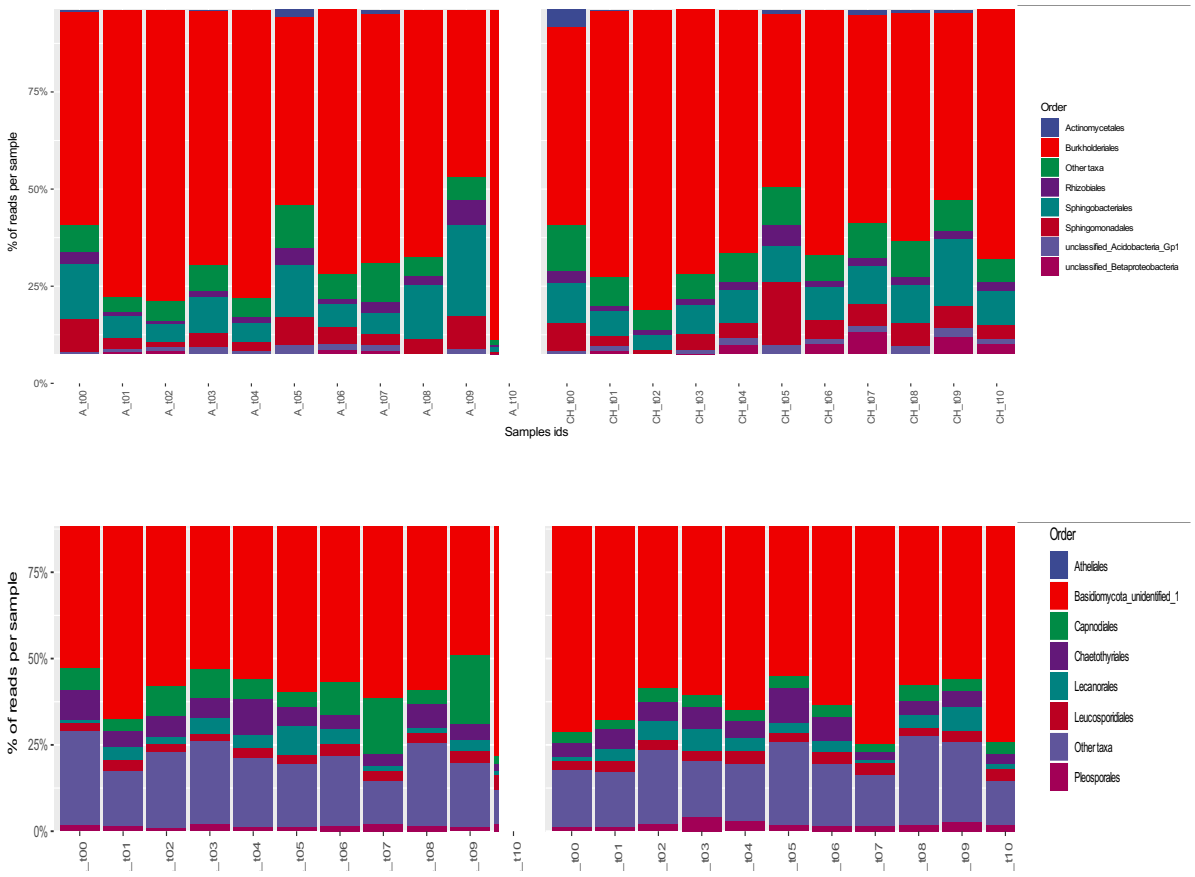


*Figure 4: **A**: Bar plot showing the relative abundance of the most abundant bacterial orders observed across both microcosm time series. The taxonomy is based on the classification of the ASV from the 16S rRNA sequencing data using RDP classifier. The minor bacterial orders represented by a low abundance of sequences in the different samples have been summed up and termed "Other taxa". **B**: Bar plot showing the relative abundance of the most abundant fungal orders observed across both microcosm time series. The taxonomy is based on the classification of the ASV from the ITS sequencing data using RDP classifier. The minor fungal orders represented by a low abundance of sequences in the different samples have been summed up and termed "Other taxa".*

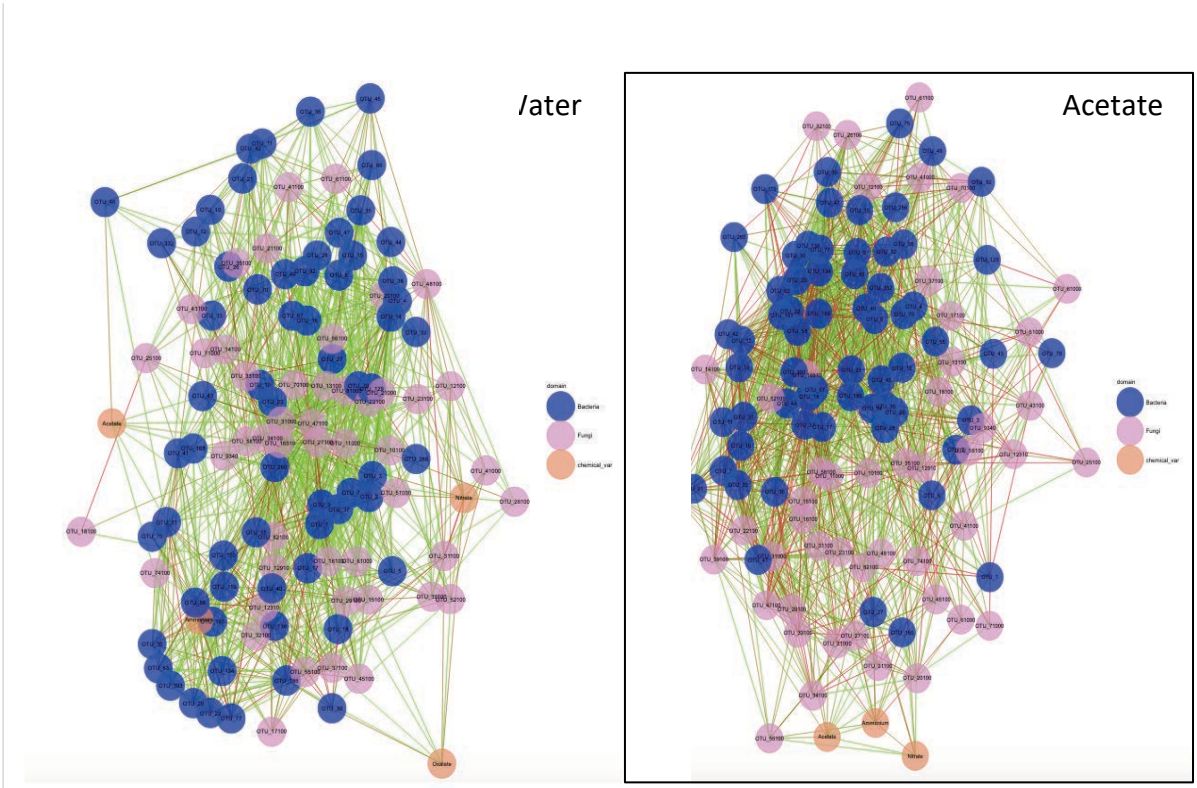
We looked at the taxonomy and relative abundances of the different ASVs. The bacterial communities from the snow were stable across time and treatments. They were dominated by *Betaproteobacteria* of the *Burkholderiales* order and the majority of them could not be classified at the genus level, but belonged to the family of *Oxalobacteraceae*.

The fungal community of the snow microcosms was also very similar across time and treatments and was dominated by taxa belonging to the phylum of the *Basidiomycota*. RDP could not classify its members further in the taxonomy with enough confidence. The second most represented phylum was the *Ascomycota*. At the order level, after the unclassified *Basidiomycota*, the *Atheliales* were the second most abundant order.

4.3.2 Bacterial-fungal interactions assessed using 16S rRNA gene and ITS sequencing in co-variance networks

The taxonomy of the ASV retrieved in the core network of each time series can be found in the supplementary material (see annex pp. 145-147). The core community ASV represented more than 60% of all the sequencing depth and could reach up to 90% in some samples (see annex p. 148).

The acetate network had a bigger number of edges (3214) than the water control network (2434) and 853 edges connected the same ASV/nutrient nodes in both networks. Of these, 134 edges changed in terms of interaction type (e.g positive to negative and vice versa) in one two networks, while 604 edges remained identical. Of the 134 edges that changed, 42 negative edges in the water control time series network became positive in the acetate amended microcosms network and 92 positive edges became negative. The nodes were more connected in the acetate amended network with an average of 55.90 edges versus 41.97 edges in the water control network. The graph showing the degree distributions for both networks are provided in the supplementary material (see annex p. 149).



ne
ot
variance (considered as a surrogate of cooperation) and each red line represents a negative co-variance (interpreted as a possible competitive interaction).

Table 4: Table summarizing the relative abundance of specific kind of co-variances used as surrogates of interactions. The kind of co-variance is named based on the taxonomy of its interacting nodes (bact = bacteria, fung= fungi) and the sign of its LSA coefficient (Pos = positive LSA value, Neg= negative LSA value).

Time series	*bact_bact Pos*	*bact_fung Pos*	*fung_fung Pos*	*bact_bact Neg*	*bact_fung Neg*	*fung_fung Neg*
Acetate	33,21%	14,95%	9,56%	21,32%	15,44%	5,51%
Water	34,04%	30,44%	13,67%	5,56%	9,82%	6,46%

We looked at the predominance of certain kinds of covariance-links in both networks. Positive interactions between bacterial ASV were dominant in both networks, followed by positive interactions between Bacteria and Fungi in the water network, representing 30%. Negative interactions between bacteria was the second most dominant type of interaction in the acetate network (21%), while this represented only 5% of the interactions in the water network. Downsampling of the original dataset was preformed to test the robustness of the networks. Similar trends were observed for most of the features for the bigger subsampling sizes (3 and 2 replicates) as compared to the original networks. Subsampling size was shown to have an impact on some of the results, especially the negative interactions that became more similar between both networks at smaller subsampling sizes (see annex pp. 150-152).
Finally, we also looked more closely at the core network (interactions that were detected in both networks). The interactions that stayed stable across both time series included most of the ASV from the network (99 ASV were connected). We observed, for example, that a fungal ASV (*Basidiomycota_sp|SH216408.06FU*, OTU_11000) interacted positively with three bacterial ASVs from the family of *Oxalobacteraceae*: two ASVs could not be classified further (OTU_1 and OTU_119) and the last one was classified as the genus *Massilia* (OTU_40). More interactions like these were also observed. Some negative interactions stayed constant across both datasets like, for example, the interaction between a bacterial ASV classified into the order of *Rhizobiales* (OTU_47) and the fungal ASV classified as *Atheliaceae_sp|SH232729.06FU* (OTU_12100). We also detected negative bacterial interactions that stayed constant between both networks, for example, an ASV classified in the family of *Oxalobacteraceae* (OTU_58) and an ASV classified as *Granulicella sp.* (OTU_37). Positive interactions were also detected for bacterial ASVs as well as for fungal ASVs in both networks.

We detected that five fungal ASVs and two bacterial ASVs were interacting with oxalate in the water control microcosm. Both bacterial ASVs (OTU_186 and OTU_39, respectively classified as *Rhizobacter sp.* and *Massilia sp.*) interacted negatively with oxalate. Most fungal ASVs also interacted negatively with oxalate, except for one classified as *Rhodotorula_sp_TP_Snow_Y129|SH212318.06FU* (OTU_52100) that showed a positive interaction. This fungal ASV interacted positively with four bacterial ASVs (OTU_129, OTU_260, OTU_286 and OTU_7: respectively classified as *Hymenobacter sp., Burkholderiales* order, *Oxalobacteraceae* family and *Mucilaginibacter sp.*). These interactions tended to disappear in the acetate amended network, with the exception of one (with *Mucilaginibacter sp.,* OTU_7) which switched to a negative interaction.

4.4 Microorganisms produce organic acid substrates at temperatures below zero and respond to nutrient additions

We compared the chemistry of water control versus sodium acetate amended time series and the main chemical variables contributing to the difference between them was oxalic acid (Figure 1), which could only be detected in the water controls. In the water controls, oxalic acid, ammonium and acetate increased significantly over time (Figure 2 B). This increase supports the hypothesis that microbial communities are active at temperatures below zero and that they can alter their chemical environment. This would suggest that microorganisms might play a non-negligible role in the seasonal increases in organic acid concentrations in arctic snow (Twickler et al. 1986), in addition to exogenous inputs and photochemistry (Grannas et al. 2007; Christner et al. 2008). Further research is needed to quantify their contribution.

Oxalogenesis has mostly been reported in fungi, with three recognized pathways (i) the cytoplasmic pathway; (ii) the tricarboxylic acid (TCA) pathway; and (iii) the glyoxylate (GLOX) pathway in which the central reaction is the fixation of carbon dioxide on a molecule of pyruvate, via pyruvate carboxylase to form oxaloacetate (Plassard and Fransson 2009). After this step, oxaloacetate can be metabolized into oxalate, citrate or various other organic acids at different rates depending on different factors such as nutrient concentrations, pH and metal concentrations. For example, Cunningham and Kuiack (1992) found that citric acid production was promoted under nitrogen-limited conditions, while oxalic acid production was promoted under carbon-limited conditions. Only a few studies have reported oxalic acid production in bacteria, with production by *Pseudomonas* from a glyoxylate intermediate in response to aluminum stress (Hamel et al. 1999) and in two species of the genus *Burkholderia* as a pathogenicity factor (Nakata, 2011; Nakata & He, 2010). In our study, pyruvate carboxylase was identified as one of the pathways that was significant via boost regression (Table 1) and KEGG pathway analysis also showed that pyruvate metabolism was significantly enriched in the water controls (Table 2), suggesting a fungal, rather than bacterial, source.

In the sodium acetate amended microcosms, oxalic acid was undetected, suggesting that a metabolic shift occurred as a result of the additions. Acetate has been shown to be an effective antifungal (Stiles et al. 2002). Kang, Park, and Go (2003) have observed that several species of the genera *Colletotrichum* were inhibited by acetate concentrations of 30 mM when cultivated in pure culture. In comparison, the initial concentration of sodium acetate added in our snow microcosms was close to 350 mM which is more than ten times the inhibition threshold reported in this study. In addition, fungal contaminations (*Penicillium*) of culture plates have also been shown to grow slower on petri dishes where strains of *Lactobacillus* were cultured and produced acetate, showing the strong inhibitory potential of this organic compound on Fungi (Guimarães, Venancio, and Abrunhosa 2018). Given that the snow system is oligotrophic, it is possible that the levels added were sufficient to inhibit oxalogenesis. The communities in the amended microcosms seemed to respond to the additions after three days, with uptake in acetate, ammonium and nitrate. The delayed response in uptake might be related to nutrient stress, as oligotrophic organisms are less reactive to abrupt resource availability (Ho et al. 2017). This is supported by the high

abundance in ABC transporter genes in both data sets (6.3% in sodium acetate amended versus 3.3% in water). These genes are reported as being prevalent in oligotrophic bacteria (Lauro et al. 2009), but have also been found to be adaptations to organic acid toxicity and antibiotic resistance (Nakano, Fukaya, and Horinouchi 2006; Greene et al. 2018; Wolfger, Mamnun, and Kuchler 2001).

4.1 Nutrient additions affect microbial metabolic pathways

Based on the machine learning algorithm, the main microbial metabolisms affected by the amended sodium acetate were related to microbial competition (16%) (Table 1). This supports the hypothesis that microbial interaction dynamics represent an important component of the response of the snow microbial community to an increase in nutrients, as previously suggested in a field study (Bergk-Pinto et al., 2019).

Differential gene abundance analysis allowed us to identify the pathways that were more abundant in each of the treatments, the majority of which were found in the control (water) microcosms (422 versus 174). A high percentage of these genes (6%) were related to the two-component system, which was the third most abundant pathway in the water control set (Table 2). This category was also observed in the acetate gene set, although the likelihood for pathway enrichment was not significant (Table 2 and Table 3). These two-component systems, prevalently found in Prokaryotes and Archea, are known to help bacteria sense and response to external signals (Capra and Laub 2012). Since the Arctic snow is a dynamic environment (Maccario et al. 2015; 2019), bacteria must adapt to rapidly changing conditions, therefore it is unsurprising to retrieve a high abundance of these genes in this ecological niche (Capra and Laub 2012). We also observed that some pathways related to yeasts (e.g. yeast meiosis and autophagy) were identified in the water gene set with a significant likelihood (Table 2) of being enriched. Yeast cells undergo meiosis under nitrogen-starved conditions and require autophagy for meiosis initiation (Matsuhara and Yamamoto, 2015). Oxalic acid has been shown to suppress autophagy (Kabbage et al., 2013) and might constitute a negative interaction between fungi in the water control microcosms. Several other metabolisms have been retrieved as likely to be enriched (likelihood < 0.05) in the water gene sets and include mainly primary metabolism (e.g. glycerolipid, pyruvate, amino acids) with the exception of chloroalkane and chloroalkene degradation (Table 2). A pathway related to siderophores (Table 2) was exclusively observed in the water gene set (but not returned as being likely to be enriched), which suggests bacterial collaboration (e.g. D'Onofrio et al. 2010).

