Comparaison et évaluation d'approches bioinformatiques et statistiques pour l'analyse du pathobiome des plantes cultivées

Charlie Pauvert

▶ **To cite this version:**

Charlie Pauvert. Comparaison et évaluation d'approches bioinformatiques et statistiques pour l'analyse du pathobiome des plantes cultivées. Bio-Informatique, Biologie Systémique [q-bio.QM]. Université de Bordeaux, 2019. Français. NNT : 2019BORD0214 . tel-02452386

HAL Id: tel-02452386
https://theses.hal.science/tel-02452386

Submitted on 23 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR

DE L'UNIVERSITÉ DE BORDEAUX

ECOLE DOCTORALE SCIENCES ET ENVIRONNEMENTS

ECOLOGIE ÉVOLUTIVE, FONCTIONNELLE ET DES COMMUNAUTÉS

Par **Charlie PAUVERT**

Comparaison et évaluation d'approches bioinformatiques et statistiques pour l'analyse du pathobiome des plantes cultivées

Sous la direction de : **Corinne VACHER**
Co-directrice : **Jessica VALLANCE**

Soutenue le 12 novembre 2019

Membres du jury :

Mme. Marie-Laure MARTIN-MAGNIETTE	Directrice de Recherche	AgroParisTech/INRA	Rapporteur
M. Christophe MOUGEL	Directeur de Recherche	INRA Rennes	Rapporteur
M. Thierry CANDRESSE	Directeur de Recherche	INRA Bordeaux	Président
Mme. Chloé DELMAS	Chargée de Recherche	INRA Bordeaux	Examinatrice
M. Frédéric MAHÉ	Ingénieur de Recherche	CIRAD Montpellier	Examinateur
Mme. Corinne VACHER	Directrice de Recherche	INRA Bordeaux	Directrice
Mme. Jessica VALLANCE	Ingénieure de Recherche	Bordeaux Sciences Agro	Invitée

Comparaison et évaluation d'approches bioinformatiques et statistiques pour l'analyse du pathobiome des plantes cultivées

Résumé : Les interactions entre micro-organismes sous-tendent de nombreux services écosystémiques, y compris la régulation des maladies des plantes cultivées. Un acteur de cette régulation est le pathobiome, défini comme le sous-ensemble des micro-organismes associés à une plante hôte et en interaction avec un agent pathogène. L'un des défis actuels consiste à reconstruire les pathobiomes à partir de données de metabarcoding, pour identifier des agents potentiels de biocontrôle et pour surveiller en temps réel leurs réponses aux changements environnementaux. Plusieurs verrous méthodologiques doivent cependant être levés pour atteindre ces objectifs. Tout d'abord, il n'existe pas de consensus concernant l'approche bioinformatique la plus fiable pour déterminer l'identité et l'abondance des micro-organismes présents dans les échantillons végétaux. De plus, les réseaux microbiens construits avec les méthodes actuellement disponibles sont des réseaux d'associations statistiques entre des comptages de séquences, non directement superposables aux réseaux d'interactions écologiques entre micro-organismes. L'objectif de la thèse était donc de déterminer les approches bioinformatiques et statistiques les plus pertinentes pour reconstruire des réseaux d'interactions microbiennes à partir de données de métabarcoding. Le modèle biologique était la vigne (*Vitis vinifera*) et l'oïdium de la vigne, *Erysiphe necator*. Nous avons tout d'abord déterminé l'approche bioinformatique la plus adaptée pour caractériser les communautés fongiques associées aux plantes, en comparant la capacité de 360 pipelines à retrouver la composition d'une communauté artificielle de 189 souches fongiques. DADA2 est apparu comme l'outil le plus performant. Nous avons ensuite évalué l'influence de la pratique culturale (viticulture conventionnelle vs. biologique) sur les communautés fongiques des feuilles de vigne et évalué le niveau de réplicabilité des réseaux microbiens construits avec une méthode d'inférence classique, SparCC. La réplicabilité était très faible, jetant ainsi un doute sur l'utilité de ces réseaux pour le biocontrôle et la biosurveillance. Nous avons donc utilisé une nouvelle approche statistique, le modèle PLN, qui permet de prendre en compte la variabilité environnementale, pour générer des hypothèses d'interactions au sein du pathobiome d'*Erysiphe necator*. Une fraction de ces hypothèses ont été testée par des co-cultures et par fouille automatique de la littérature. Quelques-unes ont été confirmées, ce qui indiquent que les réseaux microbiens, inférés adéquatement, génèrent des hypothèses plausibles d'interactions biotiques qu'il convient de tester pour développer de nouvelles stratégies de protection des cultures.
Mots-clés : Écologie microbienne, Réseaux écologiques, Métabarcoding, Inférence de réseaux, Biocontrôle, Pathobiome

---

Comparison and evaluation of bioinformatic and statistical approaches for the analysis of the pathobiome of crop plants

Abstract: Interactions among microorganisms underpin many ecosystem services, including the regulation of crop diseases. An actor in this regulation is the pathobiome, defined as the subset of microorganisms associated with a host plant in interaction with a pathogen. One of the current challenges is to reconstruct pathobiomes from metabarcoding data, in order to identify potential biocontrol agents and to monitor in real time their responses to environmental changes. However, several methodological hurdles must be overcome to achieve these objectives. First, there is no consensus on the most reliable bioinformatic approach to determine the identity and abundance of microorganisms present in plant samples. In addition, microbial networks built with currently available methods are networks of statistical associations between sequence counts, not directly related to networks of ecological interactions between microorganisms. The objective of the thesis was therefore to determine the most relevant bioinformatic and statistical approaches to reconstruct microbial interaction networks from metabarcoding data. The biological model was grapevine (*Vitis vinifera*) and the fungal agent of grapevine powdery mildew, *Erysiphe necator*. First, we determined the most appropriate bioinformatic approach to describe fungal communities associated with plants, by comparing the ability of 360 pipelines to recover the composition of an artificial community of 189 fungal strains. DADA2 appeared as the most powerful tool. We then evaluated the influence of the cropping system (conventional vs. organic viticulture) on foliar fungal communities of grapevine and assessed the level of replicability of microbial networks built with a standard inference method, SparCC. Replicability was very low, casting doubt on the usefulness of these networks for biocontrol and biomonitoring. We therefore used a new statistical approach, the PLN model, which allowed us to take into account environmental variability, to generate hypotheses of interactions in the pathobiome of *Erysiphe necator*. Some of these hypothesis were tested by co-cultures and by text-mining. A few were confirmed, indicating that microbial networks, when properly inferred, provide plausible hypotheses of biotic interactions that need to be tested to develop new strategies for crop protection.
Keywords: Microbial ecology, Ecological networks, Metarbarcoding , Network inference, Biocontrol, Pathobiome

---

Unité de recherche

UMR 1202 Biodiversité, Gènes et Communautés (BIOGECO), INRA, 33610 Pessac, France.

Remerciements

Au cours des trois années de traversée de la chaîne de montagnes “LaThèse“ différents pics et vallées doivent être gravis et traversés et ce périple ne se fait pas sans aide et soutien.

Tout d’abord, je remercie les personnes qui ont lu cette thèse, et commençant par celles sans qui elle n’existerait pas, mes directrices de thèse Corinne et Jessica. Je remercie aussi les rapporteurs Christophe Mougel et Marie-Laure Martin-Magniette pour le temps consacré à ce manuscrit. Et je les remercie d’avoir accepté de faire parti du jury aux côtés de Thierry Candresse, Chloé Delmas et Frédéric Mahé. Je remercie également les membres de mon comité de thèse, Frédéric Barraquand, Gregory Gambetta, Isabelle Lesur, Guilherme Martins et Adrien Rusch pour nos discussions scientifiques et leurs conseils avisés. Je remercie également tous les membres de l’unité BIOGECO de m’avoir accueilli ces trois années.

Un grand merci à vous deux Corinne et Jessica pour vos aides respectives avant et pendant la thèse. L’aboutissement de cette thèse et des articles vous doit beaucoup. Merci à toi Jessica, pour la découverte des manips microbio et ses difficultés cachées (Est ce qu’il y a aussi un “dieu de la microbio“ pour que ça fonctionne mieux ?). Un jour j’y crois, j’aurais ta vision “oeil-de-lynx“. Corinne, ton don de synthèse est un modèle, j’aimerais avoir ta capacité à reformuler les phrases emberlificotées que j’ai pu envoyer, les reformuler pour extraire le sens qui s’était sans doute arrêté entre mon cortex et mes mains. Mon optimisme me fait aussi espérer que je saurais un jour gérer le quart de ce que je te vois accomplir en parallèle.

Je remercie également toutes les personnes avec qui j’ai pu collaboré pour l’écriture des articles et particulièrement Isabelle, Marc, Valérie et Véronique pour vos mots d’encouragements et de sympathie.

Je remercie également les personnes qui m’ont formé puisque ce voyage scientifique est plein d’apprentissages. Céline, Patrick, Jessica et Tania pour la biologie moléculaire, Emilie et surtout Marie M. pour vos aides sur la ddPCR. Merci aux “devs“ de PLN ;) Merci à Julien C. et Stéphane R. d’avoir toujours pris le temps de répondre à mes questions et mes *issues* intempestives. Merci Corinne et Thomas de m’avoir aidé pour les analyses statistiques en écologie. Merci à Martine, Jessica et Isabelle D. de m’avoir aidé à faire pousser les acteurs principaux de la thèse. Merci Aline pour la formation planning.

Merci à toutes les personnes de l’étage au B2 pour les anecdotes régulières au cours des repas et la bonne ambiance. Je pense notamment aux membres passés du Bureau (Elsa, Marius des premiers jours), puis Hafiz et Thomas C., Tania F., Célia D., (vive le pierre-feuille-ciseaux), et Juliette (dernière recrue ^^). Merci pour toutes nos discussions (l’origine du pie-chart alors ?), et merci à Tania, Thomas C. et Juliette, mes cobayes de figures. Merci aux voisins Homero, Dèb, Xavier, Alex/Thierry et désolé pour les rires sonores. Merci à Pim et Anne, solaires, sourires, toujours, merci pour l’intégration du tout premier jour, je m’en rappelle encore. On viendra vous voir la-bas. Côté glands, merci Thomas C. pour les initiations aux ramassages comptages et merci Tania pour les “initiation“ broyages. Merci Xavier et Thierry pour les sessions sportives du midi qui défoulent. Merci pour les traquenards (Thomas C., Thierry, Marine et Tania). Merci Sophie et Virgil pour nos discussions impromptues toujours intéressantes. Gracias también a Homeo, Natalia y Marta. Merci Alix Coralie et Thierry, est ce qu’il y a eu des discussions sans mentionner un fromage dans cette pièce ?