Among the genes that were dominant in the sodium acetate amended time series, secondary metabolites were identified (18%, likelihood 0.001, Table 2). Genes related to this pathway were also observed in the water set, but the likelihood of enrichment was not significant (10%, likelihood 0.78, Table 1). Secondary metabolism has been shown to be a stress response controlled by the nutrient balance in the environment of the microorganisms (e.g. Martín et al. 2011; Fujita 2009). Concerning the pathways returned as likely to be enriched in the acetate gene set, the metabolisms were more diverse, with genes related to the metabolism of ascorbate and aldarate (organic acids not tracked during this study), and butanoate metabolism in addition to some pathways related to primary metabolism (Table 3). In addition, the degradation of geraniol was also retrieved as being likely to be enriched only in the acetate gene set (Table 3). Geraniol, as other monoterpenes secreted by

microalgae, has been shown to be an antimicrobial active compound (Santos et al. 2016). The presence of microalgae is partially confirmed by the fact that pathways related to carbon fixation and photosynthesis were also shown to be enriched in the acetate gene set (Table 3). In a study on Arctic snow from Svalbard, Zhu et al. (2020) found a positive correlation between organic acids levels and geraniol degradation and fatty acid metabolism. These results suggest that our experimental set-up reproduces field observations in a laboratory setting and can be used for hypothesis testing.

The pathways of quorum sensing, antibiotic biosynthesis (e.g. ansamycins and prodigiosin) and biofilm formation (vibrio) were significantly enriched in the acetate gene set compared to random sampling (Table 3). The enrichment of pathways related to the biosynthesis of antimicrobial compounds suggests that bacterial competition increases as a result of nutrient amendment (e.g. Gao et al. 2018) (Sánchez et al. 2010), which supports the results obtained in our field study (Bergk-Pinto et al., 2019).

4.2 Bacterial and fungal networks are impacted by the addition of organic acids

The networks built by using the ASV considered as the core microbiome of bacteria and fungi co-variances showed contrasting structures (Figure 5). The interaction changes observed could be extrapolated to a large fraction of the snow community, since 60-90% of the sequences were represented by the networks (annex p. 148). In addition, bacterial growth dynamics (Figure 3) and microbial community structure (Figure *4* and annex p. 144) were similar in both time series, therefore no other biological parameters except interactions could explain the differences observed in our networks. Downsampling of our initial dataset was applied to generate pseudo-replicates of networks to determine robustness. For a downsampling size of 3 or 2 samples, the trends were the same as our original networks, suggesting that they are robust. This was not the case for a downsampling to a size of 1 (see annex pp. 142-144), but this is probably due to the use of an algorithm (eLSA) that was originally designed for time series with replicates and thus not well adapted for non-replicated (1 sample per sampling time) data (Xia et al. 2011).

Several ASVs from our core microbiome have previously been reported in cold habitats. For example, *Basidiomycota_sp|SH216408.06FU* was identified in the maritime Antarctic region (Newsham et al. 2015), species affiliated to the genus *Granulicella sp.* have been isolated from Arctic soils (Männistö et al. 2012; Oshkin et al. 2019), *Rhodotorula_sp_TP_Snow_Y129|SH212318.06FU* was reported in a study of yeasts extracted from cold snowpacks from the Tibetan plateau (GenBank: JQ768923.1) (Clark et al. 2016), a species of *Hymenobacter* has been isolated in red snow from Antarctica (Kojima et al. 2016) and *Mucilaginibacter sp.* was previously reported in a study of Antarctic snowpacks (Antony et al. 2016). Those observations suggest that our microbial community was representative of cold ecosystems.

We did not observe a strong difference in terms of positive co-variance (considered as a proxy of collaboration) between bacteria in our networks (Table 4), but a drop in positive bacterial and fungal was observed in the sodium acetate amended microcosms. This suggests that nutrient levels do not affect bacterial collaboration, but might play a role in fungal-bacterial collaboration.

Oxalic acid is the most common light molecular weight organic acid produced among fungi and is reported to play a key role in the regulation of bacterial-fungal interactions (Deveau et al. 2018) and ecosystem functioning (Palmieri et al. 2019). Oxalic acid also serves as a public good and plays a central role in maintaining pH homeostasis (Oh et al. 2014; Plassard and Fransson 2009) and has been reported to participate in mineral weathering (Cheng et al. 2017; Becerra-Castro et al. 2013; Frey et al. 2010). Fungi, as Bacteria, can produce organic acids such as oxalate via metabolic pathways referred as overflow metabolism (Pinu et al. 2018; Geoffrey M. Gadd 1999; Palmieri et al. 2019). These metabolic byproducts can then be used by other microorganisms and initiate a cross feeding interaction (Carlson et al. 2018). Oxalate can also be used as a carbon source (Palmieri et al. 2019; Haq et al. 2018) and oxalotrophic bacteria have been shown to use oxalic acid to localize their oxalogenic fungal host and move towards it by quorum sensing-dependent chemotaxis in order for both partners to interact (Rudnick, Veen, and Boer 2015). This is supported by the network results that show almost 30% positive interactions between bacteria and fungi in the water controls. In the acetate amended samples, these positive interactions were reduced to 15% (Table 3). In our water network, we detected positive interactions between the fungal ASV (*Rhodotorula_sp_TP_Snow_Y129|SH212318.06FU*) suspected to produce the oxalate and several bacterial ASVs. In the acetate amended network, these interactions were no longer detected or switched to negative interactions. These results suggest that cross-feeding interactions occur in the snow, with cooperative interactions between oxalogenic fungi and oxalotrophic bacteria, as highlighted in the water control time series. This is in line with a recent study by Velez et al. (2018), that highlighted cross-kingdom interactions as an adaptive trait to oligotrophic environments by favouring microbial colonization and growth under low nutrient conditions.

The biggest changes were in the negative co-variances between bacterial ASVs (interpreted as a proxy of bacterial competition), where a fourfold increase in negative edges was observed in the sodium acetate amended network as compared to the water control network (Table 4). The use of the ratio of positive versus negative edges as an indicator of cooperative or competitive microbial communities was first proposed by Ding et al. (2015). This result supports the metagenomic observations that the increase in competition is related the amendment of nutrients in the snow. The increase in negative interactions changed the network representation, with an apparent tighter clustering in the bacterial nodes in the sodium acetate network than in the water network where fungal (pink) and bacterial nodes (blue) where more intermixed (Figure 5). The edges connecting bacterial nodes represented 54.53% of all the edges compared to 39.6% in the water network (Table 4). Although nutrient addition did not impact positive or negative co-variance proportions between fungi, the drop in collaboration (fewer positive edges) and increased competition (more negative edges) between fungal ASVs and bacterial ASVs in the sodium acetate amended snow networks (Table 4) suggests that fungi could become less dominant in microbial community interactions upon nutrient addition.

Nutrient addition might result in reduced dependence of bacteria on fungal exudates for survival (Velez et al. 2018). Thus, if our hypothesis of a possible cross-feeding consortium mediated by fungal production of organic acids is valid, then a switch to a non-limiting environment where competition is high could mediate a rapid change in the microbial

community. However, no such change was observed in our taxonomic data analysis (Figure 3). It is likely that a DNA-based approach was not sufficiently sensitive to capture dynamical changes at ASV level over the timing of the experiment and future work should include RNA-based approaches.

4.3 Conclusion and perspective

During this study, confirmed that the increase in organic acids in the Arctic snow affects microbial interactions as previously observed in a field-based study (Bergk Pinto et al. 2019). In addition, we retrieved specific pathways, previously shown to be positively correlated to the increase in organic acids (e.g. geraniol degradation) in our snow microcosms amended with organic acids (Zhu et al. 2020). On the other hand, our microcosm experiments did not confirm that organic acids could modulate bacterial collaboration (Bergk Pinto et al. 2019), which suggests that confounding factors were at play in the field experiment. We also investigated fungal interactions in the snow using a network approach that showed that the response to organic acids was different than for bacteria. The interactions among the fungal community stayed stable in both microcosm time series, but collaboration with bacteria was shown to drop as the competitive interactions increased in relative abundance.

The next step would be to use metatranscriptomic sequencing to track microbial activity at a finer scale as the regulation of secondary metabolism could vary rapidly over time. This would then show possible correlations between organic acid concentrations and the secondary metabolic pathways linked to microbial competition (e.g. antibiotics biosynthesis).

5 Bibliography

Albers, Pieter, Bram Weytjens, René De Mot, Kathleen Marchal, and Dirk Springael. 2018. "Molecular Processes Underlying Synergistic Linuron Mineralization in a Triple-Species Bacterial Consortium Biofilm Revealed by Differential Transcriptomics." *MicrobiologyOpen* 7 (2): e00559. https://doi.org/10.1002/mbo3.559.

Antony, Runa, Aritri Sanyal, Neelam Kapse, Prashant K. Dhakephalkar, Meloth Thamban, and Shanta Nair. 2016. "Microbial Communities Associated with Antarctic Snow Pack and Their Biogeochemical Implications." *Microbiological Research* 192 (November): 192–202. https://doi.org/10.1016/j.micres.2016.07.004.

Becerra-Castro, Cristina, Petra Kidd, Melanie Kuffner, Ángeles Prieto-Fernández, Stephan Hann, Carmela Monterroso, Angela Sessitsch, Walter Wenzel, and Markus Puschenreiter. 2013. "Bacterially Induced Weathering of Ultramafic Rock and Its Implications for Phytoextraction." *Applied and Environmental Microbiology* 79 (17): 5094–5103. https://doi.org/10.1128/AEM.00402-13.

Bergk Pinto, Benoît, Lorrie Maccario, Aurélien Dommergue, Timothy M. Vogel, and Catherine Larose. 2019. "Do Organic Substrates Drive Microbial Community Interactions in Arctic Snow?" *Frontiers in Microbiology* 10. https://doi.org/10.3389/fmicb.2019.02492.

Blasche, Sonja, Yongkyu Kim, Ana Paula Oliveira, and Kiran R. Patil. 2017. “Model Microbial Communities for Ecosystems Biology.” *Current Opinion in Systems Biology*, Systems biology of model organisms, 6 (December): 51–57. https://doi.org/10.1016/j.coisb.2017.09.002.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Callahan, Benjamin J, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes. 2016. “DADA2: High Resolution Sample Inference from Illumina Amplicon Data.” *Nature Methods* 13 (7): 581–83. https://doi.org/10.1038/nmeth.3869.

Capra, Emily J., and Michael T. Laub. 2012. “Evolution of Two-Component Signal Transduction Systems.” *Annual Review of Microbiology* 66 (1): 325–47. https://doi.org/10.1146/annurev-micro-092611-150039.

Carlson, Ross P., Ashley E. Beck, Poonam Phalak, Matthew W. Fields, Tomas Gedeon, Luke Hanley, William R. Harcombe, Michael A. Henson, and Jeffrey J. Heys. 2018. “Competitive Resource Allocation to Metabolic Pathways Contributes to Overflow Metabolisms and Emergent Properties in Cross-Feeding Microbial Consortia.” *Biochemical Society Transactions* 46 (2): 269–84. https://doi.org/10.1042/BST20170242.

Cheng, Cheng, Qi Wang, Linyan He, and Xiafang Sheng. 2017. “Change in Mineral Weathering Behaviors of a Bacterium Chitinophaga Jiangningensis JN53 under Different Nutrition Conditions.” *Journal of Basic Microbiology* 57 (4): 293–301. https://doi.org/10.1002/jobm.201600652.

Chignell, J. F., S. Park, C. M. R. Lacerda, S. K. De Long, and K. F. Reardon. 2018. “Label-Free Proteomics of a Defined, Binary Co-Culture Reveals Diversity of Competitive Responses Between Members of a Model Soil Microbial System.” *Microbial Ecology* 75 (3): 701–19. https://doi.org/10.1007/s00248-017-1072-1.

Christner, Brent C., Rongman Cai, Cindy E. Morris, Kevin S. McCarter, Christine M. Foreman, Mark L. Skidmore, Scott N. Montross, and David C. Sands. 2008. “Geographic, Seasonal, and Precipitation Chemistry Influence on the Abundance and Activity of Biological Ice Nucleators in Rain and Snow.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (48): 18854–59. https://doi.org/10.1073/pnas.0809816105.

Clark, Karen, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and Eric W. Sayers. 2016. “GenBank.” *Nucleic Acids Research* 44 (Database issue): D67–72. https://doi.org/10.1093/nar/gkv1276.

Deveau, Aurélie, Gregory Bonito, Jessie Uehling, Mathieu Paoletti, Matthias Becker, Saskia Bindschedler, Stéphane Hacquard, et al. 2018. “Bacterial-Fungal Interactions: Ecology, Mechanisms and Challenges.” *FEMS Microbiology Reviews* 42 (3): 335–52. https://doi.org/10.1093/femsre/fuy008.

Ding, Junjun, Yuguang Zhang, Ye Deng, Jing Cong, Hui Lu, Xin Sun, Caiyun Yang, et al. 2015. “Integrated Metagenomics and Network Analysis of Soil Microbial Community of the Forest Timberline.” *Scientific Reports* 5 (January): 7994. https://doi.org/10.1038/srep07994.

Dixon, Philip. 2003. “VEGAN, a Package of R Functions for Community Ecology.” *Journal of Vegetation Science* 14 (6): 927–30. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

D’Onofrio, Anthony, Jason M. Crawford, Eric J. Stewart, Kathrin Witt, Ekaterina Gavrish, Slava Epstein, Jon Clardy, and Kim Lewis. 2010. “Siderophores from Neighboring Organisms Promote the Growth of Uncultured Bacteria.” *Chemistry & Biology* 17 (3): 254–64. https://doi.org/10.1016/j.chembiol.2010.02.010.

Eren, A. Murat, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, and Tom O. Delmont. 2015. “Anvi’o: An Advanced Analysis and Visualization Platform for ‘omics Data.” *PeerJ* 3 (October): e1319. https://doi.org/10.7717/peerj.1319.

Frey, Beat, Stefan R. Rieder, Ivano Brunner, Michael Plötze, Stefan Koetzsch, Ales Lapanje, Helmut Brandl, and Gerhard Furrer. 2010. “Weathering-Associated Bacteria from the Damma Glacier Forefield: Physiological Capabilities and Impact on Granite Dissolution.” *Applied and Environmental Microbiology* 76 (14): 4788–96. https://doi.org/10.1128/AEM.00657-10.

Fujita, Yasutaro. 2009. “Carbon Catabolite Control of the Metabolic Network in <I>Bacillus Subtilis</I>.” *Bioscience, Biotechnology, and Biochemistry* 73 (2): 245–59. https://doi.org/10.1271/bbb.80479.

Gadd, Geoffrey M. 1999. “Fungal Production of Citric and Oxalic Acid: Importance in Metal Speciation, Physiology and Biogeochemical Processes.” In *Advances in Microbial Physiology*, edited by R. K. Poole, 41:47–92. Academic Press. https://doi.org/10.1016/S0065-2911(08)60165-4.

Gao, Chun-Hui, Peng Cai, Zhunjie Li, Yichao Wu, and Qiaoyun Huang. 2018. “Co-Culture of Soil Biofilm Isolates Enables the Discovery of Novel Antibiotics.” *BioRxiv*, June, 353755. https://doi.org/10.1101/353755.

Grannas, A. M., A. E. Jones, J. Dibb, M. Ammann, C. Anastasio, H. J. Beine, M. Bergin, et al. 2007. “An Overview of Snow Photochemistry: Evidence, Mechanisms and Impacts.” *Atmos. Chem. Phys.* 7 (16): 4329–73. https://doi.org/10.5194/acp-7-4329-2007.

Greene, Nicholas P., Elise Kaplan, Allister Crow, and Vassilis Koronakis. 2018. “Antibiotic Resistance Mediated by the MacB ABC Transporter Family: A Structural and Functional Perspective.” *Frontiers in Microbiology* 9. https://doi.org/10.3389/fmicb.2018.00950.

Guillonneau, Richard, Claudine Baraquet, Alexis Bazire, and Maëlle Molmeret. 2018. “Multispecies Biofilm Development of Marine Bacteria Implies Complex Relationships Through Competition and Synergy and Modification of Matrix Components.” *Frontiers in Microbiology* 9. https://doi.org/10.3389/fmicb.2018.01960.

Guimarães, Ana, Armando Venancio, and Luís Abrunhosa. 2018. “Antifungal Effect of Organic Acids from Lactic Acid Bacteria on Penicillium Nordicum.” *Food Additives & Contaminants. Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment* 35 (9): 1803–18. https://doi.org/10.1080/19440049.2018.1500718.

Hansen, Lea Benedicte Skov, Dawei Ren, Mette Burmølle, and Søren J. Sørensen. 2017. “Distinct Gene Expression Profile of *Xanthomonas Retroflexus* Engaged in Synergistic Multispecies Biofilm Formation.” *The ISME Journal* 11 (1): 300–303. https://doi.org/10.1038/ismej.2016.107.

Haq, Irshad Ul, Reto Daniel Zwahlen, Pu Yang, and Jan Dirk van Elsas. 2018. "The Response of Paraburkholderia Terrae Strains to Two Soil Fungi and the Potential Role of Oxalate." *Frontiers in Microbiology* 9. https://doi.org/10.3389/fmicb.2018.00989.

Ho, Adrian, Di Lonardo, D. Paolo, and Paul L. E. Bodelier. 2017. "Revisiting Life Strategy Concepts in Environmental Microbial Ecology." *FEMS Microbiology Ecology* 93 (3). https://doi.org/10.1093/femsec/fix006.

Hyatt, Doug, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11 (1): 119. https://doi.org/10.1186/1471-2105-11-119.

Kang, Han-Chul, Yong-Hwan Park, and Seung-Joo Go. 2003. "Growth Inhibition of a Phytopathogenic Fungus, Colletotrichum Species by Acetic Acid." *Microbiological Research* 158 (4): 321–26. https://doi.org/10.1078/0944-5013-00211.

Khan, Nymul, Yukari Maezato, Ryan S. McClure, Colin J. Brislawn, Jennifer M. Mobberley, Nancy Isern, William B. Chrisler, et al. 2018. "Phenotypic Responses to Interspecies Competition and Commensalism in a Naturally-Derived Microbial Co-Culture." *Scientific Reports* 8 (1): 297. https://doi.org/10.1038/s41598-017-18630-1.

Kojima, Hisaya, Miho Watanabe, Riho Tokizawa, Arisa Shinohara, and Manabu Fukui. 2016. "Hymenobacter Nivis Sp. Nov., Isolated from Red Snow in Antarctica." *International Journal of Systematic and Evolutionary Microbiology* 66 (11): 4821–25. https://doi.org/10.1099/ijsem.0.001435.

Langdon, W. B. 2015. "Performance of Genetic Programming Optimised Bowtie2 on Genome Comparison and Analytic Testing (GCAT) Benchmarks." *BioData Mining* 8 (1): 1. https://doi.org/10.1186/s13040-014-0034-0.

Larose, Catherine, Sibel Berger, Christophe Ferrari, Elisabeth Navarro, Aurélien Dommergue, Dominique Schneider, and Timothy M. Vogel. 2010. "Microbial Sequences Retrieved from Environmental Samples from Seasonal Arctic Snow and Meltwater from Svalbard, Norway." *Extremophiles: Life Under Extreme Conditions* 14 (2): 205–12. https://doi.org/10.1007/s00792-009-0299-2.

Lauro, Federico M., Diane McDougald, Torsten Thomas, Timothy J. Williams, Suhelen Egan, Scott Rice, Matthew Z. DeMaere, et al. 2009. "The Genomic Basis of Trophic Strategy in Marine Bacteria." *Proceedings of the National Academy of Sciences* 106 (37): 15527–33. https://doi.org/10.1073/pnas.0903507106.

Li, Dinghua, Ruibang Luo, Chi-Man Liu, Chi-Ming Leung, Hing-Fung Ting, Kunihiko Sadakane, Hiroshi Yamashita, and Tak-Wah Lam. 2016. "MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler Driven by Advanced Methodologies and Community Practices." *Methods*, Pan-omics analysis of biological data, 102 (June): 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

Maccario, Lorrie, Shelly D. Carpenter, Jody W. Deming, Timothy M. Vogel, and Catherine Larose. 2019. "Sources and Selection of Snow-Specific Microbial Communities in a Greenlandic Sea Ice Snow Cover." *Scientific Reports* 9 (1): 2290. https://doi.org/10.1038/s41598-019-38744-y.

Maccario, Lorrie, Laura Sanguino, Timothy M. Vogel, and Catherine Larose. 2015. "Snow and Ice Ecosystems: Not so Extreme." *Research in Microbiology*, Special issue on Microbial diversity, adaptation and evolution, 166 (10): 782–95. https://doi.org/10.1016/j.resmic.2015.09.002.

Männistö, Minna K., Suman Rawat, Valentin Starovoytov, and Max M. Häggblom. 2012. "Granulicella Arctica Sp. Nov., Granulicella Mallensis Sp. Nov., Granulicella Tundricola Sp. Nov. and Granulicella Sapmiensis Sp. Nov., Novel Acidobacteria from Tundra Soil." *International Journal of Systematic and Evolutionary Microbiology* 62 (Pt 9): 2097–2106. https://doi.org/10.1099/ijs.0.031864-0.

Martín, Juan F., Alberto Sola-Landa, Fernando Santos-Beneit, Lorena T. Fernández-Martínez, Carlos Prieto, and Antonio Rodríguez-García. 2011. "Cross-Talk of Global Nutritional Regulators in the Control of Primary and Secondary Metabolism in Streptomyces." *Microbial Biotechnology* 4 (2): 165–74. https://doi.org/10.1111/j.1751-7915.2010.00235.x.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.Journal* 17 (1): 10–12. https://doi.org/10.14806/ej.17.1.200.

Mitri, Sara, and Kevin Richard Foster. 2013. "The Genotypic View of Social Interactions in Microbial Communities." *Annual Review of Genetics* 47 (1): 247–73. https://doi.org/10.1146/annurev-genet-111212-133307.

Molina-Santiago, Carlos, Zulema Udaondo, Baldo F. Cordero, and Juan L. Ramos. 2017. "Interspecies Cross-Talk between Co-Cultured Pseudomonas Putida and Escherichia Coli." *Environmental Microbiology Reports* 9 (4): 441–48. https://doi.org/10.1111/1758-2229.12553.

Nakano, Shigeru, Masahiro Fukaya, and Sueharu Horinouchi. 2006. "Putative ABC Transporter Responsible for Acetic Acid Resistance in Acetobacter Aceti." *Applied and Environmental Microbiology* 72 (1): 497–505. https://doi.org/10.1128/AEM.72.1.497-505.2006.

Newsham, Kevin, David Hopkins, Lilia Carvalhais, Peter Fretwell, Steven Rushton, Anthony O'Donnell, and Paul Dennis. 2015. "Relationship between Soil Fungal Diversity and Temperature in the Maritime Antarctic." *Nature Climate Change* 6 (September). https://doi.org/10.1038/nclimate2806.

Oh, Juntaek, Eunhye Goo, Ingyu Hwang, and Sangkee Rhee. 2014. "Structural Basis for Bacterial Quorum Sensing-Mediated Oxalogenesis." *The Journal of Biological Chemistry* 289 (March). https://doi.org/10.1074/jbc.M113.543462.

Oshkin, Igor Y., Irina S. Kulichevskaya, W. Irene C. Rijpstra, Jaap S. Sinninghe Damsté, Andrey L. Rakitin, Nikolai V. Ravin, and Svetlana N. Dedysh. 2019. "Granulicella Sibirica Sp. Nov., a Psychrotolerant Acidobacterium Isolated from an Organic Soil Layer in Forested Tundra, West Siberia." *International Journal of Systematic and Evolutionary Microbiology,* 69 (4): 1195–1201. https://doi.org/10.1099/ijsem.0.003290.

Palmieri, Fabio, Aislinn Estoppey, Geoffrey L. House, Andrea Lohberger, Saskia Bindschedler, Patrick S. G. Chain, and Pilar Junier. 2019. "Chapter Two - Oxalic Acid, a Molecule at the Crossroads of Bacterial-Fungal Interactions." In *Advances in Applied Microbiology*, edited by Geoffrey Michael Gadd and Sima Sariaslani, 106:49–77. Academic Press. https://doi.org/10.1016/bs.aambs.2018.10.001.

Paulson, Joseph N., O. Colin Stine, Héctor Corrada Bravo, and Mihai Pop. 2013. "Differential Abundance Analysis for Microbial Marker-Gene Surveys." *Nature Methods* 10 (12): 1200–1202. https://doi.org/10.1038/nmeth.2658.

Pinu, Farhana R., Ninna Granucci, James Daniell, Ting-Li Han, Sonia Carneiro, Isabel Rocha, Jens Nielsen, and Silas G. Villas-Boas. 2018. "Metabolite Secretion in

Microorganisms: The Theory of Metabolic Overflow Put to the Test." *Metabolomics* 14 (4): 43. https://doi.org/10.1007/s11306-018-1339-7.

Plassard, Claude, and Petra Fransson. 2009. "Regulation of Low-Molecular Weight Organic Acid Production in Fungi." *Fungal Biology Reviews* 23 (February): 30–39. https://doi.org/10.1016/j.fbr.2009.08.002.

Raghupathi, Prem K., Wenzheng Liu, Koen Sabbe, Kurt Houf, Mette Burmølle, and Søren J. Sørensen. 2018. "Synergistic Interactions within a Multispecies Biofilm Enhance Individual Species Protection against Grazing by a Pelagic Protozoan." *Frontiers in Microbiology* 8. https://doi.org/10.3389/fmicb.2017.02649.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. "EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26 (1): 139–40. https://doi.org/10.1093/bioinformatics/btp616.

Rudnick, M. B., J. A. van Veen, and W. de Boer. 2015. "Oxalic Acid: A Signal Molecule for Fungus-Feeding Bacteria of the Genus Collimonas?" *Environmental Microbiology Reports* 7 (5): 709–14. https://doi.org/10.1111/1758-2229.12290.

Sánchez, Sergio, Adán Chávez, Angela Forero, Yolanda García-Huante, Alba Romero, Mauricio Sánchez, Diana Rocha, et al. 2010. "Carbon Source Regulation of Antibiotic Production." *The Journal of Antibiotics* 63 (8): 442–59. https://doi.org/10.1038/ja.2010.78.

Santos, Andriéli, Karem Vieira, Gabriela Nogara, Roger Wagner, Eduardo Jacob-Lopes, and Leila Zepka. 2016. "Biogeneration of Volatile Organic Compounds by Microalgae: Occurrence, Behavior, Ecological Implications and Industrial Applications." In .

Seth, Erica C., and Michiko E. Taga. 2014. "Nutrient Cross-Feeding in the Microbial World." *Frontiers in Microbiology* 5. https://doi.org/10.3389/fmicb.2014.00350.

Stiles, J., S. Penkar, M. Plocková, J. Chumchalová, and L. B. Bullerman. 2002. "Antifungal Activity of Sodium Acetate and Lactobacillus Rhamnosus." *Journal of Food Protection* 65 (7): 1188–91. https://doi.org/10.4315/0362-028x-65.7.1188.

Taylor, D. Lee, William A. Walters, Niall J. Lennon, James Bochicchio, Andrew Krohn, J. Gregory Caporaso, and Taina Pennanen. 2016. "Accurate Estimation of Fungal Diversity and Abundance through Improved Lineage-Specific Primers Optimized for Illumina Amplicon Sequencing." *Applied and Environmental Microbiology* 82 (24): 7217–26. https://doi.org/10.1128/AEM.02576-16.

Twickler, Mark S., Mary Jo Spencer, W. Berry Lyons, and Paul A. Mayewski. 1986. "Measurement of Organic Carbon in Polar Snow Samples." *Nature* 320 (6058): 156–58. https://doi.org/10.1038/320156a0.

Velez, Patricia, Laura Espinosa-Asuar, Mario Figueroa, Jaime Gasca-Pineda, Eneas Aguirre-von-Wobeser, Luis E. Eguiarte, Abril Hernandez-Monroy, and Valeria Souza. 2018. "Nutrient Dependent Cross-Kingdom Interactions: Fungi and Bacteria From an Oligotrophic Desert Oasis." *Frontiers in Microbiology* 9 (August). https://doi.org/10.3389/fmicb.2018.01755.

Wolfger, Hubert, Yasmine M Mamnun, and Karl Kuchler. 2001. "Fungal ABC Proteins: Pleiotropic Drug Resistance, Stress Response and Cellular Detoxification." *Research in Microbiology* 152 (3): 375–89. https://doi.org/10.1016/S0923-2508(01)01209-8.

Xia, Li C., Dongmei Ai, Jacob Cram, Jed A. Fuhrman, and Fengzhu Sun. 2013. "Efficient Statistical Significance Approximation for Local Similarity Analysis of High-Throughput Time Series Data." *Bioinformatics* 29 (2): 230–37. https://doi.org/10.1093/bioinformatics/bts668.

Xia, Li C., Joshua A. Steele, Jacob A. Cram, Zoe G. Cardon, Sheri L. Simmons, Joseph J. Vallino, Jed A. Fuhrman, and Fengzhu Sun. 2011. “Extended Local Similarity Analysis (ELSA) of Microbial Community and Other Time Series Data with Replicates.” *BMC Systems Biology* 5 (2): S15. https://doi.org/10.1186/1752-0509-5-S2-S15.

Yannarell, Sarah M., Gabrielle M. Grandchamp, Shih-Yuan Chen, Karen E. Daniels, and Elizabeth A. Shank. 2019. “A Dual-Species Biofilm with Emergent Mechanical and Protective Properties.” *Journal of Bacteriology*, March, JB.00670-18. https://doi.org/10.1128/JB.00670-18.

Zhu, Chengsheng, Maximilian Miller, Nicholas Lusskin, Benoît Bergk Pinto, Lorrie Maccario, Max Haggblom, Timothy Vogel, Catherine Larose, and Yana Bromberg. 2020. “Snow Microbiome Functional Analyses Reveal Novel Microbial Metabolism of Complex Organic Compounds.” *BioRxiv*, February, 2020.02.07.938555. https://doi.org/10.1101/2020.02.07.938555.

Concluding remarks for the thesis

Microbial interactions are ubiquitous in the environment and structure their communities. Their importance is starting to be uncovered due to recent studies that showed the cooperation among microorganisms can prevail in some communities and act as an environmental parameter to generate specific and complex dynamics. The current state-of-the-art concerning bacterial interactions is mainly based on laboratory co-cultures of microorganisms under different experimental conditions. As stated in the introduction of this thesis, little is known about the representability of these experiments compared to environmental microbial communities. In addition, the tools available to study these interactions in a culture-free manner are currently limited and render their study at an environmental scale more challenging.

In order to validate some of these laboratory-based observations, we investigated bacterial interactions in the snow. The secondary objective was to design and validate a protocol to study microbial interactions in the environment. In chapter one, we summarized the methods available for studying interactions and decided to use two culture-free methods to strengthen the conclusions of our experiments. We focused on the effects of organic acids on microbial interactions, because cross-feeding could be mediated by these metabolites. The choice of snow as the environment of study was appropriate as it had been reported that organic acid concentrations in the snow showed a seasonality. Our hypothesis was that the microbial collaboration would be favored at lower organic acid concentrations and that competition would be favored once organic acid concentrations increased in the snow. As a consequence, an increase in the general nutrient conditions mirrored by organic acid concentrations (during Spring melt, for example) could transform the microbial community from one dominated by collaboration to one dominated by competition.

We tested our hypothesis during our first field-based experiment. Antibiotic resistance genes (ARGs), which were considered as proxy genes for competition, were found to be correlated with organic acid concentrations in the snow in both metagenomes and metatranscriptomes. In contrast, plasmid backbone genes, which were considered as proxy genes for collaboration, were more abundant in the metagenomes sampled when organic acid concentrations were low in the snow. Co-variance networks that were used as the second culture-free method to evaluate bacterial interactions confirmed this trend. Nonetheless, given that the seasonality of organic acid concentrations was also correlated to other environmental changes, we could not exclude that the observed changes in bacterial interactions were the result of other uninvestigated factors (*e.g*. snow melting, temperature, microbial inputs …).

To further test our hypothesis, we used a microcosm time-series approach. In addition, since we switched from Roche 454 pyrosequencing to illumina MiSeq sequencing, we needed to design a specific assembly-based pipeline to annotate our metagenomes. In the third chapter, I presented this new tool and showed that the assembly of a test dataset was enhanced when this pipeline was used instead of assembly alone. We complemented the assembly annotation with a read-based annotation in order to avoid losing meaningful data. The annotation of the reads that could not be recruited in the assembly was controlled for the percentage of spurious annotations by 'learning' the threshold to set on the assembly itself. This method

improved the number of reads used in the annotation of our metagenomes while controlling the noise introduced in our data. Since our interpretations of the genes used as proxies of cooperation and competition are highly dependent on the accuracy of this assembly and annotation, this pipeline was designed to make our interpretations more reliable. This approach was used for the metagenomes generated in our microcosm study described in the last thesis chapter. We also applied our hybrid approach (tracking of genes used as proxies of microbial interactions in metagenomes and co-variance networks) on the snow microcosm data. We observed that the organic-acid-amended microcosms appeared to be enriched in genes associated with biochemical pathways related to secondary metabolites and antibiotic biosynthesis. We observed that gene-associated metabolic pathways linked to antibiotics were significantly enriched in the acetate-amended time series. We did not observe any difference in plasmid backbone genes between the control snow microcosms and the organic-acid-amended ones. This observation was confirmed further with the metagenomic data using the KEGG database. We concluded that competition was higher in the bacterial community of the organic-acid-amended snow microcosms compared to the controls. The co-variance networks were consistent with this trend but, interestingly, did not show large differences in terms of collaboration or cooperation between bacteria. We concluded that the organic acids might trigger competition between bacteria and, yet, have little or no effect on bacterial cooperation. This last observation opens a new perspective concerning cross-kingdom cooperation between fungi and bacteria. We also measured an increase in the organic acid, oxalate, (associated with fungal activity) in the control snow microcosm. Thus, the seasonal increase in organic acids observed in the Arctic snow during the winter – spring transition could be due to *in situ* metabolic activity from the endogenous snow microbial community in addition to possible aerial deposition. Specific fungal taxa were found to be the potential producers of those organic acids (oxalate) and potentially interact with bacterial taxa in the co-variance networks. Thus, both Prokaryotes and Eukaryotes need to be included when studying the interactions as both could influence the dynamics of the whole community and excluding them could lead to spurious conclusions.

These results have initiated our understanding of microbial cooperation and competition in the environment away from the controlled laboratory conditions. This development required both innovative field work and bioinformatics, although the current methodology should evolve. For example, we experienced difficulties in tracking the genes related to some microbial interactions, such as antibiotic resistance, because some databases (*e.g*., the Gene Ontology) did not cross-validate the annotations of another database (*e.g*., KEGG). Thus, database choice has a significant impact on microbial interaction results and conclusions. We believe that the definition of genes implicated in microbial interactions placed in one single database to be used as a reference for this fast growing field will sustain more reproducible and comparable investigations.