Merci également aux personnes de Pierroton qui contribuent à rendre ce lieu agréable au milieu des pins propice à toutes les discussions. Je remercie particulièrement Agathe et Arthur pour m'avoir offert ces moments parenthèses à Pierroton sur la table dehors ou dans votre bureau farci de blagues. Je suis ravi de vous avoir rencontré en thèse :) C'est l'endroit de glisser un bon courage à tous les doctorant.e.s qui vont finir cette année. Bon courage, pensez à Gandalf :"All we have to decide is what to do with the time that is given to us". Je remercie aussi la team JDED2018 pour cette expérience dans les coulisses d'un colloque et surtout vous les filles Betty C., Bettie, Claire-Elise, Manon P.

Je remercie évidemment les colocs qui ont cheminé.e.s avec moi quotidiennement dans le paysage de la thèse, Bruno et Alice pour nos débats scientifiques au cours des repas ou non qui dépassaient souvent le cadre des thèses, Charlotte pour ton soutien (c'est surtout toi qui a "subi" la rédaction) et pour avoir été le cobaye de séances de vulgarisations que j'espère réussies. Merci les broinformaticiens de Rouen (Guigui, Laura, Marion, Clem, Thib et Ben) pour les soutiens à coups de GIF et votre présence. Et merci les bros du 77 (Max, Jo, Marius, Vinz et Débo) pour les déconnades au surf qui paraissent lointaines et les sons partagés pour le soutien. Je suis ravi que cette thèse soit le moment de revoir et recontacter des ami.e.s de plus loin dans le temps, Camille F., Chloé D., Emi C., Bérénice A., Rémy L. Je remercie aussi les zozos de plus loin : Aline, Yukki, Gabriel, Baudoin et Mathilde pour vos retours d'expériences de thèse jamais sans une bière et un bonne tranche de rire saupoudrée de cynisme. Merci Anne-Laure pour tes conseils toujours avisés avant et pendant la thèse. Merci à toi et à Mahendra pour m'avoir donné envie de commencer la thèse. Merci Tania sans qui cette thèse aurait été très différente, évidemment : "Rien n'est absolu" mais il y aurait eu moins de glands broyés, de feuilles pesées sans doute (au moins des kgs), peut être moins de gueules de bois et plus de résultats "innovants" nous ne saurons jamais ;). Collègue nécessaire devenue amie, je souhaite voir ton modèle global un jour publié ^^. Merci à tous ceux qui ont rendu les jours sans avec plus d'avec, que ce soit les pizzas d'Eric, les footings "fractionnés" philosophie morale et zététique avec Thierry et Juliette, les sorties plus longues mais tout aussi bavardes avec Myriam H., les *sparring partners* de la boxe quand j'avais encore un crâne valide. L'ascension ne se serait jamais faite sans les coachs Musashi et Gandalf.

Je dois énormément à la famille, aux petits et grands lutins, à ceux plus ou moins loin : aux Pauvert (Sud et Nord), aux Serveaux, aux Bascou, aux Gonthier, aux Potiron et Moreau pour m'avoir fait voir au-delà des collines, toujours. Merci à mes parents, soutien sans faille et continu. Merci au reste de la tribu des p'tits indiens Sistreush et Brotheursh. Promis je mange des castors et des courges.

Aline, tu as été mon ravitaillement salutaire entres les pics de bonheur et les vallées difficiles, celles que j'ai parcouru à pied et celles à dos de cerveau, tu m'as soutenu et tu me soutiens encore dans les deux en ouistiti invétéré. À nous les prochains chemins tortueux avec vues imprenables (parole d'optimiste) entres torrents et forêts (parole de lutin).

To "Jack of all trades, master of none, but better than one"

Liste des productions et communications scientifiques

Publications acceptées

1. Derocles SAP, Bohan DA, Dumbrell AJ, Kitson JJN, Massol F, **Pauvert C**, Plantegenest M, Vacher C, and Evans DM (2018). "Chapter One - Biomonitoring for the 21st Century : Integrating Next-Generation Sequencing Into Ecological Network Analysis". In : *Advances in Ecological Research*. Ed. by DA Bohan, AJ Dumbrell, G Woodward, and M Jackson. Vol. 58. Next Generation Biomonitoring : Part 1. Academic Press, pp. 1–62. doi: 10.1016/bs.aecr.2017.12.001.

2. **Pauvert C**, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I, Vallance J, and Vacher C (2019). "Bioinformatics Matters : The Accuracy of Plant and Soil Fungal Community Data Is Highly Dependent on the Metabarcoding Pipeline". In : *Fungal Ecology* 41, pp. 23–33. doi:10.1016/j.funeco.2019.03.005

Publications soumises ou en préparation

1. **Pauvert C**, Vallance J, Deliere L, Buee M, and Vacher C (2019). "Microbial Networks Inferred from Metabarcoding Data Lack Replicability : Consequences for next-Generation Biomonitoring". In : *bioRxiv*, p. 642199. doi:10.1101/642199. Soumise à *Molecular Ecology Resources* en Août 2019.

2. Fort T, **Pauvert C**, Zanne AE, Ovaskainen O, Caignard T, Barret M, Compant S, Hampe A, Delzon S, and Vacher C (2019). "Maternal Effects and Environmental Filtering Shape Seed Fungal Communities in Oak Trees". In : *bioRxiv*, p. 691121. doi:10.1101/691121. Soumise au *New Phytologist* en Juillet 2019 et rejeté.

3. **Pauvert C**, Vallance J, Fort T, Calonnec A, Faivre d'Arcier J, Chancerel E, Massot M, Bohan DA, Chiquet J, Robin S, and Vacher C (2019). "Microbial association networks give relevant insights into plant pathobiomes". En préparation.

4. Fort T, Louisanna E, Chancerel E, Linares-Maurizi A, Balias P, Burlett R, **Pauvert C**, Gibbon Y, Zanne AE, Hampe A, Ogée J, Schimann H, Wingate L, and Vacher C (2019). "Within-canopy variations in the structure and function of phyllosphere microbial communities and cascading effects on leaf decomposition". En préparation.

Communications

Communications sans actes de congrès

1. **Pauvert C**, Vallance J, Delière L, Buée M and Vacher C, (2018). "Cropping systems shape foliar fungal networks of grapevine". Pathobiome 2018, Ajaccio, Corse, France, 18-20 Mars 2018 (Communication orale). Premier prix de la présentation des doctorants.

2. **Pauvert C**, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I, Vallance J, and Vacher C (2018). "ITS all fun guys : A comparison of bioinformatic pipelines for analyzing plant and soil fungal communities". International Phytobiomes Conference 2018, Montpellier, France, 4-6 Décembre 2018 (Poster). Prix jeunes chercheurs.

Autres communications

1. “Learning microbial networks from meta-barcoding data : Potential application to biocontrol and biomonitoring“. Communication orale courte pour la Journée des doctorants de l’unité BIOGECO en Avril 2017.

2. “Apprendre les réseaux microbiens grâce aux données de séquençage : une application potentielle au biocontrôle“. Communication orale de vulgarisation scientifique dans le cadre des PubHD de Bordeaux en Septembre 2017.

3. “A comparison of bioinformatic pipelines for analyzing plant and soil fungal communities“. Communication orale pour le séminaire de l’unité BIOGECO en Février 2018.

4. “Bioinformatics Matters : The Accuracy of Plant and Soil Fungal Community Data Is Highly Dependent on the Metabarcoding Pipeline“. Communication orale pour le séminaire de l’unité BIOGECO en Janvier 2019.

5. “Reconstructing microbial networks for uncovering biocontrol agents ? A test on grapevine powdery mildew“. Communication orale pour le séminaire de l’unité SAVE en Juin 2019.

Formations suivies

- École d’été en Métagénomique à l’Institut Pasteur en Septembre 2016.
- “Rédiger clairement : l’écriture scientifique“ par l’URFIST Bordeaux en octobre 2016.
- Formation avec Céline Lalanne et Patrick Léger (techniciens à l’UMR Biogeco) à la Biologie moléculaire (broyage et extraction d’ADN) en Octobre et Novembre 2016.
- “Intégrité scientifique dans les secteurs de la recherche“ en ligne par l’Université de Bordeaux en 2017
- École d’été du Labex TULIP “Biological interactions from genes to ecosystems“ du 16 au 21 juillet 2017.
- École d’été à Leuven (Belgique) “Ecological Network Inference & Analysis“ du 11 au 13 septembre 2017.
- Formation à la digital droplet PCR avec Marie Massot (assistante ingénieure à l’UMR BIOGECO) en Octobre 2017.
- Séjour à l’AgroParisTech pour collaboration sur les modèles statistiques de graphes avec Stéphane Robin et Julien Chiquet en février 2018 et février 2019.
- “Construire et développer son projet professionnel“ en février 2019 par l’Université de Bordeaux.
- Formation à l’utilisation des modèles multivariés pour l’analyse de données de comptage (PLN) en écologie et l’analyse des réseaux à l’aide des modèles à blocs stochastiques dans le cadre de l’ANR NGB en Février 2019.

- “Meet up ’To cheat or not to cheat : The dark side of data science’“ à l’URFIST de Bordeaux en mars 2019.
- “Un hackathon dataviz pour aider une ONG. De la théorie aux (bonnes) pratiques de la data visualisation“ à l’URFIST de Bordeaux en mars 2019.
- Formation à la microbiologie avec Jessica Vallance dans l’UMR SAVE en Avril 2019.

Participation à des colloques

- Réunion de lancement du réseau INRA “Phytobiomes“ à Paris en Septembre 2016.
- Colloque CARTABLE “Colloque Apprentissage de Réseaux : de la Théorie aux Applications en Biologie et Ecologie“ à Toulouse en Octobre 2016.
- Demi journée d’animation sur les approches réseaux (communautés microbiennes, invasions biologiques et émergence) au CBGP de Montpellier en Mai 2017.
- Journée thématique“Méthodes de construction et d’analyse de réseaux à partir de données de métagénomique ou métabarcoding“ du GDR Génomique Environnementale à Nantes en Juin 2017.

Autres activités

- Représentant élu des doctorants au conseil de l’Ecole doctorale en 2017.
- Coorganisateur des Journées des Doctorants de l’Ecole Doctorale 2018.
- Participation à l’encadrement du stage de Master 2 de Célia Duluc sur les communautés microbiennes du colza avec Corinne Vacher et Christophe Mougel en 2018.