Annexes

1 Chapter II annexes

Supplementary files for this chapter can be found online on the website of frontiers in microbiology using the following link: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02492/full#supplementary-material

2 Chapter IV annexes

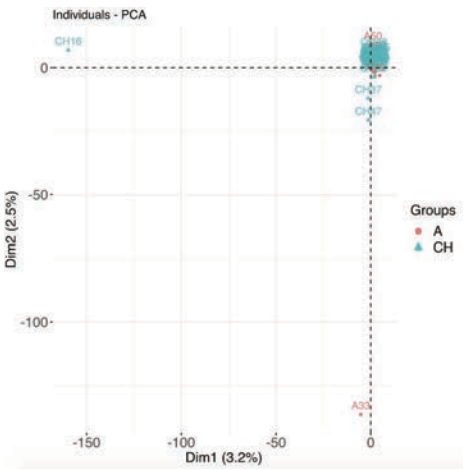


Figure 1: Principal component analysis biplot *from the* amplicon single variants (ASV) of the 16S rRNA sequencing of the samples used in this study. The samples*, green triangles* (*water control* samples) and *red dots* (*acetate amended* samples) are represented based on their respective projections.

Table 1: Table containing all the chemical measurements done on the snow microcosms samples. The first column provides the individual sample id to link each metagenome sample (fa files) to its chemical measurements. All the chemical measurements are reported in ppb and the q

Sample Fastq id	Time point	Sampling time	Sodium (ppb)	Ammonium (ppb)	Potassium (ppb)	Magnesium (ppb)	Calcium (ppb)	Acetate (ppb)	Chlore (ppb)	Nitrate (ppb)	Sulfate (ppb)	Oxalate (ppb)	qPCR16S (copies/ microL)	
43CH	t00	26/06/2018	3190.8	95.37	37.07	18.15	37.09	23.45	4043.76	1.78	841.41	5.59	3950	C
34CH	t00	26/06/2018	2276.48	26.94	31.17	48.4	87.66	124.24	3631.51	238.9	735.7	6.05	16635.5	C
37CH	t00	26/06/2018	2673.24	20.21	33.34	62.27	93.48	53.02	3871.35	314.95	806.65	5.18	13223.5	C
7CH	t00	26/06/2018	2914.51	63.2	34.43	55.87	68.67	54.31	3578.41	220.74	710.12	7.21	8575	C
40CH	t01	28/06/2018	2995.38	220.91	56.96	67.55	79.29	149.98	4147.51	208.25	768.37	15.25	63475.5	C
39CH	t01	28/06/2018	2480.12	196.27	61.85	88.3	85.18	287.22	4108.16	199.68	877.09	26.86	8049	C
16CH	t01	28/06/2018	2223.86	61.04	36.66	98.81	98.55	185.13	3761.17	134.79	776.96	14.61	12339	C
9CH	t01	28/06/2018	2897.12	256.55	62.16	63.86	63.44	399.62	3584.55	207.26	723.46	23.46	27551.5	C
17CH	t02	01/07/2018	2061.21	185.31	52.79	78.35	72.94	230.98	3386.7	183.21	721.2	30.31	25321.5	C
30CH	t02	01/07/2018	2603.64	403.28	63.33	80.92	69.4	310.25	4334.62	222.65	878.22	110.28	15180	C
4CH	t02	01/07/2018	2276.07	173.8	42.57	72.6	56.61	384.98	3450.57	176.56	692.09	16.79	33181.5	C
45CH	t02	01/07/2018	1996.44	265.98	67.96	93.29	73.96	619.83	3610.25	216.97	733.33	38.61	11655	C
28CH	t03	03/07/2018	2637.93	189.2	65.03	78.66	61.83	507.01	3892.64	235.5	818.25	27.92	11721	C
31CH	t03	03/07/2018	2668.73	26.23	42.69	99.43	77.53	23.64	4256.57	146.92	895.8	8.9	6534.5	C
6CH	t03	03/07/2018	2457.96	287.67	68.82	86.4	90.01	325.93	3422.78	200.81	691.25	56.21	56316	C
15CH	t03	03/07/2018	3248.12	375.28	67.71	76.06	90.57	352.38	4207.07	222.5	923.99	65.23	5659.5	C
11CH	t04	05/07/2018	2317.8	257.6	60.05	90.57	94.53	308.66	3959.85	210.06	830.84	14.82	3851	C
20CH	t04	05/07/2018	2407.93	480.21	65.35	77.52	74.31	312.69	3986.04	220.66	873.19	88.03	4173.5	C
21CH	t04	05/07/2018	2280.55	161	50.63	75.41	72.19	270.51	3483.88	182.69	731.82	9.51	40310.5	C
27CH	t04	05/07/2018	2335.6	316.36	64.59	119.74	175.73	621.99	3924.17	229.92	825.43	64.41	11907.5	C

26CH	t05	08/07/2018	2389.03	386.14	63.87	100.86	150.21	291.81	3919.24	214.26	832.04	98.64	31973	CH
2CH	t05	08/07/2018	2048.36	59.97	26.16	105.47	167.78	5.48	3603.81	66.93	725.93	3.59	40679.5	CH
8CH	t05	08/07/2018	2198.94	279.05	58.72	88.73	1256.46	318.53	3627.17	203.61	712.19	27.54	110092	CH
23CH	t05	08/07/2018	2078.53	262.49	46.64	47.63	127.46	434.36	3432.91	191.26	714.17	57.62	24561	CH
19CH	t06	10/07/2018	2852.13	295.58	59.81	63.57	103.68	546.31	3836.11	249.16	832.86	90.57	14720	CH
38CH	t06	10/07/2018	2323.02	268.75	63.77	88.18	94.38	929.34	3731.66	254.99	753.05	50.95	12189.5	CH
25CH	t06	10/07/2018	2346.02	204.22	59.67	84.25	118.16	343.02	3877.77	235.96	795.48	23.37	7174	CH
32CH	t06	10/07/2018	2178.69	221.42	61.43	85.1	61.17	289.21	3834.72	215.99	825.72	47.16	5589.5	CH
12CH	t07	12/07/2018	2420.17	309.33	63.37	101.96	76.08	358.36	4117.48	242.97	884.83	82.97	15708	CH
41CH	t08	12/07/2018	2321.95	81.67	45.96	111.37	92.75	22.76	3978.85	264.75	805.58	29.33	5039	CH
3CH	t07	12/07/2018	2071.86	219.29	56.07	79.99	51.03	249.87	3487.9	201.15	723.98	62.25	54422	CH
33CH	t07	12/07/2018	2469.31	250.91	56.69	78.09	53.18	21.3	3875.05	217.12	802.85	18.06	23591	CH
50CH	t08	16/07/2018	2203.24	243.8	73.95	97.92	89.46	266.45	3860.84	246.53	711.63	59.75	13700.5	CH
44CH	t08	16/07/2018	2573.38	223.91	61.17	91.65	67.33	251.44	4176.99	298.14	856.79	73.38	4344.5	CH
14CH	t07	16/07/2018	2094.89	250.84	61.49	113.74	94.67	375.9	3809.59	246.85	816.91	55.03	46081	CH
47CH	t08	16/07/2018	2170.59	333.17	58.22	111.47	107.49	492.6	3940.39	252.72	768.66	81.2	12367.5	CH
29CH	t09	18/07/2018	2422.77	411.34	58.23	72.13	62.54	359.84	3591.32	243.47	743.02	75.37	69589	CH
1CH	t09	18/07/2018	2173.22	237.16	64.37	96.83	86.44	298.57	3766.01	225.38	762.9	44.99	46029	CH
22CH	t09	18/07/2018	2329.37	167.37	54.58	90.78	75.16	1210.47	3469.84	247.17	697.9	37	98836	CH
18CH	t09	18/07/2018	2081.03	242.22	63.61	81.41	66.49	449.19	3632.7	209.69	715.4	68.6	23792	CH
24CH	t10	19/07/2018	2621.77	286.87	64.74	107.47	111.86	967.98	4140.81	282.65	919.48	85.82	13158	CH
13CH	t10	19/07/2018	2582.08	76.32	61.91	108.8	96.55	1032.99	4147.09	243.13	866.46	28.1	7603.5	CH
10CH	t10	19/07/2018	1988.53	288.34	57.86	88.78	78.97	433.08	3429.3	192.49	724.92	48.24	9514.5	CH
5CH	t10	19/07/2018	2006.92	149.86	61.45	92.39	74.37	837.09	3518.51	210.99	735.52	34.86	17796.5	CH
29A	t00	26/06/2018	3317.37	354.5	31.71	54.06	239.76	3354.92	3895.46	179.88	742.96	0.5	382	A
28A	t00	26/06/2018	3487.18	426.53	21.68	79.98	321.59	1741.95	3663.24	147.61	700.1	0.5	23170	A
24A	t00	26/06/2018	3316.07	415.59	13.19	94.5	330.64	1507.82	3545.65	48.79	665.01	0.5	15597	A
1A	t00	26/06/2018	3378.53	669.29	37.19	111.09	514	49.98	3643.44	0.5	698.26	0.5	20176.5	A

42A	t01	28/06/2018	3417.79	581.31	50.83	117.72	474.02	3277.87	3606.96	190.42	669.35	0.5	12110.5	A
4A	t01	28/06/2018	3591.14	62.79	46.56	66.27	96.6	2752.84	3780.61	196.41	747.38	0.5	7064	A
11A	t01	28/06/2018	3342.45	622.51	52.86	159.63	607.3	2590.03	3454.66	156.75	644.2	0.5	26652	A
3A	t01	28/06/2018	3987.67	429.74	31.56	125.31	452.23	324.59	4254.19	11.27	782.33	0.5	48081	A
19A	t02	01/07/2018	3813.82	295.87	37.64	121.98	370.4	814.89	3888.03	0.5	794.83	0.5	11282	A
31A	t02	01/07/2018	3514.22	110.77	55.92	86.72	207.57	507.71	3833.31	20.26	734	0.5	18371	A
15A	t02	01/07/2018	3594.31	58.32	49.18	84.39	173.04	2381.07	3647.04	70.96	650.06	0.5	44465	A
44A	t02	01/07/2018	3864.22	32.05	44.49	68.91	120.46	552.44	3954.22	83.82	791.82	0.5	6181	A
23A	t03	03/07/2018	4086.14	18.22	38.87	62.8	105.12	73.1	3690.98	0.5	782.13	0.5	5944	A
12A	t03	03/07/2018	3801.15	17.25	49.87	73.5	112.33	1343.26	3934.81	27.43	830.51	0.5	4223.5	A
30A	t03	03/07/2018	3317.01	22.85	39.75	83.09	139.4	1151.75	3883.08	109.93	844.22	0.5	2080.5	A
10A	t03	03/07/2018	3705.15	12.08	29.26	60.72	81.56	0.5	3628.09	0.5	642.88	0.5	18048	A
34A	t04	05/07/2018	3758.37	13.06	31.76	60.77	77.76	31.28	3688.58	0.5	766.08	0.5	9686	A
32A	t04	05/07/2018	3675.44	8.5	352.59	45.82	51.12	30.82	3855.1	74.37	730.94	0.5	26176.5	A
20A	t04	05/07/2018	3424.9	8.71	43.46	50.78	53.68	147.11	3562.84	60.86	705.76	0.5	11392.5	A
48A	t04	05/07/2018	3754.97	19.63	40.78	107.28	135.01	0.5	4133.48	0.5	815.69	0.5	4099	A
33A	t05	08/07/2018	3773.66	20.08	47.74	85.95	125.32	1111.67	3820.49	66.15	726.44	0.5	36492	A
46A	t05	08/07/2018	4206.85	10.13	75.32	52.97	55.34	717.53	3948.48	105.16	717.18	0.5	80933	A
37A	t05	08/07/2018	3336.42	14.37	49.31	77.66	95.57	0.5	3575.65	0.5	616.3	0.5	70701	A
26A	t05	08/07/2018	3209.98	8.57	26.53	30.86	2800.82	71.27	3556.38	60.83	650.68	0.5	77019.5	A
22A	t06	10/07/2018	3985.21	16.57	47.6	53.9	75.92	0.5	3907.9	0.5	713.77	0.5	7946.5	A
6A	t06	10/07/2018	3389.01	11.85	41.43	63.58	96.89	0.5	3710.61	0.5	656.89	0.5	11639.5	A
51A	t06	10/07/2018	3523.65	11.3	66.17	65.32	92.67	91.15	3765.83	0.5	608.83	0.5	9383	A
13A	t06	10/07/2018	3826.41	28.4	56.72	74.54	76.21	2664.99	4012.1	212.4	701.02	0.5	10781.5	A
35A	t07	12/07/2018	3407.4	14.17	54.9	74.89	117.01	0.5	3642.58	0.5	627.34	0.5	15363	A
50A	t07	12/07/2018	3378.56	11.31	144.75	65.85	76.5	775.5	3793.77	125.29	669.11	0.5	40631	A
9A	t07	12/07/2018	3798.78	12.57	87.91	55.88	77.99	1915.68	3787.55	69.11	954.2	0.5	26633	A
36A	t07	12/07/2018	3403.74	8.22	48.05	56.35	62.46	2742.4	3653.78	0.5	642.73	0.5	27936	A

25A	t08	16/07/2018	3440.33	13.93	48.85	75.36	95.15	2244	3488.07	0.5	629.61	0.5	4646	A
40A	t08	16/07/2018	3583.58	11.07	39.01	60.85	67.36	0.5	3704.7	0.5	664.34	0.5	54690	A
8A	t08	16/07/2018	3574.38	8.85	63.63	52.7	55.81	0.5	3649.66	0.5	596.63	0.5	4770	A
5A	t08	16/07/2018	3584.71	13.38	60.35	50.24	62.31	0.5	3654.03	0.5	963.19	0.5	8255.5	A
16A	t09	18/07/2018	3530.98	12.91	43.52	61.25	76.83	0.5	3902.16	0.5	1078.33	0.5	31725	A
14A	t09	18/07/2018	3752.53	11.05	46.72	52.31	59.8	0.5	3825.89	0.5	680.95	0.5	12452	A
39A	t09	18/07/2018	3353.49	9.8	42.72	58.26	57.09	0.5	3602.97	0.5	616.94	0.5	50529.5	A
45A	t09	18/07/2018	3466.42	10.04	46.4	61.96	61.29	0.5	3848.44	0.5	675.05	0.5	20444	A
21A	t10	19/07/2018	3439.28	9.74	56.99	61.97	56.81	143.18	3783.6	0.5	633.54	0.5	12872.5	A
2A	t10	19/07/2018	3540.78	11.79	42.12	65.72	59.04	0.5	3833.53	0.5	646.91	0.5	37322	A
38A	t10	19/07/2018	3473.62	10.33	57.85	62.96	51.79	0.5	3629.62	0.5	651.37	0.5	12789.5	A
7A	t10	19/07/2018	3393.91	10.34	49.85	69.21	52.73	28.76	3810.86	0.5	642.08	0.5	4479.5	A

Table 2:Table showing the GO terms tracked in our metagenome annotations as proxies of plasmids.

GO term ID	GO term classification	GO term name
GO:0060910	biological_process	negative regulation of DNA replication initiation involved in plasmid copy number maintenance
GO:0060908	biological_process	plasmid copy number maintenance
GO:0060909	biological_process	regulation of DNA replication initiation involved in plasmid copy number maintenance
GO:0075530	biological_process	establishment of latency as a linear episome
GO:0075529	biological_process	establishment of latency as a circular episome
GO:0075720	biological_process	establishment of episomal latency
GO:0030541	biological_process	plasmid partitioning
GO:0030543	biological_process	2-micrometer plasmid partitioning
GO:0006276	biological_process	plasmid maintenance
GO:0042150	biological_process	plasmid recombination

Table 3: Table showing the GO terms tracked in our metagenome annotations as proxies of acetate metabolism.

GO term ID	GO term classification	GO term name
GO:0006083	biological_process	acetate metabolic process
GO:0006846	biological_process	acetate transport
GO:0010034	biological_process	response to acetate
GO:0019654	biological_process	acetate fermentation

Table 4: Table showing the GO terms tracked in our metagenome annotations as proxies of antibiotics metabolisms.