Sommaire **7**

I Introduction **2**

1 Les micro-organismes associés aux plantes forment des réseaux écologiques 3
1.1 Quels micro-organismes colonisent les plantes ? 3
1.2 Comment interagissent-ils avec la plante hôte et entre eux ? 6
2 Comment et pourquoi reconstruire les réseaux microbiens ? 9
2.1 Comment identifier les micro-organismes ? 9
2.2 Comment identifier les interactions ? 20
2.3 Applications potentielles 27
3 Objectifs de la thèse 29
4 Modèle d'étude : la vigne 30

II L'exactitude des données sur les communautés fongiques des plantes et du sol dépend fortement du pipeline de métabarcoding **34**

1 Résumé 35
2 Introduction 36
3 Matériels et méthodes 39
3.1 Une communauté fongique artificielle. 39
3.2 Pipelines bioinformatiques. 40
4 Résultats 44
4.1 Évaluation de la sensibilité maximale 44
4.2 Influence des pipelines sur les séquences 44
4.3 Comparaison des pipelines. 46
5 Discussion 50
6 Informations supplémentaires 54

III Les réseaux microbiens inférés manquent de réplicabilité : implications pour la biosurveillance de nouvelle génération **59**

1 Résumé 60
2 Introduction 61
3 Matériels et méthodes 64
3.1 Site de l'étude et échantillonnage 64
3.2 Analyses bioinformatiques. 65
3.3 Analyses statistiques 66
4 Résultats 70
4.1 Effet de la pratique culturale sur la diversité α et β de la communauté 70
4.2 Effet de la pratique culturale sur la propriétés α et β des réseaux 70

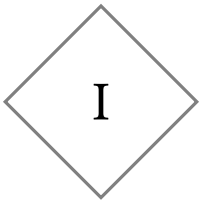
5 Discussion 74
6 Informations supplémentaires 79

IV Les associations des réseaux microbiens offrent des indications pertinentes sur les pathobiomes des plantes 84
1 Résumé 85
2 Introduction 86
3 Matériels et méthodes 88
3.1 Site de l'étude et échantillonnage 88
3.2 Extraction de l'ADN et séquençage Illumina MiSeq 90
3.3 Quantification des ADN fongiques avec la ddPCR^{TM} 91
3.4 Analyses bioinformatiques. 91
3.5 Tests des hypothèses 92
4 Résultats 97
5 Discussion 102
6 Informations supplémentaires 106
6.1 Sélection des espèces dans les réseaux 114
6.2 Code pour la fouille automatique de la littérature 114

V Conclusion et perspectives 118
1 Points forts et limites des approches utilisées dans la thèse 119
2 Pistes de recherche futures 124
2.1 Quel échantillonnage pour inférer les réseaux microbiens ?. 124
2.2 Comment réduire les biais moléculaires ? 125
2.3 Comment réduire les biais statistiques ? 126
2.4 Comment mieux valider les interactions prédites ?. 127

VI Bibliographie 130

VII Annexes 166
1 La biosurveillance pour le 21ème siècle : Intégrer le séquençage nouvelle-génération dans l'analyse des réseaux écologiques 167
2 Les effets maternels et l'environnement façonnent les communautés fongiques des graines chez les chênes 168



Introduction

1 Les micro-organismes associés aux plantes forment des réseaux écologiques

1.1 Quels micro-organismes colonisent les plantes ?

Les micro-organismes sont définis comme les êtres vivants invisibles à l'œil nu (Maczulak, 2011, p. 514) Comme les macro-organismes, les micro-organismes peuvent être classés selon une phylogénie, un arbre du vivant, qui indique les relations ancestrales entre eux (e.g., Hug et al., 2016). On distingue ainsi trois grands domaines au sein du vivant : les bactéries, les archées et les eucaryotes (e.g., Hug et al., 2016). Ces domaines sont les racines d'une classification hiérarchique utilisée pour classer les organismes vivants selon les rangs taxonomiques suivants : Domaine > Règne > Embranchement (ou Phylum) > Classe > Ordre > Famille > Genre > Espèce. Les micro-organismes forment dans la nature des *communautés*. Selon la définition de Whittaker (1975) une communauté est « . . . un ensemble de populations de plantes, d'animaux, de bactéries et de champignons qui vivent dans un environnement et interagissent les uns avec les autres, formant ensemble un système vivant distinct avec sa propre composition, structure, relations environnementales, développement et fonction. » (voir Morin, 2011, pour une discussion des définitions alternatives). Les micro-organismes qui colonisent les plantes sont présentés dans les paragraphes suivants.

Les bactéries sont des micro-organismes unicellulaires procaryotes – sans noyau et organites au sein de leur cellule – dont la taille est de l'ordre du micromètre, soit un millième de millimètre (Maczulak, 2011, p. 76). Tous les compartiments des plantes – racines, feuilles, tiges, fleurs et fruits – sont peuplés de bactéries et nombreuses sont celles qui sont retrouvées dans plusieurs compartiments (Amend et al., 2019 ; Hamonts et al., 2018 ; Lundberg et al., 2012 ; Ottesen et al., 2013 ; Zarraonaindia et al., 2015). La densité bactérienne est estimée à près d'un million d'individus par centimètre carré sur les feuilles (Figure 1) (Vorholt, 2012) et jusqu'à cent fois plus au niveau des racines (Bulgarelli et al., 2013). Ces communautés sont dynamiques et leur composition change tout au long du développement de la plante et de ses organes (Copeland et al., 2015 ; Maignien et al., 2014 ; Shade et al., 2013). Parmi cette diversité microbienne associée à la plante, les bactéries détectées fréquemment à la surface et au sein des tissus végétaux appartiennent aux phylums des *Actinobacteria*, $\alpha-$ et $\gamma-$*Proteobacteria* (Hardoim et al., 2015). Des communautés d'archées sont également associées aux plantes (Buée et al., 2009a ; Hassani et al., 2018 ; Knief et al., 2012 ; Moissl-Eichinger et al., 2017). L'absence de détection d'agents pathogènes chez les archées pourrait expliquer pourquoi elles ont moins été étudiées en association avec un hôte, mais elles sont désormais de plus en plus incluses (Moissl-Eichinger et al., 2017).

Les champignons sont des micro-organismes eucaryotes pluricellulaires ou unicellulaires (levures) qui tirent leur énergie de nutriments absorbés (hétérotrophie) puisqu'ils n'ont pas de chlorophylle et ne réalisent pas la photosynthèse (Maczulak, 2011, p. 326). Les champignons sont plus grands que les bactéries, avec des tailles comprises entre la dizaine et la centaine de micromètres (Figure 2). Dans le langage commun, ils font référence aux organes de fructification

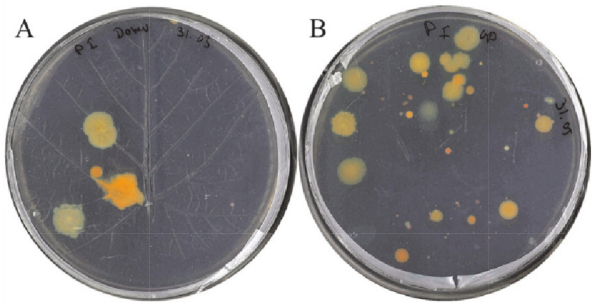


Figure 1 – Bactéries en culture issues d'empreintes de feuilles de vigne (*Vitis vinifera* L. cv. Pinot noir). Source : Vionnet et al. (2018).

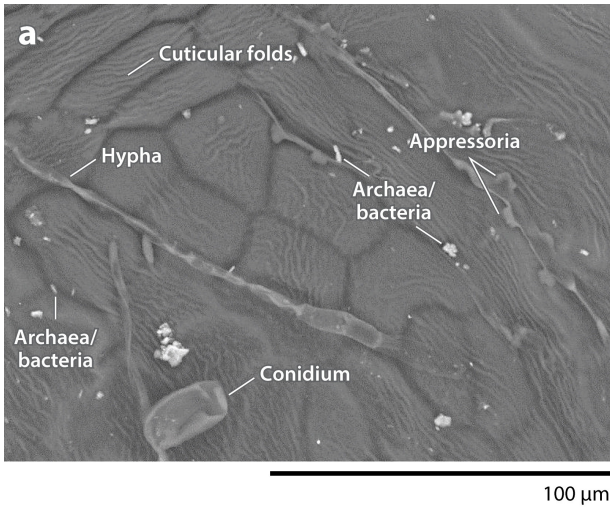


Figure 2 – Bactéries et champignons (notamment l'agent de l'oïdium, *Erysiphe necator*) colonisant la surface d'une feuille de vigne (*Vitis vinifera* L. cv. Neuburger) en microscopie électronique. Source : Vacher et al. (2016a).

visibles et comestibles dans les sous-bois (Dupont et al., 2017), Le terme “levures“ désignent des champignons unicellulaires qui se reproduisent par bourgeonnement ou fission tels que la levure de boulanger *Saccharomyces cerevisiae* (Kurtzman et al., 2011). Les champignons sont détectés dans les différents compartiments des plantes, notamment les racines où les mycorhizes forment des symbioses avec les racines de la majorité des espèces de plantes, et échangent nutriments du sol contre des sucres dérivés de la photosynthèse (Bonfante et Genre, 2010 ; Peay et al., 2016). Ils sont détectés jusque dans le feuillage de la cime des arbres les plus hauts du monde (Harrison et al., 2016), dans les habitats variés des forêts (Baldrian, 2017) et au sein de nombreuses plantes cultivées (Chen et al., 2016 ; Fisher et Petrini, 1992 ; Hertz et al., 2016 ; Pancher et al., 2012 ; Sapkota et al., 2015). Contrairement aux bactéries, il n’existe pas d’estimation de la densité des champignons dans les différents compartiments des plantes, leur biomasse est néanmoins estimée entre 0.05 et 0.2% du poids sec des tissus photosynthétiques (Davey et al., 2009). Les champignons associés aux parties aériennes appartiennent principalement aux phylums *Ascomycota* et *Basidiomycota* tandis que ceux associés aux racines sont majoritairement des *Glomeromycota* (Hardoim et al., 2015 ; Peay et al., 2016).

Les oomycètes ont longtemps été classés parmi les champignons mais appartiennent désormais aux Straménopiles (dont font partie les algues brunes) (Ruggiero et al., 2015). Les oomycètes comprennent entre 800 et 1000 espèces dont de nombreux agents pathogènes des plantes cultivées, comme par exemple *Phytophthora infestans* (mildiou de la pomme de terre), *Plasmopara viticola* (mildiou de la vigne) ou *Bremia lactucae* (mildiou de la salade) (Kamoun et al., 2015). Des oomycètes non-pathogènes sont également détectés dans les échantillons végétaux (Agler et al., 2016 ; Hassani et al., 2018 ; Schlatter et al., 2017).

Les virus sont des particules inframicroscopiques constituées d’acide nucléique entouré d’une couche de protéines et qui doivent infecter des cellules vivantes pour se multiplier (Maczulak, 2011, p. 779). Leur appartenance aux micro-organismes ainsi qu’aux êtres vivants divise la communauté scientifique (Pradeu et al., 2016). Les virus associés aux plantes infectent les cellules végétales ou les micro-organismes associés aux plantes (Roossinck, 2019). La classe des mycovirus englobe des virus capables d’infecter des cellules de champignons (Pandey et al., 2018), ce qui peut conférer à la plante une résistance accrue aux stress biotiques et abiotiques (Márquez et al., 2007 ; Nuss, 2005).

Des cortèges complexes de micro-organismes sont associés aux plantes, et ces micro-organismes sont désignés comme *épiphytes* ou *endophytes* selon qu’ils colonisent l’extérieur ou l’intérieur des tissus végétaux. Certains micro-organismes sont néanmoins capables de changer de compartiment (Rodriguez et al., 2009).

Les micro-organismes sont si intimement et fréquemment associés aux plantes que la notion qu’ils devaient être étudiés ensemble s’est formée : l’holobionte (Margulis, 1991 ; Zilber-Rosenberg et Rosenberg, 2008). L’holobionte constitue l’ensemble d’un hôte (plante ou animal) et les micro-organismes qui le colonise (Margulis, 1991 ; Zilber-Rosenberg et Rosenberg, 2008). L’ensemble des génomes de l’holobionte, l’hologénome, est parfois considéré comme une unité de sélection en évolution (Zilber-Rosenberg et Rosenberg, 2008). Cette théorie de l’hologénome suppose que les adaptations de l’holobionte aux changements de l’environnement peuvent se faire grâce aux organismes ayant les temps de générations les plus courts et donc par les micro-organismes (Bordenstein et Theis, 2015). Certains aspects de cette théorie sont

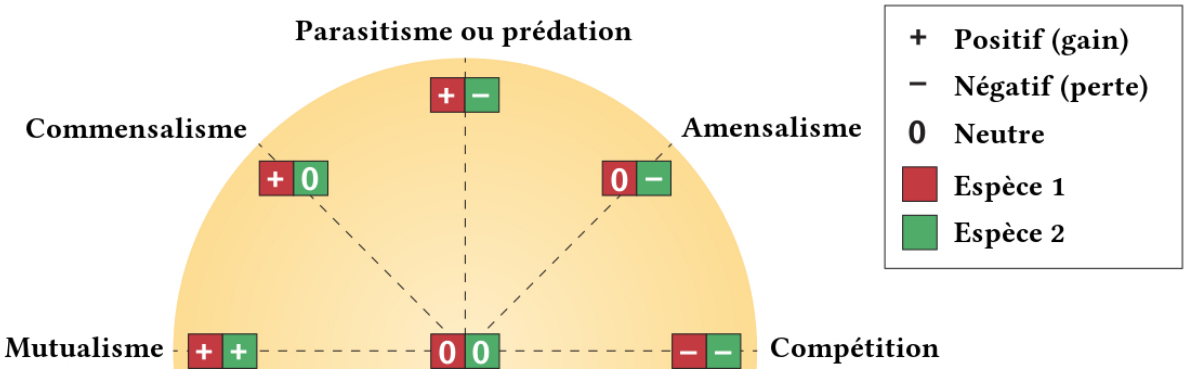


FIGURE 3 – Représentation des interactions écologiques entre deux espèces. Pour chaque partenaire l'issue de l'interaction peut être un gain (+) ou une perte (-) de fitness, ou bien être neutre (0). Adaptée de FAUST et RAES (2012) et LIDICKER (1979).

critiqués et demandent à être démontrés comme par exemple l'hérédité des micro-organismes associés à un hôte ou encore les effets positifs des micro-organismes sur la valeur adaptative de l'hôte – la *fitness* – (ROSENBERG et ZILBER-ROSENBERG, 2018). Cette théorie permet toutefois de guider les études sur les plantes hôtes et de ne pas omettre les micro-organismes qui en sont une partie intégrante (VANDENKOORNHUYSE et al., 2015).

1.2 Comment interagissent-ils avec la plante hôte et entre eux ?

Les micro-organismes ont tout d'abord interagit entre eux avant d'interagir avec les plantes (HASSANI et al., 2018). Ces interactions peuvent être représentées par un réseau dans lequel les nœuds représentent les micro-organismes et les liens représentent leurs interactions biotiques.

Les différents types d'interactions peuvent être classés selon l'impact sur la fitness de chacun des partenaires : positif (+), négatif (-) ou neutre (0) (Figure 3 FAUST et RAES, 2012 ; LIDICKER, 1979). Dans une interaction entre deux espèces, trois types d'interactions augmentent la fitness d'une des deux espèces : la coopération (++), le commensalisme (+0) et la prédation ou le parasitisme (+-). À l'inverse, trois types d'interactions nuisent à l'une des deux espèces : l'amensalisme (-0), la compétition (–), et la prédation ou le parasitisme (-+).

Les modèles d'écologie théorique montrent que la compétition stabilise les communautés microbiennes, c'est-à-dire que les abondances retournent à leur équilibre après une perturbation (COYTE et al., 2015). En effet, l'introduction croissante d'interactions coopératives – comme le mutualisme (Figure 3) – rend ces réseaux interdépendants et donc instables dès que l'abondance d'une espèce s'effondre (COYTE et al., 2015). Ces prédictions issues de simulations sont corroborées par des études expérimentales où la compétition entre espèces bactériennes domine (FOSTER et BELL, 2012). Les micro-organismes qui nécessitent les mêmes ressources – sucres, acides aminés – entrent ainsi en compétition par exploitation lorsque ces ressources sont limitées (GHOUL et MITRI, 2016). La nature d'une interaction entre deux partenaires peut varier selon les conditions environnementales (HOEK et al., 2016). Dans une relation mutualiste dans laquelle les acides aminés synthétisés par une levure bénéficient à une autre, le changement

de concentration de ces acides aminés dans le milieu modifie leur interaction (Hoek et al., 2016). Dans une relation mutualiste dans laquelle les acides aminés synthétisés par une levure bénéficient à une autre, si ces acides aminées sont introduit dans le milieu, l'interaction est modifiée (Hoek et al., 2016). Plus cette ressource augmente, plus il est coûteux de synthétiser pour autrui, le coût du mutualisme augmente, et l'interaction passe du mutualisme à une compétition menant à l'extinction d'une des deux levures (Hoek et al., 2016). Les micro-organismes peuvent également déployer un arsenal chimique comme la synthèse de composés anti-microbiens chez les champignons pour rivaliser avec les bactéries. Cet arsenal induit de la compétition par interférence ou à de l'amensalisme selon que les bactéries ripostent ou non (Ghoul et Mitri, 2016 ; McCormack et al., 1994 ; Soliman et al., 2013). Une compétition apparente peut survenir entre des espèces n'interagissant pas si ces espèces partagent un prédateur ou si elles sont sensibles aux mêmes composés anti-microbiens (Holt et Bonsall, 2017). Des relations de parasitisme au sein du même domaine microbien se produisent aussi, par exemple lorsque des pathogènes de plantes sont eux-mêmes parasités par des champignons (mycoparasites) (Kiss, 1998 ; Parratt et Laine, 2016). Toutefois, des coopérations existent entre micro-organismes que ce soit pour partager des nutriments (Zengler et Zaramela, 2018) ou se disperser. Par exemple, les hyphes fongiques favorisent la dispersion et les échanges de matériel génétique des bactéries dans le sol (Berthold et al., 2016) et même dans le fromage (Zhang et al., 2018). Il reste à démontrer dans ce cas précis s'il s'agit de mutualisme ou de commensalisme en évaluant si les champignons bénéficient ou non de l'interaction. De telles relations pourraient exister à la surface de la feuille également puisque le marquage par fluorescence des micro-organismes en microscopie montre la présence de bactéries vivantes au sein des hyphes de champignons isolés de plantes (Hoffman et Arnold, 2010). Les interactions bactéries-champignons ne se limitent pas aux cas présentés et relèvent de toute la gamme d'interactions (Deveau et al., 2018 ; Kobayashi et Crouch, 2009, Figure 3)

La présence de bactéries ou champignons spécifiques peut bénéficier à la plante en la protégeant, en augmenteant sa croissance (Chen et al., 2017 ; Fávaro et al., 2012), ou en induisant une résistance plus importante à des facteurs abiotiques tels que la sécheresse (Naveed et al., 2014). Par exemple, l'inoculation de certaines bactéries sur les fleurs du colza influence la visite des pollinisateurs et donc la fitness de ces plantes (Farré-Armengol et Junker, 2019). Ces micro-organismes sont retrouvés sans inoculation sur les pollinisateurs qui servent leur dispersion (Russell et al., 2019). Les mycorhizes établissent une relation de mutualisme avec la plante en lui fournissant des éléments minéraux (phosphate, azote) contre des composés carbonés issus de la photosynthèse (Bonfante et Genre, 2010). Une étude récente a démontré que la présence de bactéries commensales dans les racines est nécessaire à la survie de *Arabidopsis thaliana* et que cette survie dépend probablement de la compétition entre les bactéries commensales d'une part et des champignons et oomycètes d'autre part (Durán et al., 2018). Mais les micro-organismes sont aussi les agents responsables de nombreuses maladies des plantes (Buée et al., 2009a), ce qui peut mettre en péril les plantes cultivées et les personnes qui en dépendent (Fisher et al., 2012).

Le lien pathogène-hôte a été étendu par le concept de *pathobiome* défini comme « un agent pathogène intégré dans son environnement biotique » (Vayssier-Taussat et al., 2014). Brader et al. (2017) ont reformulé ce concept sous la forme « d'un pathogène, son environnement microbien et leurs interactions qui mènent à l'état de maladie ». Bass et al. (2019) le définissent comme « l'ensemble des organismes associés à l'hôte – comprenant les procaryotes, les eucaryotes et les virus – associés à un état de santé réduit (ou potentiellement réduit), en raison

des interactions entre les membres de cet ensemble et l'hôte ». Ces deux dernières définitions insistent sur le fait que l'état de santé de l'hôte peut être modulé par les membres du pathobiome (Kemen, 2014). Ainsi ce concept a initié l'émergence d'études plus holistiques prenant en compte cet environnement microbien (Carmichael et al., 2019 ; Kerdraon et al., 2019 ; Koskella et al., 2017 ; Lebreton et al., 2019 ; Suda et al., 2009). Dans la suite du manuscrit, le terme de pathobiome fait référence au sous-ensemble des micro-organismes associés à une plante hôte en interaction avec un agent pathogène (Brader et al., 2017 ; Vayssier-Taussat et al., 2014).