GO term ID	GO term classification	GO term name
GO:0016999	biological_process	antibiotic metabolic process
GO:0042891	biological_process	antibiotic transport
GO:0042895	molecular_function	antibiotic transmembrane transporter activity
GO:0030651	biological_process	peptide antibiotic biosynthetic process
GO:0030650	biological_process	peptide antibiotic metabolic process
GO:0030653	biological_process	beta-lactam antibiotic metabolic process
GO:0030652	biological_process	peptide antibiotic catabolic process
GO:0030655	biological_process	beta-lactam antibiotic catabolic process
GO:0030654	biological_process	beta-lactam antibiotic biosynthetic process
GO:0030648	biological_process	aminoglycoside antibiotic biosynthetic process
GO:0030647	biological_process	aminoglycoside antibiotic metabolic process
GO:0030649	biological_process	aminoglycoside antibiotic catabolic process
GO:0046677	biological_process	response to antibiotic
GO:0046353	molecular_function	aminoglycoside 3-N-acetyltransferase activity
GO:0071236	biological_process	cellular response to antibiotic
GO:0042740	biological_process	exogenous antibiotic catabolic process
GO:0042741	biological_process	endogenous antibiotic catabolic process
GO:0017000	biological_process	antibiotic biosynthetic process
GO:0017001	biological_process	antibiotic catabolic process

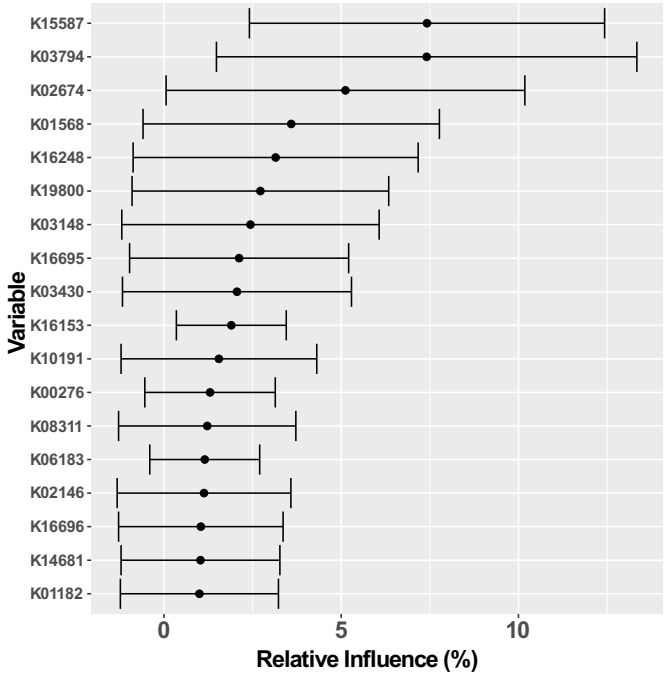


Figure 2: most influent Kegg Orthologs (KO) retrieved during the training of the boost regression algorithm. The dots show the respective average influence of the KO observed to have an average influence above 1% on the 1000 validations done during this experiment. The error bars display the standard deviation of this influence.

Table 5: Summary of the definitions and pathways of the KO retrieved by the boost regression as having the highest influence (see 7 for detailed influence).

Kegg ortholog id	Name	Definition	Pathway
K15587	nikD, cntD	nickel transport system ATP-binding protein [EC:7.2.2.11]	ABC transporters
K03794	sirB	sirohydrochlorin ferrochelatase [EC:4.99.1.4]	ko00860 Porphyrin and chlorophyll metabolism ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites
K02674	pilY1	type IV pilus assembly protein PilY1	Secretion system Bacterial motility proteins
K01568	PDC, pdc	pyruvate decarboxylase [EC:4.1.1.1]	ko00010 Glycolysis / Gluconeogenesis ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites ko01130 Biosynthesis of antibiotics
K16248	gutA, gutP	probable glucitol transport protein GutA	09183 Protein families: signaling and cellular processes Transporters [BR:ko02000]
K19800	SCH9	serine/threonine protein kinase SCH9 [EC:2.7.11.1]	ko04138 Autophagy - yeast ko04213 Longevity regulating pathway - multiple species

K03148	thiF	sulfur carrier protein ThiS adenylyltransferase [EC:2.7.7.73]	ko00730 Thiamine metabolism ko01100 Metabolic pathways ko04122 Sulfur relay system
K16695	wzxC	lipopolysaccharide exporter	Protein families: signaling and cellular processes Transporters [BR:ko02000]
K03430	phnW	2-aminoethylphosphonate-pyruvate transaminase [EC:2.6.1.37]	ko00440 Phosphonate and phosphinate metabolism ko01100 Metabolic pathways ko01120 Microbial metabolism in diverse environments
K16153	K16153	glycogen phosphorylase/synthase [EC:2.4.1.1 2.4.1.11]	ko00500 Starch and sucrose metabolism ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites
K10191	lacK	lactose/L-arabinose transport system ATP-binding protein	ko02010 ABC transporters
K00276	AOC3, AOC2, tynA	primary-amine oxidase [EC:1.4.3.21]	ko00260 Glycine, serine and threonine metabolism ko00350 Tyrosine metabolism ko00360 Phenylalanine metabolism ko00410 beta-Alanine metabolism ko00950 Isoquinoline alkaloid biosynthesis ko00960 Tropane, piperidine and pyridine alkaloid biosynthesis ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites
K08311	nudH	putative (di)nucleoside polyphosphate hydrolase [EC:3.6.1.-]	ko03018 RNA degradation
K06183	rsuA	16S rRNA pseudouridine516 synthase [EC:5.4.99.19]	09182 Protein families: genetic information processing 03009 Ribosome biogenesis
K02146	ATPeV0D, ATP6D	V-type H+-transporting ATPase subunit d	ko00190 Oxidative phosphorylation ko01100 Metabolic pathways
K16696	amsL	exopolysaccharide (amylovoran) exporter	09183 Protein families: signaling and cellular processes 02000 Transporters
K14681	argHA	argininosuccinate lyase / amino-acid N-acetyltransferase [EC:4.3.2.1 2.3.1.1]	ko00220 Arginine biosynthesis ko00250 Alanine, aspartate and glutamate metabolism ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites ko01130 Biosynthesis of antibiotics ko01210 2-Oxocarboxylic acid metabolism ko01230 Biosynthesis of amino acids
K01182	IMA, malL	oligo-1,6-glucosidase [EC:3.2.1.10]	ko00052 Galactose metabolism ko00500 Starch and sucrose metabolism ko01100 Metabolic pathways

Table 6: pathways returned as detected in the water control gene set. The first column describes the pathway id, its name and, between brackets, the number of ko retrieved for this particular pathway. The second column gives the total number of genes identified in the water gene set with one or more ko related to this particular pathway. The third column gives, as a fraction, the number of genes retrieved as being part of this pathway among all the genes from the water gene set (422 genes). The last column shows the p-value computed by random distribution method (see material and method) to detect if the number of genes related to this pathway was bigger than it would be due to a random event.

Definition	Ngenes Per pathway	Fraction Genes enriched	Likelihood
map01100 Metabolic pathways (128)	89	NA	NA
map01110 Biosynthesis of secondary metabolites (48)	45	0.1066	0.785
map01120 Microbial metabolism in diverse environments (57)	39	0.0924	0.638
map01130 Biosynthesis of antibiotics (29)	33	0.0782	0.647
map01200 Carbon metabolism (18)	17	0.0403	0.664
map01210 2-Oxocarboxylic acid metabolism (4)	4	0.0095	0.576
map01212 Fatty acid metabolism (7)	10	0.0237	0.478
map01230 Biosynthesis of amino acids (10)	9	0.0213	0.924
map01220 Degradation of aromatic compounds (12)	11	0.0261	0.101
map00010 Glycolysis / Gluconeogenesis (8)	10	0.0237	0.233
map00020 Citrate cycle (TCA cycle) (8)	7	0.0166	0.103
map00030 Pentose phosphate pathway (4)	3	0.0071	0.729
map00040 Pentose and glucuronate interconversions (3)	2	0.0047	0.855
map00051 Fructose and mannose metabolism (1)	1	0.0024	0.988
map00053 Ascorbate and aldarate metabolism (2)	6	0.0142	0.076
map00500 Starch and sucrose metabolism (6)	3	0.0071	0.93
map00520 Amino sugar and nucleotide sugar metabolism (2)	2	0.0047	0.991
map00620 Pyruvate metabolism (14)	15	0.0355	0.021
map00630 Glyoxylate and dicarboxylate metabolism (9)	8	0.0190	0.612
map00640 Propanoate metabolism (5)	6	0.0142	0.788
map00650 Butanoate metabolism (15)	12	0.0284	0.134
map00660 C5-Branched dibasic acid metabolism (2)	2	0.0047	0.407
map00562 Inositol phosphate metabolism (2)	2	0.0047	0.634
map00190 Oxidative phosphorylation (10)	7	0.0166	0.401
map00710 Carbon fixation in photosynthetic organisms (1)	1	0.0024	0.898
map00720 Carbon fixation pathways in prokaryotes (5)	5	0.0118	0.509
map00680 Methane metabolism (4)	3	0.0071	0.934

map00910 Nitrogen metabolism (3)	2	0.0047	0.822
map00920 Sulfur metabolism (3)	3	0.0071	0.819
map00061 Fatty acid biosynthesis (1)	5	0.0118	0.609
map00062 Fatty acid elongation (2)	2	0.0047	0.069
map00071 Fatty acid degradation (9)	13	0.0308	0.064
map00072 Synthesis and degradation of ketone bodies (3)	3	0.0071	0.237
map00140 Steroid hormone biosynthesis (2)	2	0.0047	0.242
map00561 Glycerolipid metabolism (4)	8	0.0190	0.023
map00564 Glycerophospholipid metabolism (2)	2	0.0047	0.879
map00565 Ether lipid metabolism (1)	1	0.0024	0.629
map00592 alpha-Linolenic acid metabolism (1)	1	0.0024	0.671
map01040 Biosynthesis of unsaturated fatty acids (1)	1	0.0024	0.645
map00230 Purine metabolism (5)	3	0.0071	0.983
map00240 Pyrimidine metabolism (3)	2	0.0047	0.919
map00250 Alanine, aspartate and glutamate metabolism (3)	2	0.0047	0.935
map00260 Glycine, serine and threonine metabolism (8)	8	0.0190	0.416
map00270 Cysteine and methionine metabolism (3)	3	0.0071	0.912
map00280 Valine, leucine and isoleucine degradation (10)	14	0.0332	0.056
map00290 Valine, leucine and isoleucine biosynthesis (1)	1	0.0024	0.85
map00300 Lysine biosynthesis (1)	1	0.0024	0.913
map00310 Lysine degradation (6)	9	0.0213	0.144
map00330 Arginine and proline metabolism (3)	6	0.0142	0.509
map00340 Histidine metabolism (3)	7	0.0166	0.016
map00350 Tyrosine metabolism (4)	4	0.0095	0.763
map00360 Phenylalanine metabolism (2)	4	0.0095	0.819
map00380 Tryptophan metabolism (7)	11	0.0261	0.042
map00400 Phenylalanine, tyrosine and tryptophan biosynthesis (2)	2	0.0047	0.687
map00410 beta-Alanine metabolism (3)	8	0.0190	0.093
map00440 Phosphonate and phosphinate metabolism (1)	1	0.0024	0.573
map00450 Selenocompound metabolism (2)	2	0.0047	0.621
map00460 Cyanoamino acid metabolism (2)	1	0.0024	0.83
map00480 Glutathione metabolism (3)	4	0.0095	0.461
map00510 N-Glycan biosynthesis (1)	1	0.0024	0.76
map00532 Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate (1)	1	0.0024	0.105
map00534 Glycosaminoglycan biosynthesis - heparan sulfate / heparin (1)	1	0.0024	0.067

map00550 Peptidoglycan biosynthesis (1)	1	0.0024	0.672
map00511 Other glycan degradation (1)	1	0.0024	0.766
map00571 Lipoarabinomannan (LAM) biosynthesis (2)	1	0.0024	0.396
map00740 Riboflavin metabolism (1)	1	0.0024	0.654
map00760 Nicotinate and nicotinamide metabolism (6)	4	0.0095	0.357
map00770 Pantothenate and CoA biosynthesis (2)	2	0.0047	0.648
map00780 Biotin metabolism (1)	5	0.0118	0.252
map00790 Folate biosynthesis (3)	3	0.0071	0.417
map00830 Retinol metabolism (2)	3	0.0071	0.251
map00860 Porphyrin and chlorophyll metabolism (4)	2	0.0047	0.877
map00130 Ubiquinone and other terpenoid-quinone biosynthesis (1)	1	0.0024	0.904
map00900 Terpenoid backbone biosynthesis (2)	2	0.0047	0.67
map00981 Insect hormone biosynthesis (1)	5	0.0118	0.004
map00903 Limonene and pinene degradation (1)	5	0.0118	0.065
map00281 Geraniol degradation (5)	5	0.0118	0.072
map01052 Type I polyketide structures (2)	1	0.0024	0.271
map01054 Nonribosomal peptide structures (3)	1	0.0024	0.356
map01053 Biosynthesis of siderophore group nonribosomal peptides (5)	1	0.0024	0.575
map00232 Caffeine metabolism (1)	1	0.0024	0.249
map00311 Penicillin and cephalosporin biosynthesis (1)	1	0.0024	0.438
map00333 Prodigiosin biosynthesis (1)	5	0.0118	0.108
map00999 Biosynthesis of various secondary metabolites - part 1 (1)	1	0.0024	0.05
map00998 Biosynthesis of various secondary metabolites - part 2 (1)	1	0.0024	0.477
map00362 Benzoate degradation (11)	10	0.0237	0.065
map00627 Aminobenzoate degradation (1)	1	0.0024	0.99
map00364 Fluorobenzoate degradation (1)	1	0.0024	0.624
map00625 Chloroalkane and chloroalkene degradation (5)	8	0.0190	0.013
map00361 Chlorocyclohexane and chlorobenzene degradation (1)	1	0.0024	0.784
map00623 Toluene degradation (2)	2	0.0047	0.545
map00622 Xylene degradation (2)	3	0.0071	0.121
map00633 Nitrotoluene degradation (1)	1	0.0024	0.629
map00642 Ethylbenzene degradation (1)	1	0.0024	0.546

map00643 Styrene degradation (1)	3	0.0071	0.32
map00930 Caprolactam degradation (1)	1	0.0024	0.894
map00626 Naphthalene degradation (2)	2	0.0047	0.555
map00365 Furfural degradation (1)	1	0.0024	0.296
map00980 Metabolism of xenobiotics by cytochrome P450 (3)	5	0.0118	0.161
map00982 Drug metabolism - cytochrome P450 (2)	4	0.0095	0.372
map00983 Drug metabolism - other enzymes (3)	4	0.0095	0.373
map03040 Spliceosome (4)	5	0.0118	0.338
map03010 Ribosome (10)	7	0.0166	0.262
map00970 Aminoacyl-tRNA biosynthesis (4)	3	0.0071	0.666
map03013 RNA transport (3)	3	0.0071	0.667
map03015 mRNA surveillance pathway (2)	2	0.0047	0.605
map03008 Ribosome biogenesis in eukaryotes (4)	5	0.0118	0.156
map04141 Protein processing in endoplasmic reticulum (4)	3	0.0071	0.766
map04120 Ubiquitin mediated proteolysis (6)	4	0.0095	0.273
map04122 Sulfur relay system (1)	1	0.0024	0.742
map03018 RNA degradation (2)	2	0.0047	0.824
map03030 DNA replication (3)	3	0.0071	0.355
map03430 Mismatch repair (1)	1	0.0024	0.84
map03440 Homologous recombination (1)	1	0.0024	0.85
map02010 ABC transporters (45)	14	0.0332	0.664
map03070 Bacterial secretion system (4)	3	0.0071	0.793
map02020 Two-component system (40)	27	0.0640	0.022
map04014 Ras signaling pathway (3)	3	0.0071	0.241
map04015 Rap1 signaling pathway (2)	2	0.0047	0.389
map04010 MAPK signaling pathway (2)	2	0.0047	0.514
map04013 MAPK signaling pathway - fly (1)	1	0.0024	0.731
map04016 MAPK signaling pathway - plant (1)	3	0.0071	0.084
map04012 ErbB signaling pathway (2)	2	0.0047	0.217
map04310 Wnt signaling pathway (1)	1	0.0024	0.739
map04340 Hedgehog signaling pathway (1)	1	0.0024	0.573
map04341 Hedgehog signaling pathway - fly (2)	2	0.0047	0.165
map04390 Hippo signaling pathway (1)	1	0.0024	0.733
map04391 Hippo signaling pathway - fly (1)	1	0.0024	0.66
map04370 VEGF signaling pathway (1)	1	0.0024	0.572
map04371 Apelin signaling pathway (2)	2	0.0047	0.506
map04630 JAK-STAT signaling pathway (2)	2	0.0047	0.07
map04668 TNF signaling pathway (2)	2	0.0047	0.133
map04066 HIF-1 signaling pathway (1)	1	0.0024	0.88
map04068 FoxO signaling pathway (6)	4	0.0095	0.191
map04020 Calcium signaling pathway (2)	2	0.0047	0.508

map04070 Phosphatidylinositol signaling system (1)	1	0.0024	0.726
map04072 Phospholipase D signaling pathway (2)	2	0.0047	0.459
map04071 Sphingolipid signaling pathway (1)	1	0.0024	0.889
map04024 cAMP signaling pathway (3)	3	0.0071	0.407
map04022 cGMP-PKG signaling pathway (1)	1	0.0024	0.839
map04151 PI3K-Akt signaling pathway (6)	5	0.0118	0.163
map04152 AMPK signaling pathway (3)	3	0.0071	0.507
map04150 mTOR signaling pathway (3)	3	0.0071	0.325
map04144 Endocytosis (3)	4	0.0095	0.511
map04145 Phagosome (5)	5	0.0118	0.084
map04142 Lysosome (7)	5	0.0118	0.199
map04146 Peroxisome (2)	2	0.0047	0.945
map04140 Autophagy - animal (5)	6	0.0142	0.068
map04138 Autophagy - yeast (4)	5	0.0118	0.073
map04136 Autophagy - other (1)	1	0.0024	0.577
map04137 Mitophagy - animal (1)	2	0.0047	0.154
map04139 Mitophagy - yeast (2)	2	0.0047	0.408
map04110 Cell cycle (6)	5	0.0118	0.143
map04111 Cell cycle - yeast (7)	6	0.0142	0.134
map04112 Cell cycle - Caulobacter (1)	1	0.0024	0.791
map04113 Meiosis - yeast (7)	7	0.0166	0.059
map04114 Oocyte meiosis (4)	4	0.0095	0.271
map04210 Apoptosis (5)	3	0.0071	0.325
map04214 Apoptosis - fly (1)	1	0.0024	0.78
map04217 Necroptosis (2)	2	0.0047	0.59
map04115 p53 signaling pathway (2)	2	0.0047	0.237
map04218 Cellular senescence (1)	1	0.0024	0.889
map04510 Focal adhesion (3)	3	0.0071	0.223
map04520 Adherens junction (1)	1	0.0024	0.605
map04530 Tight junction (2)	2	0.0047	0.574
map04540 Gap junction (2)	2	0.0047	0.313
map04550 Signaling pathways regulating pluripotency of stem cells (1)	1	0.0024	0.502
map02024 Quorum sensing (17)	14	0.0332	0.186
map05111 Biofilm formation - Vibrio cholerae (1)	1	0.0024	0.94
map02025 Biofilm formation - Pseudomonas aeruginosa (5)	4	0.0095	0.514
map02026 Biofilm formation - Escherichia coli (4)	5	0.0118	0.141
map02030 Bacterial chemotaxis (7)	2	0.0047	0.825
map04810 Regulation of actin cytoskeleton (3)	3	0.0071	0.225
map04611 Platelet activation (3)	3	0.0071	0.129

map04620 Toll-like receptor signaling pathway (2)	2	0.0047	0.087
map04621 NOD-like receptor signaling pathway (4)	4	0.0095	0.072
map04625 C-type lectin receptor signaling pathway (2)	2	0.0047	0.191
map04650 Natural killer cell mediated cytotoxicity (2)	2	0.0047	0.117
map04612 Antigen processing and presentation (3)	2	0.0047	0.176
map04660 T cell receptor signaling pathway (2)	2	0.0047	0.259
map04659 Th17 cell differentiation (1)	1	0.0024	0.454
map04657 IL-17 signaling pathway (2)	1	0.0024	0.454
map04662 B cell receptor signaling pathway (2)	2	0.0047	0.128
map04664 Fc epsilon RI signaling pathway (2)	2	0.0047	0.195
map04666 Fc gamma R-mediated phagocytosis (1)	1	0.0024	0.766
map04670 Leukocyte transendothelial migration (2)	2	0.0047	0.17
map04062 Chemokine signaling pathway (3)	3	0.0071	0.108
map04911 Insulin secretion (1)	1	0.0024	0.401
map04910 Insulin signaling pathway (4)	4	0.0095	0.19
map04922 Glucagon signaling pathway (2)	2	0.0047	0.69
map04923 Regulation of lipolysis in adipocytes (2)	2	0.0047	0.072
map04920 Adipocytokine signaling pathway (1)	1	0.0024	0.855
map03320 PPAR signaling pathway (2)	2	0.0047	0.921
map04929 GnRH secretion (1)	1	0.0024	0.424
map04912 GnRH signaling pathway (2)	2	0.0047	0.291
map04913 Ovarian steroidogenesis (1)	1	0.0024	0.367
map04915 Estrogen signaling pathway (5)	4	0.0095	0.049
map04914 Progesterone-mediated oocyte maturation (5)	5	0.0118	0.038
map04917 Prolactin signaling pathway (2)	2	0.0047	0.147
map04921 Oxytocin signaling pathway (2)	2	0.0047	0.611
map04926 Relaxin signaling pathway (4)	4	0.0095	0.024
map04935 Growth hormone synthesis, secretion and action (3)	3	0.0071	0.083
map04918 Thyroid hormone synthesis (2)	2	0.0047	0.185
map04919 Thyroid hormone signaling pathway (5)	5	0.0118	0.035
map04928 Parathyroid hormone synthesis, secretion and action (1)	1	0.0024	0.637
map04916 Melanogenesis (1)	1	0.0024	0.63
map04924 Renin secretion (1)	1	0.0024	0.492

map04925 Aldosterone synthesis and secretion (2)	2	0.0047	0.251
map04927 Cortisol synthesis and secretion (1)	1	0.0024	0.308
map04261 Adrenergic signaling in cardiomyocytes (2)	2	0.0047	0.552
map04270 Vascular smooth muscle contraction (3)	2	0.0047	0.283
map04970 Salivary secretion (2)	2	0.0047	0.177
map04971 Gastric acid secretion (1)	1	0.0024	0.476
map04972 Pancreatic secretion (1)	1	0.0024	0.744
map04976 Bile secretion (2)	2	0.0047	0.363
map04973 Carbohydrate digestion and absorption (1)	1	0.0024	0.592
map04978 Mineral absorption (1)	1	0.0024	0.361
map04962 Vasopressin-regulated water reabsorption (2)	2	0.0047	0.31
map04960 Aldosterone-regulated sodium reabsorption (1)	1	0.0024	0.453
map04961 Endocrine and other factor-regulated calcium reabsorption (3)	3	0.0071	0.057
map04964 Proximal tubule bicarbonate reclamation (1)	1	0.0024	0.453
map04966 Collecting duct acid secretion (1)	1	0.0024	0.44
map04724 Glutamatergic synapse (1)	1	0.0024	0.73
map04727 GABAergic synapse (2)	2	0.0047	0.376
map04725 Cholinergic synapse (2)	2	0.0047	0.203
map04728 Dopaminergic synapse (1)	1	0.0024	0.839
map04726 Serotonergic synapse (1)	1	0.0024	0.671
map04720 Long-term potentiation (1)	1	0.0024	0.684
map04723 Retrograde endocannabinoid signaling (2)	2	0.0047	0.449
map04721 Synaptic vesicle cycle (3)	3	0.0071	0.192
map04722 Neurotrophin signaling pathway (2)	2	0.0047	0.424
map04745 Phototransduction - fly (1)	1	0.0024	0.398
map04740 Olfactory transduction (1)	1	0.0024	0.351
map04742 Taste transduction (1)	1	0.0024	0.41
map04750 Inflammatory mediator regulation of TRP channels (2)	2	0.0047	0.157
map04320 Dorso-ventral axis formation (1)	1	0.0024	0.425
map04360 Axon guidance (1)	1	0.0024	0.624
map04361 Axon regeneration (2)	2	0.0047	0.396
map04380 Osteoclast differentiation (2)	2	0.0047	0.15
map04211 Longevity regulating pathway (2)	2	0.0047	0.403
map04212 Longevity regulating pathway - worm (2)	3	0.0071	0.533

map04213 Longevity regulating pathway - multiple species (2)	2	0.0047	0.711
map04713 Circadian entrainment (1)	1	0.0024	0.558
map04714 Thermogenesis (7)	7	0.0166	0.328
map04626 Plant-pathogen interaction (5)	4	0.0095	0.04
map05200 Pathways in cancer (9)	9	0.0213	0.045
map05202 Transcriptional misregulation in cancer (2)	2	0.0047	0.313
map05206 MicroRNAs in cancer (3)	3	0.0071	0.394
map05205 Proteoglycans in cancer (5)	5	0.0118	0.046
map05204 Chemical carcinogenesis (3)	4	0.0095	0.142
map05203 Viral carcinogenesis (3)	3	0.0071	0.512
map05230 Central carbon metabolism in cancer (2)	2	0.0047	0.587
map05231 Choline metabolism in cancer (2)	2	0.0047	0.43
map05235 PD-L1 expression and PD-1 checkpoint pathway in cancer (1)	1	0.0024	0.612
map05210 Colorectal cancer (2)	2	0.0047	0.342
map05212 Pancreatic cancer (1)	1	0.0024	0.626
map05225 Hepatocellular carcinoma (5)	6	0.0142	0.064
map05226 Gastric cancer (2)	2	0.0047	0.449
map05214 Glioma (2)	2	0.0047	0.332
map05221 Acute myeloid leukemia (2)	2	0.0047	0.17
map05220 Chronic myeloid leukemia (2)	2	0.0047	0.179
map05218 Melanoma (1)	1	0.0024	0.468
map05211 Renal cell carcinoma (3)	3	0.0071	0.07
map05215 Prostate cancer (4)	3	0.0071	0.126
map05213 Endometrial cancer (2)	2	0.0047	0.214
map05224 Breast cancer (2)	2	0.0047	0.298
map05222 Small cell lung cancer (2)	2	0.0047	0.2
map05223 Non-small cell lung cancer (2)	2	0.0047	0.191
map05323 Rheumatoid arthritis (3)	2	0.0047	0.198
map05010 Alzheimer disease (2)	2	0.0047	0.841
map05012 Parkinson disease (3)	3	0.0071	0.563
map05016 Huntington disease (4)	4	0.0095	0.786
map05017 Spinocerebellar ataxia (2)	2	0.0047	0.388
map05020 Prion diseases (1)	1	0.0024	0.477
map05030 Cocaine addiction (1)	1	0.0024	0.45
map05031 Amphetamine addiction (1)	1	0.0024	0.682
map05032 Morphine addiction (1)	1	0.0024	0.376
map05034 Alcoholism (2)	2	0.0047	0.552
map05418 Fluid shear stress and atherosclerosis (7)	7	0.0166	0.027
map05410 Hypertrophic cardiomyopathy (HCM) (1)	1	0.0024	0.506

map05412 Arrhythmogenic right ventricular cardiomyopathy (ARVC) (2)	2	0.0047	0.099
map05414 Dilated cardiomyopathy (DCM) (2)	2	0.0047	0.131
map05416 Viral myocarditis (1)	1	0.0024	0.433
map04930 Type II diabetes mellitus (1)	1	0.0024	0.458
map04932 Non-alcoholic fatty liver disease (NAFLD) (2)	2	0.0047	0.658
map04931 Insulin resistance (2)	2	0.0047	0.488
map04933 AGE-RAGE signaling pathway in diabetic complications (1)	1	0.0024	0.524
map04934 Cushing syndrome (2)	2	0.0047	0.412
map05110 Vibrio cholerae infection (3)	3	0.0071	0.193
map05120 Epithelial cell signaling in Helicobacter pylori infection (1)	1	0.0024	0.716
map05130 Pathogenic Escherichia coli infection (3)	3	0.0071	0.433
map05132 Salmonella infection (6)	6	0.0142	0.114
map05131 Shigellosis (2)	2	0.0047	0.817
map05135 Yersinia infection (3)	3	0.0071	0.143
map05133 Pertussis (3)	3	0.0071	0.249
map05134 Legionellosis (2)	2	0.0047	0.495
map05152 Tuberculosis (3)	4	0.0095	0.203
map05100 Bacterial invasion of epithelial cells (3)	3	0.0071	0.074
map05166 Human T-cell leukemia virus 1 infection (3)	3	0.0071	0.501
map05170 Human immunodeficiency virus 1 infection (2)	2	0.0047	0.671
map05162 Measles (2)	2	0.0047	0.384
map05164 Influenza A (2)	2	0.0047	0.544
map05161 Hepatitis B (2)	2	0.0047	0.416
map05160 Hepatitis C (2)	2	0.0047	0.528
map05168 Herpes simplex virus 1 infection (3)	3	0.0071	0.162
map05163 Human cytomegalovirus infection (3)	3	0.0071	0.226
map05167 Kaposi sarcoma-associated herpesvirus infection (1)	1	0.0024	0.781
map05169 Epstein-Barr virus infection (2)	2	0.0047	0.621
map05165 Human papillomavirus infection (4)	4	0.0095	0.486
map05146 Amoebiasis (3)	4	0.0095	0.016
map05145 Toxoplasmosis (1)	1	0.0024	0.766
map05142 Chagas disease (American trypanosomiasis) (1)	1	0.0024	0.622
map05143 African trypanosomiasis (1)	1	0.0024	0.305
map01501 beta-Lactam resistance (10)	7	0.0166	0.129

map01503 Cationic antimicrobial peptide (CAMP) resistance (4)	4	0.0095	0.312
map01521 EGFR tyrosine kinase inhibitor resistance (2)	2	0.0047	0.279
map01524 Platinum drug resistance (4)	7	0.0166	0.011
map01523 Antifolate resistance (1)	1	0.0024	0.662
map01522 Endocrine resistance (3)	3	0.0071	0.093

Table 7: pathways returned as being detected in the acetate gene set. The first column describes the pathway id, its name and, between brackets, the number of ko retrieved for this particular pathway. The second column gives the total number of genes identified in the acetate gene set with one or more ko related to this particular pathway. The third column gives, as a fraction, the number of genes retrieved as being part of this pathway among all the genes from the acetate gene set (422 genes). The last column shows the p-value computed by random distribution method (see material and method) to detect if the number of genes related to this pathway was bigger than it would be due to a random event.

Definition	Ngenes Per pathway	Fraction Genes enriched	likelihood
map01100 Metabolic pathways (93)	60	NA	NA
map01110 Biosynthesis of secondary metabolites (38)	32	0.1839	0.001
map01120 Microbial metabolism in diverse environments (31)	21	0.1207	0.166
map01130 Biosynthesis of antibiotics (26)	21	0.1207	0.035
map01200 Carbon metabolism (15)	11	0.0632	0.134
map01212 Fatty acid metabolism (7)	12	0.0690	0.001
map01230 Biosynthesis of amino acids (14)	9	0.0517	0.079
map01220 Degradation of aromatic compounds (6)	6	0.0345	0.077
map00010 Glycolysis / Gluconeogenesis (4)	4	0.0230	0.376
map00020 Citrate cycle (TCA cycle) (3)	2	0.0115	0.451
map00030 Pentose phosphate pathway (2)	3	0.0172	0.231
map00040 Pentose and glucuronate interconversions (4)	3	0.0172	0.182
map00051 Fructose and mannose metabolism (3)	1	0.0057	0.871
map00052 Galactose metabolism (1)	1	0.0057	0.83
map00053 Ascorbate and aldarate metabolism (4)	4	0.0230	0.036
map00500 Starch and sucrose metabolism (2)	2	0.0115	0.681
map00520 Amino sugar and nucleotide sugar metabolism (9)	4	0.0230	0.328
map00620 Pyruvate metabolism (6)	5	0.0287	0.244
map00630 Glyoxylate and dicarboxylate metabolism (3)	3	0.0172	0.676
map00640 Propanoate metabolism (3)	3	0.0172	0.646
map00650 Butanoate metabolism (10)	8	0.0460	0.026
map00562 Inositol phosphate metabolism (4)	3	0.0172	0.047
map00190 Oxidative phosphorylation (5)	3	0.0172	0.43

map00710 Carbon fixation in photosynthetic organisms (2)	3	0.0172	0.042
map00720 Carbon fixation pathways in prokaryotes (7)	5	0.0287	0.042
map00680 Methane metabolism (3)	3	0.0172	0.452
map00910 Nitrogen metabolism (1)	1	0.0057	0.771
map00061 Fatty acid biosynthesis (3)	8	0.0460	0.003
map00071 Fatty acid degradation (7)	9	0.0517	0.006
map00121 Secondary bile acid biosynthesis (2)	1	0.0057	0.121
map00140 Steroid hormone biosynthesis (1)	1	0.0057	0.328
map00561 Glycerolipid metabolism (3)	3	0.0172	0.197
map00564 Glycerophospholipid metabolism (3)	3	0.0172	0.177
map00600 Sphingolipid metabolism (1)	1	0.0057	0.493
map00230 Purine metabolism (5)	4	0.0230	0.387
map00240 Pyrimidine metabolism (1)	1	0.0057	0.843
map00250 Alanine, aspartate and glutamate metabolism (1)	1	0.0057	0.814
map00260 Glycine, serine and threonine metabolism (2)	2	0.0115	0.795
map00270 Cysteine and methionine metabolism (1)	1	0.0057	0.885
map00280 Valine, leucine and isoleucine degradation (4)	5	0.0287	0.28
map00310 Lysine degradation (5)	4	0.0230	0.248
map00220 Arginine biosynthesis (1)	1	0.0057	0.709
map00330 Arginine and proline metabolism (2)	2	0.0115	0.687
map00340 Histidine metabolism (4)	3	0.0172	0.118
map00350 Tyrosine metabolism (4)	4	0.0230	0.163
map00360 Phenylalanine metabolism (5)	5	0.0287	0.091
map00380 Tryptophan metabolism (3)	3	0.0172	0.513
map00400 Phenylalanine, tyrosine and tryptophan biosynthesis (10)	5	0.0287	0.001
map00410 beta-Alanine metabolism (3)	3	0.0172	0.32
map00440 Phosphonate and phosphinate metabolism (1)	1	0.0057	0.332
map00450 Selenocompound metabolism (1)	1	0.0057	0.576
map00510 N-Glycan biosynthesis (1)	1	0.0057	0.498
map00513 Various types of N-glycan biosynthesis (1)	1	0.0057	0.418
map00603 Glycosphingolipid biosynthesis - globo and isoglobo series (1)	1	0.0057	0.237
map00730 Thiamine metabolism (1)	1	0.0057	0.556
map00750 Vitamin B6 metabolism (1)	1	0.0057	0.309
map00760 Nicotinate and nicotinamide metabolism (3)	3	0.0172	0.136