Ainsi, dès qu'un agent pathogène envahit un hôte et interagit avec sa communauté microbienne, ils constituent un pathobiome. En cas d'invasion fructueuse, le pathogène peut perturber la communauté microbienne, et certains auteurs postulent que cela pourrait induire alors une plus grande hétérogénéité dans la communauté (Zaneveld et al., 2017). Ce processus, dénommé le *principe Anna Karénine*, provient de la lecture de l'incipit [1] du roman de Tolstoï par le prisme de l'écologie microbienne : « Tous les *microbiomes sains* se ressemblent, mais chaque *microbiome altéré* l'est à sa façon » (Zaneveld et al., 2017). Bien que la composition du microbiote des plantes change entre les tissus sains et les tissus infectés, les compositions bactériennes et fongiques semblent plus homogènes dans les tissus infectés d'après des études effectuées sur différentes plantes, *i.e.* concombre, arabette, marronnier, fusain, radis (Agler et al., 2016 ; Koskella et al., 2017 ; Rezki et al., 2016 ; Suda et al., 2009 ; Zhang et al., 2019). Ces résultats remettent donc en question l'existence de le principe Anna Karénine chez les plantes par rapport aux études reportées chez les animaux (Zaneveld et al., 2017).

Le succès de l'établissement d'un pathogène sur une plante dépend de la plante et des conditions abiotiques, notamment les niveaux d'humidité et de température, qui sont susceptibles de varier fortement dans le cadre des changements globaux (Anderson et al., 2004 ; Cavicchioli et al., 2019). Ces conditions abiotiques seront favorables ou défavorables aux pathogènes selon leur physiologie (Cavicchioli et al., 2019). Au-delà de conditions abiotiques défavorables, l'invasion infructueuse d'un agent pathogène peut être liée à un *effet barrière* de la communauté microbienne qui confère des fonctions immunitaires étendues à l'hôte (Vannier et al., 2019). Les micro-organismes réalisent un effet barrière direct ou indirect via l'activation des mécanismes de défenses des plantes (Hacquard et al., 2017 ; Lee et al., 2017 ; Vogel et al., 2016). L'effet barrière direct peut être lié à différents types d'interactions entre micro-organismes : la compétition par interférence, la compétition indirecte pour la ressource (espace ou nutriments), de la compétition apparente ou du parasitisme (Arnold et al., 2003 ; Durán et al., 2018 ; Kemen, 2014 ; Li et al., 2019).

Une meilleure compréhension des mécanismes en jeu au sein des pathobiomes paraît importante pour limiter les dégâts causés par les pathogènes voire pour maintenir la survie des hôtes considérés notamment les plantes cultivées. Les hôtes et leurs micro-organismes sont en effet affectés par les changements globaux (Cavicchioli et al., 2019). Le maintien des interactions hôte–micro-organismes est donc crucial puisque les micro-organismes sont responsables de la vie telle que nous la connaissons (Falkowski et al., 2008 ; Gilbert et Neufeld, 2014 ; Selosse, 2017).

[1] « Toutes les familles heureuses se ressemblent, mais chaque famille malheureuse l'est à sa façon. » Tolstoï. *Anna Karénine*, 1885.

2 Comment et pourquoi reconstruire les réseaux microbiens ?

2.1 Comment identifier les micro-organismes ?

L'identification taxonomique des organismes basée sur la description de leurs morphologies et habitats est plus difficile pour les micro-organismes que pour les macro-organismes. La microbiologie a pour objectif d'identifier les micro-organismes via leur isolement depuis l'écosystème étudié puis leur culture sur des milieux artificiels (Crowther et al., 2018). Mais ces méthodes sont longues et laborieuses et nécessitent une forte expertise (Hibbett et al., 2016). De plus, les micro-organismes biotrophes obligatoires se nourrissent via les cellules vivantes de la plante hôte donc les milieux artificiels ne permettent pas leur croissance (Kemen et Jones, 2012). Seule une fraction, entre 0.1% et 50% maximum, des micro-organismes associés aux plantes est ainsi considérée cultivable (Müller et Ruppel, 2014). Mais les innovations récentes en microbiologie visent à augmenter cette fraction via des approches haut-débit (Kehe et al., 2019 ; Lagier et al., 2018). Ces méthodes de culturomique testent automatiquement plusieurs milieux artificiels ainsi que des combinaisons de nutriments (Lagier et al., 2018) ou d'espèces (Kehe et al., 2019).

Tout en reconnaissant ces progrès, nous nous sommes basés pour cette thèse sur une approche indépendante de la culture microbienne : le *métabarcoding*. Cette approche se base sur l'identification d'un organisme par une séquence d'ADN représentative, un *barcode* ou marqueur moléculaire (Hebert et al., 2003). Elle est nommée métabarcoding lorsqu'elle est appliquée à la communauté pour identifier plusieurs organismes simultanément(Taberlet et al., 2018). La facilité d'accès croissante aux nouvelles technologies de séquençage haut-débit a permis l'utilisation massive du métabarcoding pour étudier les communautés associées aux plantes (Di Bella et al., 2013) (Figure 4). L'assignation d'une séquence à un rang taxonomique se base sur la comparaison avec des banques de référence (Richter et Rosselló-Móra, 2009 ; Varghese et al., 2015).

2.1.1 *Des micro-organismes aux séquences d'ADN*

La première étape de l'approche de métabarcoding est la sélection du marqueur moléculaire. Ce choix va contraindre la résolution taxonomique, à savoir quel groupe microbien sera identifié et jusqu'à quel rang taxonomique. Un marqueur qui identifie correctement une espèce microbienne doit posséder une variabilité génétique au sein de l'espèce inférieure à celle entre espèces (Schoch et al., 2012).

L'identification moléculaire des micro-organismes au sein du règne des bactéries repose sur la petite sous-unité de l'ARN ribosomique dénommée 16S dont la longueur totale avoisine 1 500 nucléotides (nt) (Karst et al., 2018). La séquence de cette petite sous-unité est composée d'une alternance de régions conservées et de courtes régions variables (Yarza et al., 2014). Ces régions variables sont considérées comme des substituts moléculaires à l'espèce bactérienne (Edgar, 2018 ; Stackebrandt et Goebel, 1994). Mais elles sont présentes aussi dans les chloroplastes,

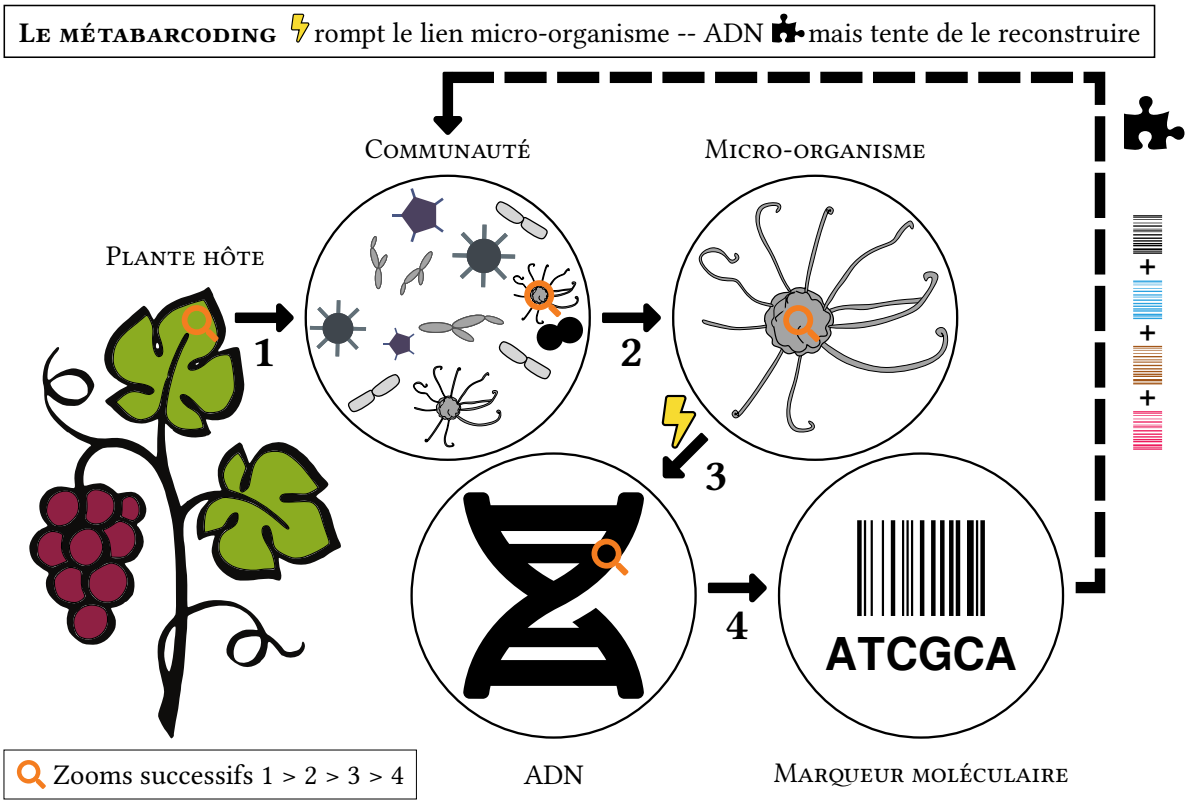


Figure 4 – Principe du métabarcoding appliqué à l'étude des communautés de micro-organismes associées aux plantes. La plante hôte est échantillonnée (1) afin de caractériser ses communautés composées de différents micro-organismes (2). L'approche de métabarcoding vise à recenser les marqueurs moléculaires au sein de l'ADN des micro-organismes (3,4) comme un substitut à leur observation. Mais cette approche se heurte à la rupture du lien micro-organismes et ADN (symbolisée par un éclair) qui va nécessiter des traitements supplémentaires (en ligne pointillée) pour retrouver la composition de la communauté.

ces organites issus d'endosymbiose responsables de la photosynthèse au sein des cellules végétales (Rastogi et al., 2010). Leurs séquences peuvent ainsi saturer les échantillons prélevés sur des plantes et limiter l'identification des espèces microbiennes (Rastogi et al., 2010). Pour amplifier les communautés microbiennes des tissus photosynthétiques, une amplification de l'ADN excluant les chloroplastes est nécessaire ou bien un blocage de leur amplification (Fitzpatrick et al., 2018). La caractérisation des ces communautés peut s'appuyer également sur des marqueurs bactériens alternatifs comme les gènes codant la sous-unité B de l'ADN gyrase (gyrB) (Barret et al., 2015) ou la sous-unité β de l'ARN polymérase (rpoB) (Di Bella et al., 2013).