map00770 Pantothenate and CoA biosynthesis (1)	1	0.0057	0.614
map00780 Biotin metabolism (1)	4	0.0230	0.049
map00830 Retinol metabolism (1)	1	0.0057	0.521
map00860 Porphyrin and chlorophyll metabolism (3)	1	0.0057	0.781
map00130 Ubiquinone and other terpenoid-quinone biosynthesis (2)	2	0.0115	0.238
map00900 Terpenoid backbone biosynthesis (1)	1	0.0057	0.623
map00906 Carotenoid biosynthesis (1)	1	0.0057	0.28
map00981 Insect hormone biosynthesis (1)	1	0.0057	0.396
map00903 Limonene and pinene degradation (3)	3	0.0172	0.071
map00281 Geraniol degradation (3)	4	0.0230	0.018
map01051 Biosynthesis of ansamycins (1)	2	0.0115	0.031
map00960 Tropane, piperidine and pyridine alkaloid biosynthesis (2)	2	0.0115	0.095
map00401 Novobiocin biosynthesis (1)	1	0.0057	0.205
map00333 Prodigiosin biosynthesis (1)	4	0.0230	0.022
map00998 Biosynthesis of various secondary metabolites - part 2 (2)	2	0.0115	0.029
map00362 Benzoate degradation (7)	7	0.0402	0.017
map00627 Aminobenzoate degradation (2)	2	0.0115	0.559
map00364 Fluorobenzoate degradation (1)	1	0.0057	0.364
map00625 Chloroalkane and chloroalkene degradation (2)	2	0.0115	0.375
map00361 Chlorocyclohexane and chlorobenzene degradation (1)	1	0.0057	0.479
map00623 Toluene degradation (3)	3	0.0172	0.05
map00622 Xylene degradation (1)	1	0.0057	0.411
map00791 Atrazine degradation (1)	1	0.0057	0.234
map00930 Caprolactam degradation (2)	2	0.0115	0.193
map00626 Naphthalene degradation (1)	1	0.0057	0.561
map00980 Metabolism of xenobiotics by cytochrome P450 (1)	1	0.0057	0.698
map00982 Drug metabolism - cytochrome P450 (1)	1	0.0057	0.722
map00983 Drug metabolism - other enzymes (1)	1	0.0057	0.743
map03010 Ribosome (1)	1	0.0057	0.872
map00970 Aminoacyl-tRNA biosynthesis (2)	2	0.0115	0.388
map03008 Ribosome biogenesis in eukaryotes (1)	1	0.0057	0.639
map04141 Protein processing in endoplasmic reticulum (1)	1	0.0057	0.802

map04122 Sulfur relay system (2)	2	0.0115	0.077
map03050 Proteasome (1)	1	0.0057	0.536
map03420 Nucleotide excision repair (1)	1	0.0057	0.647
map03430 Mismatch repair (1)	1	0.0057	0.559
map02010 ABC transporters (28)	11	0.0632	0.066
map02060 Phosphotransferase system (PTS) (4)	1	0.0057	0.219
map03070 Bacterial secretion system (2)	2	0.0115	0.503
map02020 Two-component system (15)	12	0.0690	0.053
map04014 Ras signaling pathway (1)	1	0.0057	0.445
map04015 Rap1 signaling pathway (1)	1	0.0057	0.365
map04010 MAPK signaling pathway (1)	1	0.0057	0.469
map04012 ErbB signaling pathway (1)	1	0.0057	0.273
map04370 VEGF signaling pathway (1)	1	0.0057	0.242
map04064 NF-kappa B signaling pathway (1)	1	0.0057	0.118
map04066 HIF-1 signaling pathway (1)	1	0.0057	0.531
map04020 Calcium signaling pathway (2)	1	0.0057	0.443
map04070 Phosphatidylinositol signaling system (3)	2	0.0115	0.076
map04072 Phospholipase D signaling pathway (1)	1	0.0057	0.364
map04152 AMPK signaling pathway (1)	1	0.0057	0.641
map04146 Peroxisome (1)	3	0.0172	0.281
map04140 Autophagy - animal (2)	2	0.0115	0.266
map04110 Cell cycle (1)	1	0.0057	0.664
map04111 Cell cycle - yeast (1)	1	0.0057	0.754
map04216 Ferroptosis (1)	3	0.0172	0.027
map04115 p53 signaling pathway (1)	1	0.0057	0.322
map04218 Cellular senescence (1)	1	0.0057	0.521
map02024 Quorum sensing (11)	10	0.0575	0.025
map05111 Biofilm formation - Vibrio cholerae (4)	4	0.0230	0.028
map02025 Biofilm formation - Pseudomonas aeruginosa (1)	1	0.0057	0.765
map02030 Bacterial chemotaxis (7)	5	0.0287	0.011
map02040 Flagellar assembly (1)	1	0.0057	0.479
map04650 Natural killer cell mediated cytotoxicity (1)	1	0.0057	0.206
map04660 T cell receptor signaling pathway (1)	1	0.0057	0.288
map04658 Th1 and Th2 cell differentiation (1)	1	0.0057	0.157
map04659 Th17 cell differentiation (1)	1	0.0057	0.211
map04664 Fc epsilon RI signaling pathway (1)	1	0.0057	0.214
map04666 Fc gamma R-mediated phagocytosis (1)	1	0.0057	0.418
map04670 Leukocyte transendothelial migration (1)	1	0.0057	0.227

map04920 Adipocytokine signaling pathway (2)	4	0.0230	0.008
map03320 PPAR signaling pathway (4)	6	0.0345	0.007
map04921 Oxytocin signaling pathway (1)	1	0.0057	0.537
map04935 Growth hormone synthesis, secretion and action (1)	1	0.0057	0.289
map04919 Thyroid hormone signaling pathway (2)	1	0.0057	0.489
map04722 Neurotrophin signaling pathway (1)	1	0.0057	0.407
map04750 Inflammatory mediator regulation of TRP channels (1)	1	0.0057	0.238
map04360 Axon guidance (1)	1	0.0057	0.372
map04361 Axon regeneration (1)	1	0.0057	0.44
map04211 Longevity regulating pathway (1)	1	0.0057	0.447
map04714 Thermogenesis (1)	3	0.0172	0.384
map04626 Plant-pathogen interaction (1)	1	0.0057	0.38
map05200 Pathways in cancer (1)	1	0.0057	0.813
map05206 MicroRNAs in cancer (1)	1	0.0057	0.534
map05205 Proteoglycans in cancer (1)	1	0.0057	0.521
map05231 Choline metabolism in cancer (1)	1	0.0057	0.423
map05235 PD-L1 expression and PD-1 checkpoint pathway in cancer (1)	1	0.0057	0.297
map05225 Hepatocellular carcinoma (1)	1	0.0057	0.635
map05214 Glioma (1)	1	0.0057	0.319
map05223 Non-small cell lung cancer (1)	1	0.0057	0.23
map05340 Primary immunodeficiency (1)	1	0.0057	0.085
map05034 Alcoholism (1)	1	0.0057	0.544
map04933 AGE-RAGE signaling pathway in diabetic complications (2)	1	0.0057	0.237
map05110 Vibrio cholerae infection (1)	1	0.0057	0.377
map05120 Epithelial cell signaling in Helicobacter pylori infection (1)	1	0.0057	0.416
map05130 Pathogenic Escherichia coli infection (1)	1	0.0057	0.543
map05131 Shigellosis (2)	1	0.0057	0.693
map05135 Yersinia infection (1)	1	0.0057	0.366
map05134 Legionellosis (1)	1	0.0057	0.491
map05166 Human T-cell leukemia virus 1 infection (1)	1	0.0057	0.671
map05170 Human immunodeficiency virus 1 infection (1)	1	0.0057	0.591
map05167 Kaposi sarcoma-associated herpesvirus infection (1)	1	0.0057	0.399
map01501 beta-Lactam resistance (2)	2	0.0115	0.51
map01521 EGFR tyrosine kinase inhibitor resistance (1)	1	0.0057	0.319

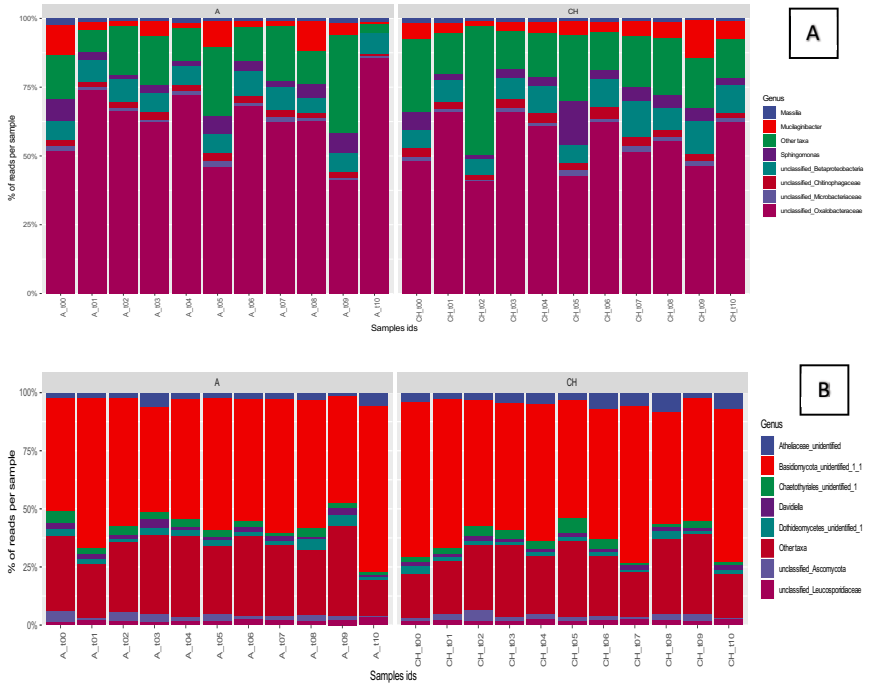


*Figure 3: **A**: Bar plot showing the relative abundance of the eight most abundant bacterial genus observed across both microcosm time series. The taxonomy is based on the classification of the ASV from the 16S rRNA sequencing data using RDP classifier. The minor bacterial genus represented by a low abundance of sequences in the different samples have been summed up and termed "Other taxa". **B**: Bar plot showing the relative abundance of the most abundant fungal orders observed across both microcosm time series. The taxonomy is based on the classification of the ASV from the ITS sequencing data using RDP classifier. The minor fungal orders represented by a low abundance of sequences in the different samples have been summed up and termed "Other taxa".*

Table 8: Taxonomy table of the core ASV (ASV appearing in at least 22 samples of the considered time series). The first column shows the ASV name displayed on the networks in the main article. The two last columns show the presence (X) or absence (-) of the ASV in each core community from both time series. Note: Only the ASVs being present in both core communities were used to build the core networks in the article.

ASV name	domain	phylum	class	order	family	genus	species	waterCore	acetateCore
OTU_1	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_10	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Mucilaginibacter	Unclassified	X	X
OTU_10100	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_11	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Unclassified	X	X
OTU_11000	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_111	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	unclassified_Microbacteriaceae	Unclassified	X	-
OTU_11610	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	-	X
OTU_117	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	unclassified_Chitinophagaceae	Unclassified	X	-
OTU_119	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_12	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium	Unclassified	X	X
OTU_12100	Fungi	Basidiomycota	Agaricomycetes	Atheliales	Atheliaceae	Atheliaceae_unidentified	Atheliaceae_sp|SH232729.06FU	X	X
OTU_12110	Fungi	Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	unclassified_Leucosporidiaceae	unclassified_Leucosporidiaceae	-	X
OTU_12310	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH219414.06FU	X	X
OTU_129	Bacteria	Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	Hymenobacter	Unclassified	X	X
OTU_12910	Fungi	Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	X	X
OTU_13100	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_134	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_138	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_14	Bacteria	Proteobacteria	Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	Unclassified	X	X
OTU_14100	Fungi	Ascomycota	Lecanoromycetes	Lecanorales	Parmeliaceae	Pseudevernia	Pseudevernia_furfuracea|SH230099.06FU	X	X
OTU_15	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	unclassified_Microbacteriaceae	Unclassified	X	X
OTU_150	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia	Unclassified	X	X
OTU_15100	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_16	Bacteria	Proteobacteria	Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	Unclassified	X	X
OTU_161	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_16100	Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Chaetothyriales_unidentified	Chaetothyriales_unidentified_1	Chaetothyriales_sp|SH228288.06FU	X	X
OTU_16510	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_168	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Subtercola	Unclassified	X	X
OTU_17	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia	Unclassified	X	X
OTU_17010	Fungi	Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	X	-
OTU_17100	Fungi	Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Davidiella	Davidiella_tassiana|SH196750.06FU	X	X
OTU_18	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Rhizobacter	Unclassified	X	X
OTU_18100	Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	unclassified_Herpotrichiellaceae	unclassified_Herpotrichiellaceae	X	X
OTU_186	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Rhizobacter	Unclassified	X	X
OTU_19	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Unclassified	X	X
OTU_19100	Fungi	Ascomycota	Lecanoromycetes	Lecanorales	Parmeliaceae	Pseudevernia	Pseudevernia_furfuracea|SH230099.06FU	X	-
OTU_2	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_20	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_20100	Fungi	Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	unclassified_Leucosporidiaceae	unclassified_Leucosporidiaceae	X	X
OTU_21	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Unclassified	X	X
OTU_21000	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_21100	Fungi	Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	unclassified_Leucosporidiaceae	unclassified_Leucosporidiaceae	X	X
OTU_22	Bacteria	Proteobacteria	Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	Unclassified	X	X
OTU_22100	Fungi	Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	unclassified_Leucosporidiaceae	unclassified_Leucosporidiaceae	X	X
OTU_23	Bacteria	Proteobacteria	Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	Unclassified	X	X
OTU_23100	Fungi	Ascomycota	Dothideomycetes	Pleosporales	Venturiaceae_1	unclassified_Venturiaceae_1	unclassified_Venturiaceae_1	X	X
OTU_23610	Fungi	unclassified_Fungi	unclassified_Fungi	unclassified_Fungi	unclassified_Fungi	unclassified_Fungi	unclassified_Fungi	X	-
OTU_25100	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_26	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Unclassified	X	X
OTU_260	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified_Burkholderiales	unclassified_Burkholderiales	Unclassified	X	X
OTU_26100	Fungi	Basidiomycota	Agaricomycetes	Atheliales	Atheliaceae	Atheliaceae_unidentified	Atheliaceae_sp|SH232729.06FU	X	X
OTU_27	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	unclassified_Chitinophagaceae	Unclassified	X	X
OTU_27100	Fungi	Ascomycota	Dothideomycetes	Dothideales	Incertae_sedis_22	Celosporium	Celosporium_sp|SH231451.06FU	X	X
OTU_272	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	-	X
OTU_28	Bacteria	Proteobacteria	Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	Unclassified	X	X
OTU_286	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_29100	Fungi	Ascomycota	unclassified_Ascomycota	unclassified_Ascomycota	unclassified_Ascomycota	unclassified_Ascomycota	unclassified_Ascomycota	X	X
OTU_3	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_30	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_31	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Unclassified	X	X
OTU_31000	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_31100	Fungi	Ascomycota	Lecanoromycetes	Lecanorales	Lecanorales_unidentified	Lecanorales_unidentified_1	Lecanorales_sp|SH227434.06FU	X	X

OTU_32	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	Ferruginibacter	Unclassified	X	X
OTU_32100	Fungi	Ascomycota	Dothideomycetes	Dothideales	Dothideaceae	Endoconidioma	Endoconidioma_populi|SH231447.06FU	X	X
OTU_33	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium	Unclassified	X	X
OTU_33100	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_332	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_34100	Fungi	Basidiomycota	Microbotryomycetes	Leucosporidiales	unclassified_Leucosporidiales	unclassified_Leucosporidiales	unclassified_Leucosporidiales	X	X
OTU_35	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	unclassified_Microbacteriaceae	Unclassified	X	X
OTU_35100	Fungi	Basidiomycota	Agaricomycetes	Polyporales	Polyporales_unidentified	Polyporales_unidentified_1	Polyporales_sp|SH202320.06FU	X	X
OTU_36	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium	Unclassified	X	X
OTU_364	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	-
OTU_37	Bacteria	Acidobacteria	Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	Granulicella	Unclassified	X	X
OTU_37100	Fungi	Basidiomycota	Microbotryomycetes	Leucosporidiales	unclassified_Leucosporidiales	unclassified_Leucosporidiales	unclassified_Leucosporidiales	X	X
OTU_38	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Polaromonas	Unclassified	X	X
OTU_387	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	-
OTU_39	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia	Unclassified	X	X
OTU_39100	Fungi	unclassified_Fungi	unclassified_Fungi	unclassified_Fungi	unclassified_Fungi	unclassified_Fungi	unclassified_Fungi	X	X
OTU_393	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified_Burkholderiales	unclassified_Burkholderiales	Unclassified	X	X
OTU_4	Bacteria	Proteobacteria	Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	Unclassified	X	X
OTU_40	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia	Unclassified	X	X
OTU_40100	Fungi	Ascomycota	Dothideomycetes	Venturiales	Venturiaceae	Venturiaceae_unidentified	Venturiaceae_sp|SH238426.06FU	-	X
OTU_41	Bacteria	Acidobacteria	Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	Granulicella	Unclassified	X	X
OTU_41000	Fungi	Basidiomycota	Agaricomycetes	Atheliales	Atheliaceae	Atheliaceae_unidentified	Atheliaceae_sp|SH232729.06FU	X	X
OTU_41100	Fungi	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Mycosphaerellaceae_unidentified	Mycosphaerellaceae_sp|SH238976.06FU	X	X
OTU_42	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Mucilaginibacter	Unclassified	X	X
OTU_43	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	unclassified_Chitinophagaceae	Unclassified	X	X
OTU_43100	Fungi	Ascomycota	Dothideomycetes	Dothideomycetes_unidentified	Dothideomycetes_unidentified_1	Dothideomycetes_unidentified_1	Dothideomycetes_sp|SH231472.06FU	X	X
OTU_44	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Luteibacter	Unclassified	X	X
OTU_45	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	unclassified_Rhizobiales	unclassified_Rhizobiales	Unclassified	X	X
OTU_45100	Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	unclassified_Chaetothyriales	unclassified_Chaetothyriales	unclassified_Chaetothyriales	X	X
OTU_46	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Burkholderia	Unclassified	X	X
OTU_47	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	unclassified_Rhizobiales	unclassified_Rhizobiales	Unclassified	X	X
OTU_47100	Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Herpotrichiellaceae_unidentified	Herpotrichiellaceae_sp|SH241308.06FU	X	X
OTU_48	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	Unclassified	-	X
OTU_48100	Fungi	Fungi_unidentified	Fungi_unidentified_1	Fungi_unidentified_1	Fungi_unidentified_1_1	Fungi_unidentified_1_1	Fungi_sp|SH216411.06FU	X	X
OTU_49	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_5	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_51000	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_52100	Fungi	Basidiomycota	Microbotryomycetes	Sporidiobolales	Incertae_sedis_25	Rhodotorula	Rhodotorula_sp_TP_Snow_Y129|SH212318.06FU	X	X
OTU_55	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Mucilaginibacter	Unclassified	X	X
OTU_55100	Fungi	Ascomycota	Leotiomycetes	Helotiales	Incertae_sedis_2	Incertae_sedis_2_unidentified	Helotiales_sp|SH234732.06FU	X	X
OTU_56100	Fungi	Basidiomycota	Agaricomycetes	Agaricales	Entolomataceae	Clitopilus	unclassified_Clitopilus	X	X
OTU_58	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_58100	Fungi	Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	unclassified_Leucosporidiaceae	unclassified_Leucosporidiaceae	X	X
OTU_6	Bacteria	Proteobacteria	Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	unclassified_Betaproteobacteria	Unclassified	X	X
OTU_60	Bacteria	Acidobacteria	Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	Unclassified	X	-
OTU_60100	Fungi	Ascomycota	Dothideomycetes	unclassified_Dothideomycetes	unclassified_Dothideomycetes	unclassified_Dothideomycetes	unclassified_Dothideomycetes	-	X
OTU_61	Bacteria	Acidobacteria	Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	unclassified_Acidobacteria_Gp1	Terriglobus	Unclassified	X	-
OTU_61000	Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Chaetothyriales_unidentified	Chaetothyriales_unidentified_1	Chaetothyriales_sp|SH228288.06FU	X	X
OTU_61100	Fungi	Ascomycota	Sordariomycetes	Lulworthiales	Lulworthiaceae	Zalerion	Zalerion_sp_T2N16c|SH224770.06FU	X	X
OTU_63	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_64100	Fungi	Ascomycota	Leotiomycetes	Helotiales	Helotiales_unidentified	Helotiales_unidentified_1	Helotiales_sp|SH209225.06FU	-	X
OTU_65	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiales_incertae_sedis	unclassified_Burkholderiales_incertae_sedis	Unclassified	X	-
OTU_66	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Polaromonas	Unclassified	X	X
OTU_67	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	Ferruginibacter	Unclassified	X	X
OTU_68	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	unclassified_Rhizobiales	unclassified_Rhizobiales	Unclassified	-	X
OTU_69	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	-	X
OTU_69100	Fungi	Ascomycota	Eurotiomycetes	Incertae_sedis_18	Incertae_sedis_40	Sarcinomyces	Sarcinomyces_crustaceus|SH210411.06FU	X	-
OTU_7	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Mucilaginibacter	Unclassified	X	X
OTU_70	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	unclassified_Sphingomonadaceae	Unclassified	X	X
OTU_70100	Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Udeniomyces	Udeniomyces_sp_XJ_8B3|SH207893.06FU	X	X
OTU_71	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	unclassified_Acetobacteraceae	Unclassified	X	-
OTU_71000	Fungi	Basidiomycota	Basidiomycota_unidentified	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1	Basidiomycota_unidentified_1_1	Basidiomycota_sp|SH216408.06FU	X	X
OTU_74100	Fungi	Basidiomycota	Agaricomycetes	Agaricales	Entolomataceae	Clitopilus	Clitopilus_hobsonii|SH205041.06FU	X	X
OTU_75	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia	Unclassified	X	X
OTU_77	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_77100	Fungi	Ascomycota	unclassified_Ascomycota	unclassified_Ascomycota	unclassified_Ascomycota	unclassified_Ascomycota	unclassified_Ascomycota	-	X
OTU_79	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_81000	Fungi	Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Davidiella	Davidiella_tassiana|SH196750.06FU	X	X

OTU_82100	Fungi	Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	unclassified_Leucosporidiaceae	unclassified_Leucosporidiaceae	X	X
OTU_88100	Fungi	Basidiomycota	Tremellomycetes	Tremellales	Incertae_sedis_12	Cryptococcus_1	Cryptococcus_victoriae|SH198055.06FU	X	-
OTU_89	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	Unclassified	X	-
OTU_89100	Fungi	Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	unclassified_Basidiomycota	-	X
OTU_9	Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	unclassified_Chitinophagaceae	Unclassified	X	X
OTU_90	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	unclassified_Sphingomonadaceae	Unclassified	X	-
OTU_90100	Fungi	Ascomycota	Ascomycota_unidentified	Ascomycota_unidentified_1	Ascomycota_unidentified_1	Ascomycota_unidentified_1_1	Ascomycota_sp|SH222905.06FU	X	-
OTU_92	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified_Oxalobacteraceae	Unclassified	X	X
OTU_92100	Fungi	Basidiomycota	Agaricomycetes	Atheliales	Atheliaceae	Atheliaceae_unidentified	Atheliaceae_sp|SH232729.06FU	X	-
OTU_9340	Fungi	Basidiomycota	Agaricomycetes	Atheliales	Atheliaceae	Byssocorticium	Byssocorticium_sp|SH233176.06FU	X	X

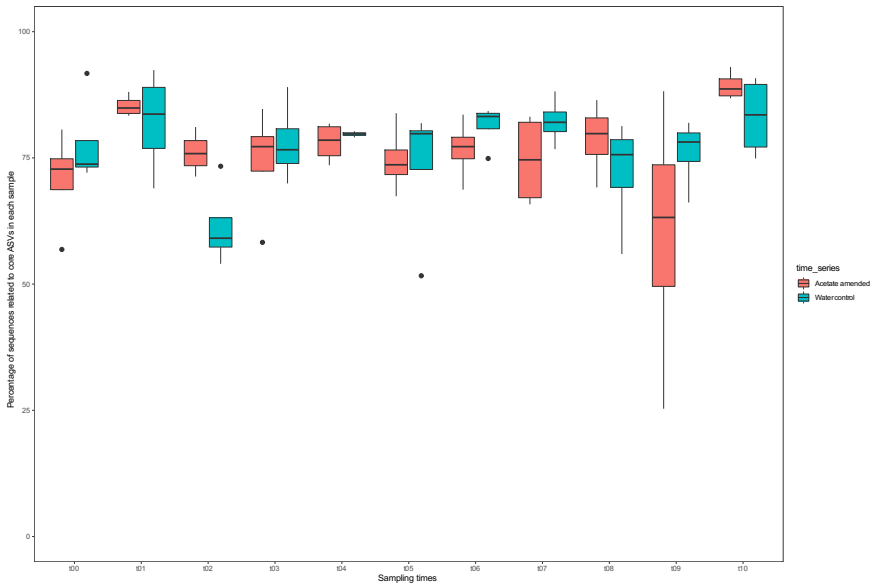


Figure 4: Boxplot showing the relative abundance (as a percentage of total sequences retrieved in the 16S rRNA and ITS sequencing) of the ASV composing the core community in the samples from both time series microcosms. The blue boxplots represent the abundance of the core ASV in the replicates from the water control snow microcosms while the red boxplots display the relative abundance of the core ASV in the acetate amended microcosms. As we can the core community represents more than 50% of the sequence pool in almost all the samples and can sometimes even reach up to 90% in some samples of the last sampling time (t10). Thus, the trends observed in our networks will be representative of the dynamics of a substantial part of the snow communities tracked during this time series experiment.

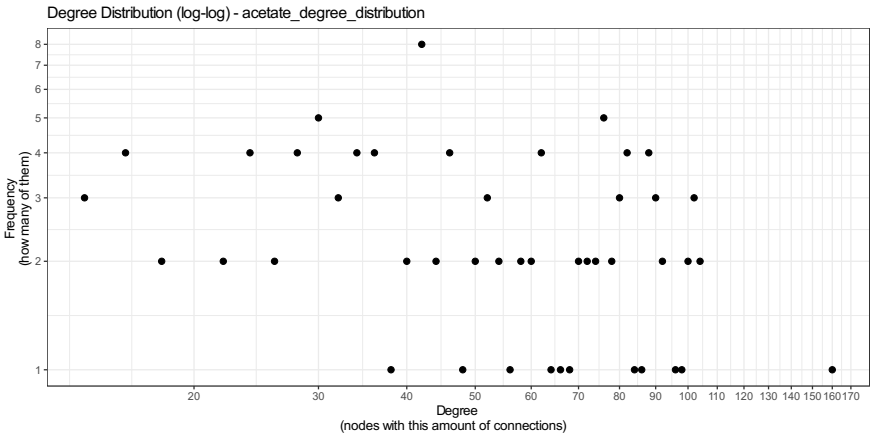


Figure 5: Histogram showing the degree distribution (i.e the number of edges connecting the different nodes = ASV or nutrients) from the acetate amended time series network. The average connectivity of this network is 55.90.

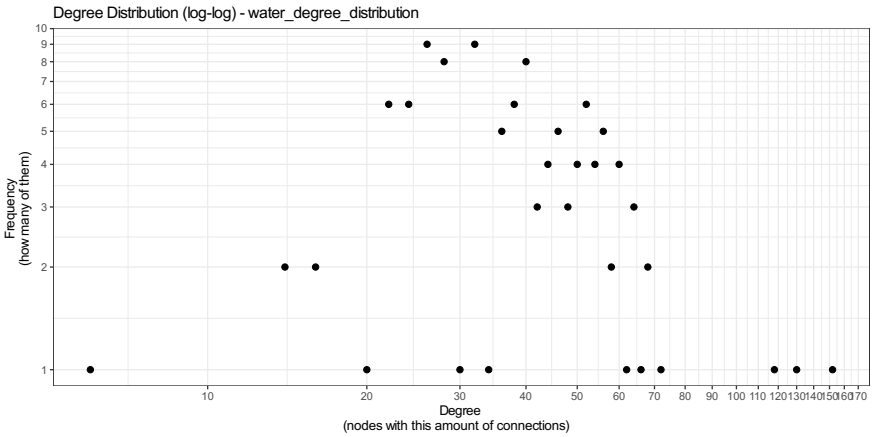


Figure 6: Histogram showing the degree distribution (i.e the number of edges connecting the different nodes = ASV or nutrients) from the water control time series network. The average connectivity of this network is 41.97.

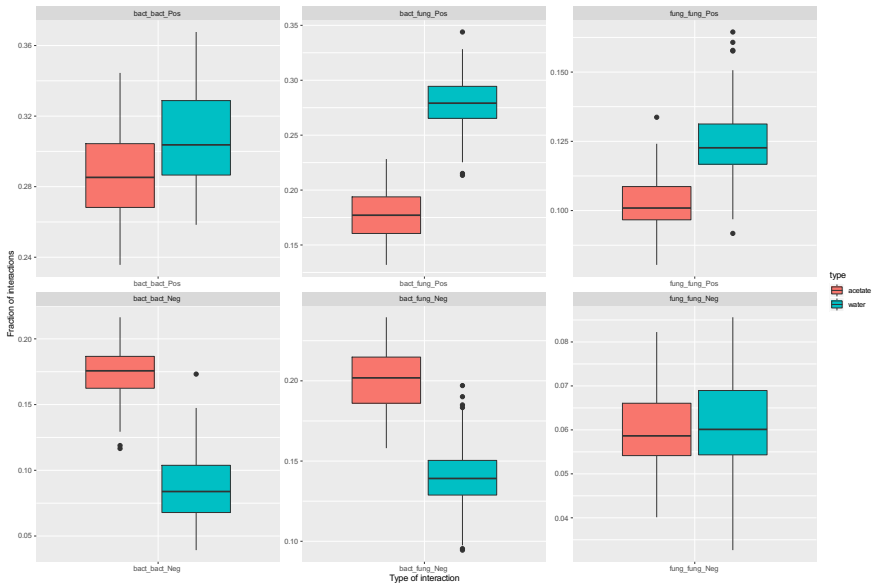


Figure 7: Boxplots showing the fraction of edges (used here as a possible clue of biological interaction) representing different kind of co-variances based on the fact that the sign of the LSA coefficient (Pos = positive LSA value, Neg= negative LSA value) and the taxonomy of the interacting nodes (bact = bacteria, fung= fungi)represented in the subsampled networks. Those networks were built by sampling randomly three replicates among the four replicates present at each time point (100 networks built for each time series). The kind of co-variance is named based on the taxonomy of its interacting nodes and the sign of the LSA coefficient.

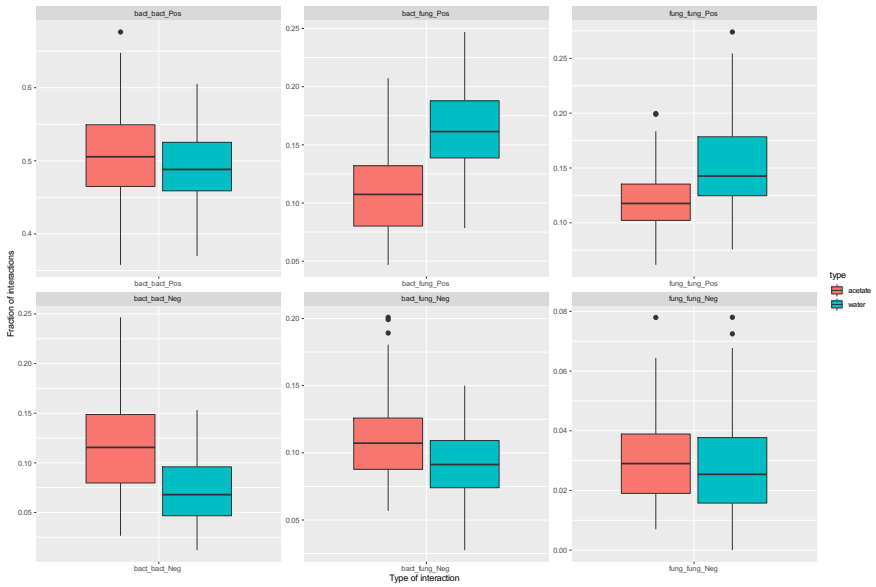


Figure 8: Boxplots showing the fraction of edges (used here as a possible clue of biological interaction) representing different kind of co-variances based on the fact that the sign of the LSA coefficient (Pos = positive LSA value, Neg= negative LSA value) and the taxonomy of the interacting nodes (bact = bacteria, fung= fungi)represented in the subsampled networks. Those networks were built by sampling randomly two replicates among the four replicates present at each time point (100 networks built for each time series). The kind of co-variance is named based on the taxonomy of its interacting nodes and the sign of the LSA coefficient.

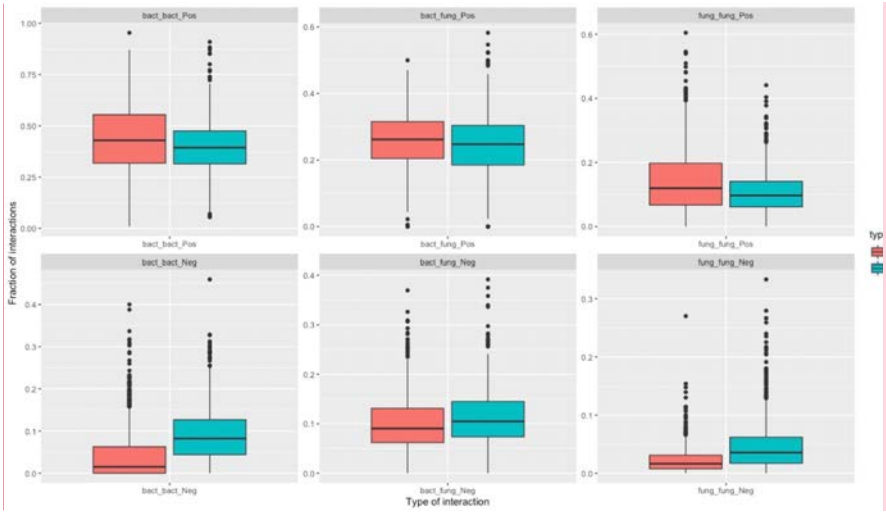


Figure 9: Boxplots showing the fraction of edges (used here as a possible clue of biological interaction) representing different kind of co-variances based on the fact that the sign of the LSA coefficient (Pos = positive LSA value, Neg= negative LSA value) and the taxonomy of the interacting nodes (bact = bacteria, fung= fungi)represented in the subsampled networks. Those networks were built by sampling randomly only one replicate among the four replicates present at each time point (1000 networks built for each time series). The kind of co-variance is named based on the taxonomy of its interacting nodes and the sign of the LSA coefficient.

Commenté [BBP1]: Supp