A l'instar des bactéries, une portion de l'ADN ribosomique est également utilisée pour l'identification des champignons. La région *Internal Transcribed Spacer* (ITS) est située entre les gènes qui codent pour les deux sous-unités du ribosome : la petite sous-unité (voir SSU Figure 5) en amont contenant le gène 18S et la grande sous-unité (voir LSU Figure 5) en aval contenant les gènes 5.8S et 28S (Halwachs et al., 2017). L'ITS est considéré comme le marqueur moléculaire fongique de référence (Schoch et al., 2012) et est également utilisé pour la détection des oomycètes (Legeay et al., 2019). La région ITS complète a une longueur variable entre 150 et 800 nt environ (Palmer et al., 2018). Bien que ce marqueur moléculaire soit considéré comme suffisamment résolutif pour identifier les champignons au niveau de l'espèce (Kõljalg et al., 2013 ; Vu et al., 2019), certaines de ses spécificités peuvent compliquer l'identification correcte des champignons (Palmer et al., 2018). La variabilité importante de la longueur de la région ITS complète est due aux longueurs variables des deux régions la constituant : l'ITS1 (entre le 18S et le 5.8S) et l'ITS2 (entre le 5.8S et le 28S) (Palmer et al., 2018 ; Tedersoo et al., 2015). La possibilité d'insertion d'un intron de grande taille au sein de la région ITS1 augmente la variabilité de la longeur (Taylor et al., 2016). Ces variations de longueur associées à un taux important de répétitions de nucléotides, les homopolymères, limitent également la possibilité de construire un arbre phylogénétique (Nilsson et al., 2019 ; Palmer et al., 2018) et donc d'utiliser des outils basés sur les distances phylogénétiques comme UniFrac (Lozupone et Knight, 2005). La région 18S a été proposée comme marqueur fongique alternatif pour réaliser ce type d'analyses phylogénétiques mais ce marqueur est faiblement résolutif (Lindahl et al., 2013). Le gène TEF1α a également été récemment suggéré pour combiner les avantages de la phylogénie et de la résolution taxonomique à l'espèce (Stielow et al., 2015). L'existence de banques dédiées aux séquences ITS expertisées comme UNITE (Abarenkov et al., 2010) et son usage répandu favorisant les comparaisons, font que la région ITS reste un marqueur moléculaire de choix pour la majorité des études sur la diversité et la structure des communautés de champignons (Lindahl et al., 2013 ; Yahr et al., 2016).

Une fois les échantillons récoltés, l'extraction de l'ADN microbien est réalisée conjointement grâce à des actions physiques (température, broyage etc...) et à une lyse chimique des cellules (Lindahl et al., 2013). Le protocole d'extraction va ainsi influencer la quantité et la qualité de l'ADN extrait, et la détection de certains micro-organismes donc la composition de la communauté étudiée (Dopheide et al., 2019 ; Maropola et al., 2015 ; Pollock et al., 2018 ; Rittenour et al., 2012).

Suivant l'extraction, une étape d'amplification, la *polymerase chain reaction* ou PCR, augmente la quantité d'ADN microbien grâce à une réaction en chaîne cyclique qui sépare les brins d'ADN, hybride les amorces et synthétise le marqueur ciblé (Saiki et al., 1988). Les amorces sont des courts fragments d'ADN construits pour s'hybrider aux régions adjacentes conservées du

marqueur ciblé comme le 18S et le 5.8S pour l'ITS1 (Figure 5) et limiter l'amplification de l'ADN végétal (Bellemain et al., 2010). Le choix des amorces est crucial pour identifier correctement les micro-organismes et plusieurs couples d'amorces ont été proposés pour répondre au compromis entre une détection large – avec des faux positifs – et très ciblée – avec des faux négatifs – des espèces étudiées (par exemple Gardes et Bruns, 1993 ; Ihrmark et al., 2012 ; Taylor et al., 2016 ; Tedersoo et al., 2015 ; White et al., 1990). Lorsque suffisamment d'ADN microbien de qualité est obtenu, le séquençage de l'ADN permet de révéler la composition en nucléotides des marqueurs moléculaires (voir Figure 4) sous forme de séquences appelées *reads* ou lectures (Bálint et al., 2016 ; Lindahl et al., 2013). La concentration d'ADN est préalablement ajustée avant le séquençage via des dilutions pour être similaire dans les différents échantillons. De nombreuses technologies de séquençage existent et elles ont évolué rapidement au cours des deux dernières décennies (Goodwin et al., 2016). Actuellement, les technologies disponibles sont celles de deuxième et troisième génération, qui génèrent respectivement des millions de courtes lectures (150-300 nt, Illumina) ou des milliers de lectures plus longues (Oxford Nanopore et Pacific Bioscience) (Figure 5 ; Goodwin et al., 2016 ; Nilsson et al., 2019). Dans cette thèse, la technologie Illumina MiSeq a été utilisée. Il s'agit en effet de l'approche la plus commune dans les études de métabarcoding (Nilsson et al., 2019), et plus particulièrement dans sa version *paired-end* (bidirectionnelle), lorsque le marqueur moléculaire est séquencé simultanément dans les deux sens, nommés *forward* et *reverse*, via des lectures de 250 nt qui peuvent se chevaucher. Les ADN des différents échantillons sont préalablement marqués avec un index, une étiquette moléculaire unique par échantillon, ce qui permet de les séquencer simultanément sur un même *run* de séquençage (Lindahl et al., 2013). Le nombre de lectures générées lors d'un run est une valeur fixe de l'ordre de 20 millions pour le MiSeq. Le nombre de lectures par échantillon dénommé *effort de séquençage* est souvent variable entres les différents échantillons composant le run malgré l'ajustement de la concentration en ADN (McMurdie et Holmes, 2014).

2.1.2 *Analyses bioinformatiques : des séquences au relevé des micro-organismes*

En écologie des communautés des macro-organismes, l'observation directe de l'écosystème étudié permet d'obtenir un relevé des espèces présentes, voire leurs abondances, sur plusieurs sites. Les approches de métabarcoding rompent le lien entre micro-organismes et séquences d'ADN, ce qui implique des analyses supplémentaires (Figure 4). Les analyses bioinformatiques appliquées aux données de séquences tentent de rétablir au mieux ces liens rompus à l'aide d'un ensemble d'étapes impliquant plusieurs programmes informatiques qui s'enchaînent, et dénommés collectivement *pipeline* (Bálint et al., 2014 ; Lindahl et al., 2013). Ces pipelines produisent un relevé des micro-organismes présents à partir des séquences (Bálint et al., 2016). Certains pipelines développés initialement pour l'analyse des séquences 16S ont été adaptés par la suite pour l'utilisation des marqueurs ITS et sont désormais largement utilisés à travers le monde. Parmi eux se trouvent des outils tels que mothur (Schloss et al., 2009), QIIME 1 et 2 (Bolyen et al., 2019 ; Caporaso et al., 2010), FROGS (Escudie et al., 2015 ; Escudié et al., 2018), VSEARCH (Rognes et al., 2016) et DADA2 (Callahan et al., 2016). D'autres pipelines ont été développés spécifiquement pour l'analyse des séquences ITS tels que SCATA (Durling et al., 2011), CloVR-ITS (White et al., 2013) et PIPITS (Gweon et al., 2015).

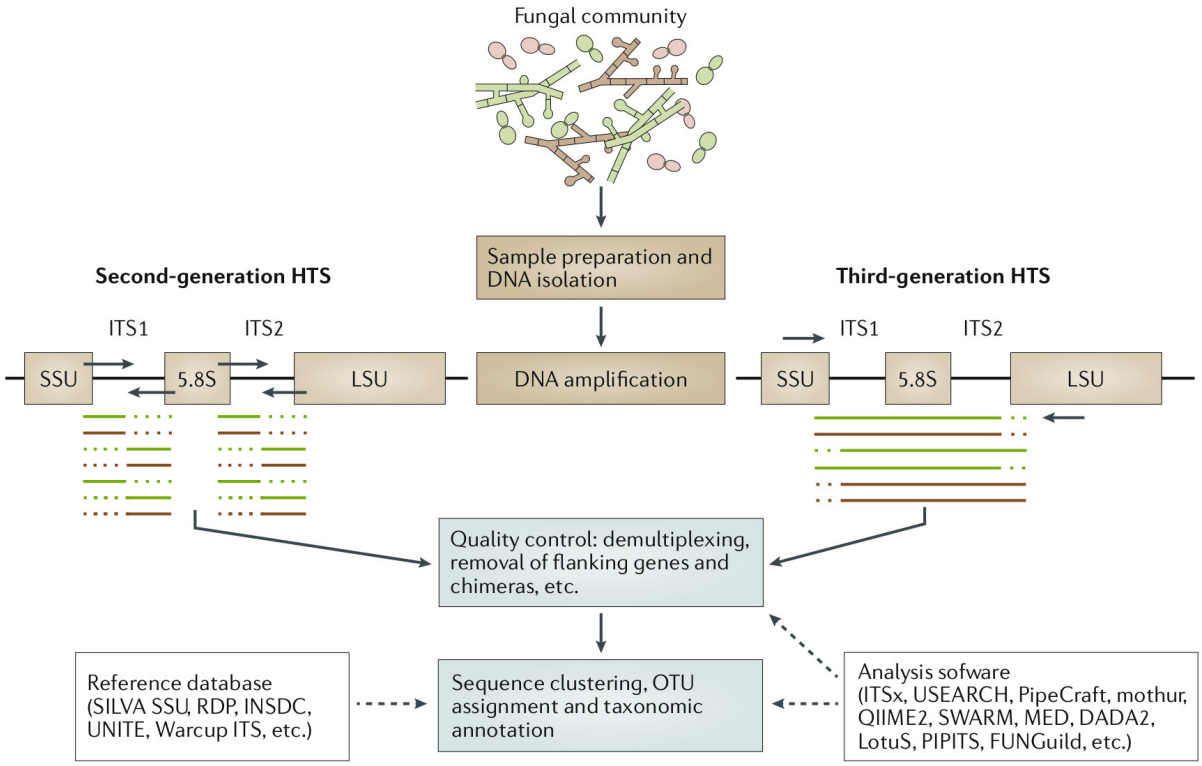


FIGURE 5 – Principales étapes du métabarcoding illustrées pour les champignons. Une communauté fongique est échantillonnée, l'ADN est extrait puis un marqueur moléculaire issu de l'ADN ribosomique est amplifié (petite, *i.e.* SSU, ou grande, *i.e.*LSU, sous-unité, ou encore ITS1 et/ou ITS2 illustrés par les flèches noires). Le séquençage à haut-débit (HTS) est ensuite réalisé par des technologies de seconde (à gauche) et troisième génération (à droite) qui produisent chacune des séquences de tailles variables. Un grand nombre d'outils est ensuite nécessaire pour traiter les séquences obtenues et identifier les micro-organismes présents. Source : NILSSON et al. (2019).

Les quatre étapes principales d'un pipeline pour caractériser les communautés microbiennes sont :

Étape 1 le pré-traitement des séquences pour obtenir des séquences de qualité

Étape 2 l'analyse des variations génétiques des séquences de qualité

Étape 3 l'élimination des séquences artéfactuelles et des contaminants

Étape 4 l'incrément d'informations sur les séquences.

Étape 1 – Pré-traitement des séquences pour obtenir des séquences de qualité

Les fichiers bruts issus du séquenceur doivent être traités afin d'attribuer les séquences à chacun des échantillons en utilisant les index utilisés pour combiner plusieurs échantillons dans un même run (Bálint et al., 2014). Les amorces qui ont permis l'amplification spécifique du marqueur moléculaire sont également séquencées et présentes (parfois partiellement, parfois en plusieurs exemplaires) dans les séquences (Bálint et al., 2014). Certains auteurs recommandent de les retirer afin de limiter l'inflation de séquences artéfactuelles (Bálint et al., 2014 ; Lindahl et al., 2013) et un outil répandu, cutadapt, réalise parfaitement cette tâche (Martin, 2011).

Dans le cas d'un séquençage bidirectionnel, les lectures *forward* et *reverse* correspondent à la même séquence lue dans les deux sens, et ces deux paires peuvent être assemblées pour être davantage sûr de la séquence (Nguyen et al., 2015). Pour les marqueurs moléculaires dont la longueur est peu variable tels que les régions du 16S bactérien (Palmer et al., 2018), il est possible de définir une longueur minimale attendue de chevauchement entre les paires comme option des outils bioinformatiques, voire d'exiger un consensus complet entre les paires (Eren et al., 2013). Cependant, les variations de longueur des séquences d'ITS limitent l'établissement de tels seuils fixes (Tedersoo et al., 2015) et impliquent l'utilisation d'outils d'assemblage des paires flexibles, tels que PEAR (Zhang et al., 2014). Le marqueur est parfois tellement long que les paires ne peuvent se chevaucher, ou tellement court que les paires dépassent le marqueur moléculaire de chaque côté (Zhang et al., 2014). L'absence d'assemblage des paires au profit de l'utilisation d'une des deux seulement peut se justifier dans le cas où aucun chevauchement entres les paires n'est possible ou si la qualité des séquences d'une des paires est trop faible (Nguyen et al., 2015).

Un score de qualité est attribué par nucléotide pour chaque séquence. Ce score, dénommé Phred score et noté Q, est défini tel que $Q = -10 \log_{10} p$ où p indique la probabilité d'erreur de séquençage du nucléotide. Par exemple, un Phred score de $Q = 30$ reflète une probabilité d'erreur de $p = 0.001$ (Ewing et Green, 1998). A partir des scores de qualité, plusieurs approches existent pour pré-traiter les séquences. Il est par exemple possible de retirer les séquences si plus de 75% des nucléotides consécutifs sont de mauvaise qualité ($Q < 3$). Il est également possible de tronquer les séquences si les nucléotides de mauvaise qualité représentent moins de 75% de la longueur (Bokulich et al., 2013). D'autres auteurs ont plutôt opté pour conserver les séquences dans leur intégralité mais en calculant le nombre d'erreurs attendues dans une séquence, noté E et défini comme $E = \sum_i 10^{-Q/10}$ (Edgar et Flyvbjerg, 2015). Les séquences avec plus d'une erreur attendue sont ensuite retirées des données (Edgar et Flyvbjerg, 2015).

Une dernière étape possible de pré-traitement consiste à vérifier que les séquences correspondent bien au marqueur ciblé en identifiant les sous-régions conservées et à centrer l'analyse sur les sous-régions variables qui sont les plus informatives (Bálint et al., 2014 ; Lindahl et al., 2013). Dans le cas du marqueur ITS, ces régions conservées – 18S, 5.8S ou 26S – sont détectables même partiellement par l'outil ITSx grâce à une base de données contenant des alignements de ces régions (Bengtsson-Palme et al., 2013). ITSx réalise ensuite une extraction sélective de sous-régions variables telles que l'ITS1 (Bengtsson-Palme et al., 2013). Une extension récente de cet outil conserve les scores de qualité des séquences pour chaque nucléotide permettant leur utilisation par d'autres outils en aval du pipeline (Rivers et al., 2018). L'outil metaxa2 est similaire à ITSx mais a été développé pour les marqueurs bactériens (Bengtsson-Palme et al., 2015).

Étape 2 – Analyse des variations génétiques des séquences de qualités

A cette étape, les séquences similaires peuvent être regroupées en unités bioinformatiques nommées *Operational Taxonomic Unit* (OTU ; Bálint et al., 2016). Suivant la résolution taxonomique du marqueur et le seuil de similarité choisis, ces entités sont considérées comme des substituts bioinformatiques d'un genre ou d'une espèce de micro-organismes. La valeur de 97% de similarité entre des séquences de 16S a été proposée comme un seuil séparant les espèces bactériennes (Stackebrandt et Goebel, 1994), puis augmentée à environ 99% grâce à des banques de séquences bien plus fournies (Edgar, 2018). Chez les champignons, un seuil fixe à 97% est fréquemment utilisé pour les deux régions ITS puisqu'il permet d'identifier correctement les espèces dans des communautés artificielles (Bakker, 2018 ; Tedersoo et al., 2015). Il est cependant possible que des séquences d'espèces distinctes – mais proches – soient agglomérées entres elles avec ce seuil (Ryberg, 2015). Mais le choix exact du seuil n'a pas d'effet sur les conclusions écologiques lorsqu'un processus structure fortement les communautés étudiées (Botnen et al., 2018). Certains auteurs soutiennent qu'un seuil variable selon les groupes taxonomiques de champignons devraient tenir lieu de définition bioinformatique de l'espèce, un système implémenté dans la banque de séquences ITS UNITE sous la forme d'espèces hypothétiques, les *species hypothesis* (Kõljalg et al., 2013).

Trois stratégies existent pour générer des OTU à partir des séquences (Navas-Molina et al., 2013). La stratégie *closed reference* consiste à comparer les séquences obtenues avec une banque de séquences existantes avec des OTU pré-définies telle que UNITE (Kõljalg et al., 2013) ou SILVA (Quast et al., 2012). À l'inverse, la stratégie *de novo* démarre des séquences directement et rassemble les séquences similaires entre elles à un seuil donné pour générer des OTU selon différents algorithmes possibles. Ces algorithmes de *clustering* tels que *nearest, furthest or average neighbor* peuvent respectivement regrouper les séquences dans une OTU si la similarité d'une séquence au moins, de toutes les séquences, ou des séquences en moyenne sont inférieures au seuil fixé (Navas-Molina et al., 2013 ; Westcott et Schloss, 2017). D'autres algorithmes limitent les calculs coûteux – en temps et en mémoire – de ces approches en commençant les comparaisons avec un type de séquences : les plus longues ou les plus abondantes comme dans *VSEARCH* (Rognes et al., 2016). Finalement, la stratégie *open reference* est hybride et applique une stratégie *closed reference* sur les séquences qui le peuvent suivie d'une stratégie *de novo* sur les restantes. Dans le cas du marqueur fongique ITS, les stratégies *closed reference* avec une banque de séquences parviennent mieux à identifier les espèces fongiques dans une communauté artificielle de séquences d'ITS de UNITE (Halwachs et al., 2017) ou d'ADN fongique séquencé (Cline et al., 2017). Malgré ces avantages, les séquences représentatives des OTU générées par cette stratégie dépendent de l'ordre d'apparition des séquences dans la

banque de référence et les premières séquences sont favorisées (Westcott et Schloss, 2015).

Les distances entre séquences au sein d'une même OTU ne correspondent pas toujours au seuil fourni pour les reconstruire mais la stratégie *de novo* ou l'algorithme de clustering *average neighbor* présentent de bons résultats (Westcott et Schloss, 2015). D'autres approches se développent donc pour définir les OTU de façon alternative en prenant en compte cette contrainte. Par exemple, *OptiClust* place initialement les séquences uniques au hasard dans des OTU générées avec un seuil fixe (Westcott et Schloss, 2017). Ce placement est ensuite réévalué en continu jusqu'à maximiser une métrique qui évalue si deux séquences doivent bien être regroupées dans la même OTU, en sachant leur distance à toutes les autres séquences (Westcott et Schloss, 2017). *Swarm* a été développé comme alternative aux approches *de novo* nécessitant un seuil fixe. En effet, *Swarm* démarre avec des séquences uniques qui sont connectées entre elles tant qu'elles possèdent une seule différence de nucléotides (insertion, substitution ou délétion) pour former les limites naturelles des OTU (Mahé et al., 2014 ; Mahé et al., 2015). Les ensembles de séquences reliées entre elles sont ensuite pondérés par leurs comptages respectifs pour identifier les séquences faiblements abondantes afin de séparer ces ensembles en OTU au niveau de ces séquences (Mahé et al., 2014 ; Mahé et al., 2015).

Une unité bioinformatique supplémentaire nommée *Amplicon Sequence Variants* (ASV) a été créée récemment. Elle regroupe les variants exacts de séquences, ce qui correspond à un seuil de 100% de similarité (Callahan et al., 2017). Le principe consiste à identifier parmi la multitude de séquences uniques celles qui comportent des erreurs afin de les éliminer et faire ressortir la variabilité naturelle des marqueurs moléculaires (Callahan et al., 2016). Plusieurs outils permettent de générer ces unités tels que *DADA2* (Callahan et al., 2016), *deblur* (Amir et al., 2017) ou *UNOISE2* (Edgar, 2016), bien que ce dernier n'incorpore pas la qualité des nucléotides durant l'identification des ASV.

La comparaison de DADA2, deblur, UNOISE2 et d'une stratégie *open reference* basée sur VSEARCH indique que toutes retrouvent la composition attendue d'une communauté artificielle, mais que le nombre d'unités incorrectement identifiées est bien supérieur avec l'*open reference* (Nearing et al., 2018). Les ASV permettent d'identifier les variants de façon unique à l'aide de leurs séquences et donc sans banque de référence (Callahan et al., 2017). Par conséquent, la comparaison des occurrences de ces ASV entre différents habitats et études ne nécessitent pas de devoir réanalyser la totalité des séquences (Thompson et al., 2017). De plus, la fine variabilité génétique détectée par les ASV est de l'ordre des similarités entre marqueurs de la même espèce – environ 99% – issues des analyses des banques de séquences bactériennes et fongiques (Edgar, 2018 ; Vu et al., 2019). Toutefois, ni OTU ni ASV ne peuplent les communautés microbiennes et le choix de l'unité bioinformatique doit dépendre principalement du rang taxonomique concerné par la question scientifique (McLaren et Callahan, 2018).

Étape 4 – Elimination des séquences artéfactuelles et des contaminants

L'amplification d'un brin d'ADN peut être interrompue à mi-parcours par le détachement des amorces et se poursuivre sur un brin d'ADN d'un autre micro-organisme, créant ainsi un marqueur moléculaire hybride : une chimère. Ces séquences chimériques nuisent aux analyses en créant des OTU ou des ASV artefactuels. La fréquence des chimères, estimée sur des communautés artificielles, est relativement faible (environ 0.2%) pour la région ITS (Aas et al., 2017 ; Bakker, 2018), mais beaucoup plus élevée (de 8 à 70%) pour la région 16S (Haas et al., 2011 ; Schloss et al., 2011). La formation des chimères est influencée par la nature de la polymérase

et est contre-intuitivement diminuée avec des polymérases plus propices aux erreurs (Bakker, 2018). UNITE compile et met à jour les séquences chimériques d'ITS (Nilsson et al., 2015). Plusieurs programmes bioinformatiques ont été développés pour essayer d'identifier et de retirer les séquences considérées chimériques, comme UCHIME qui est également implémentés dans VSEARCH et DADA2 (Callahan et al., 2016 ; Edgar et al., 2011 ; Nilsson et al., 2010 ; Rognes et al., 2016).

La grande sensibilité des technologies de séquençage haut-débit ainsi que le nombre élevé de séquences générées (plusieurs milliers/millions) rendent nécessaire l'inclusion de différents témoins – positifs et négatifs – lors d'un run afin de s'assurer de l'absence de contaminants provenant des étapes d'échantillonnage, de biologie moléculaire ou encore du séquençage (Hornung et al., 2019 ; Pollock et al., 2018). Galan et al. (2016) ont réalisé une exploration pionnière du contenu de différents témoins (d'extraction, d'amplification, de séquençage) qui leur a permis de développer une méthode de filtre de la table d'ASV. Dans cette méthode, les seuils de contamination croisée T_{CC} et de fausse attribution T_{FA} ont été utilisés pour retirer les ASV de tous les échantillons où leurs comptages comportent moins de séquences que l'un ou l'autre seuil. Le seuil de contamination croisée T_{CC} est défini comme le nombre maximal de séquences de chaque ASV trouvés dans les échantillons témoins négatifs ou positifs. Le seuil de fausse attribution T_{FA} est défini comme $T_{FA} = N \times R_{fa}$ où R_{fa} indique le nombre de séquences le plus élevé d'une souche témoin positif dans un échantillon non témoin, divisé par le nombre total de séquences de la souche dans le run, et N le nombre total de séquences de chaque ASV. Une autre méthode, *decontam*, identifie les contaminants comme des ASV fréquemment détectés dans les échantillons avec une faible concentration d'ADN et dans les échantillons témoins négatifs (Davis et al., 2018). Une autre méthode, *decontam*, identifie les contaminants comme les ASV plus fréquemment détectés dans les témoins négatifs (Davis et al., 2018). L'inclusion des quantités d'ADN de chaque échantillon avant le séquençage, si elles sont disponibles, permet d'améliorer la détection des contaminants avec decontam (Davis et al., 2018).

Après les étapes de nettoyage et filtrage des données présentées précédemment, les OTU ou ASV obtenues dans les différents échantillons servent à la construction d'une table, qui sera le point de départ de multiples analyses. La table d'OTU ou d'ASV est un tableau à double entrée qui contient le nombre de séquences par OTU ou ASV, et par échantillon. Ces comptages correspondent à une approximation biaisée de l'abondance des espèces (Lamb et al., 2019). Néanmoins, la fiabilité de ces comptages doit être assurée au maximum. En effet, des erreurs peuvent être introduites dans les séquences par la polymérase utilisée lors de l'amplification (Shagin et al., 2017) ou lors du séquençage (Schirmer et al., 2015) qu'il est nécessaire de corriger – pendant l'analyse de variations de séquences (p. 15) – ou d'éliminer pour réaliser des inférences écologiques de qualité (Sommeria-Klein et al., 2016). Postuler que les séquences erronées sont beaucoup moins abondantes que les séquences microbiennes ciblées semble raisonnable mais il reste cependant difficile de distinguer les séquences erronées des séquences issues de micro-organismes "rares", c'est-à-dire en faible abondance dans l'environnement (Lynch et Neufeld, 2015). A l'heure actuelle, il n'existe pas de consensus sur la manière de filtrer les ASV (ou OTU) dites rares, la méthode la plus communément utilisée étant d'enlever les unités avec un faible nombre de séquences (1 à 10 séquences) ou une faible abondance relative (inférieure à 0.005%) (Bokulich et al., 2013 ; Brown et al., 2015). Une approche récente, LULU, propose de regrouper les OTU (ou ASV) potentiellement erronées avec les OTU (ou ASV) parents dont ils sont issus (Frøslev et al., 2017). Cette méthode a permis d'accentuer la

concordance entre le relevé de biodiversité effectué visuellement et via le métabarcoding sur des plantes, et elle possède l'avantage de conserver le nombre de séquences total de la table d'OTU (ou ASV) et donc de limiter la perte d'information pour les analyses en aval (Frøslev et al., 2017).

Étape 4 – Incrément d'informations sur les séquences

Une fois la table d'ASV (ou d'OTU) filtrée, l'étape suivante consiste à incrémenter les ASV ou les OTU avec des informations complémentaires telles que leur taxonomie ou leurs fonctions potentielles. Pour cela, les séquences sont analysées et comparées à des banques de séquences de référence – telle que UNITE (Kõljalg et al., 2013) ou SILVA (Quast et al., 2012) – grâce à différents outils et approches tels que BLAST (Altschul et al., 1990), *Ribosomal Database Project* (RDP) *Classifier*(Wang et al., 2007), metaxa2 (Bengtsson-Palme et al., 2015) ; ou des approches consensuelles entres plusieurs outils avec CONSTAX (Gdanetz et al., 2017). Concernant l'étude des communautés fongiques, le RDP Classifier, qui utilise la composition en nucléotides des séquences de la banque pour assigner les ASV (ou OTU) dans la hiérarchie taxonomique, présente d'excellentes performances de classification sur l'ITS (Bokulich et al., 2018). La performance des outils dépend fortement de l'exhaustivité et de la qualité des banques de référence (Nilsson et al., 2017). Des approches récentes tentent ainsi de limiter une confiance excessive dans l'assignation taxonomique donnée par un outil en incluant l'incertitude sur les banques de référence sous la forme d'une probabilité qu'aucune séquence représentative ne soit dans la banque (Abarenkov et al., 2018 ; Somervuo et al., 2016 ; Somervuo et al., 2017). Une fois une taxonomie établie pour les ASV ou OTU, il est possible d'inférer les fonctions et rôles potentiels des espèces identifiées en comparant leur taxonomie à celle d'autres espèces dont les fonctions sont documentées dans la littérature. Les base de données FUNGuild (Nguyen et al., 2016) et FAPROTAX (Louca et al., 2016a) documentent les liens taxonomie-fonction pour les champignons et les bactéries.

2.1.3 *Analyses statistiques : extraire de l'information du relevé des micro-organismes*

Les comptages de séquences obtenus par ASV (ou OTU) correspondent à des proportions par rapport au nombre total de séquences (effort de séquençage) de l'échantillon considéré (voir le dernier §, section 2.1.1, p. 12 ; Gloor et al., 2017). Ces données sont contraintes par une somme fixe et sont dites compositionnelles (Aitchison, 1982 ; Quinn et al., 2018). Les nombres de séquences des unités (ASV ou OTU) d'un même échantillon ne sont donc pas indépendants. Par exemple si la proportion d'un ASV augmente, les proportions des autres ASV seront mécaniquement diminuées (Gloor et al., 2017). De plus, l'effort de séquençage varie entre échantillons pour de multiples raisons : la difficulté d'ajuster uniformément la concentration d'ADN, le choix de la technologie de séquençage ou encore de façon aléatoire (Gloor et al., 2017 ; Nilsson et al., 2019). La compositionnalité des données ainsi que la variation de l'effort de séquençage entre échantillons sont des caractéristiques des données de séquençage à considérer obligatoirement dans les analyses ultérieures de diversité ou de réseaux (Bálint et al., 2016 ; McMurdie et Holmes, 2014).

La diversité α correspond à la diversité des espèces dans un échantillon (Legendre et Legendre, 2012, p. 258), et plusieurs indices permettent de l'estimer (Daly et al., 2018). Par exemple, la richesse spécifique est définie comme le nombre d'espèces présentes dans un échantillon. Dans le cas des analyses de communautés microbiennes, la richesse correspond au nombre d'unités taxonomiques (OTU ou ASV) dans un échantillon. L'abondance relative des

espèces peut être incluse dans l'estimation de la diversité. De nombreux autres indices existent et pondèrent différemment les espèces rares et dominantes (Hill, 1973 ; Lucas et al., 2017). Par exemple, l'indice Inverse Simpson ${}^2D$ est défini suivant la notation unifiée proposée par Hill (Hill, 1973) comme ${}^2D = 1/\sum_i p_i^2$ où p_i indique la proportion de l'espèce i dans la communauté (Simpson, 1949). Cet indice offre l'avantage d'être interprétable directement comme le nombre effectif d'espèces d'une communauté en donnant plus de poids au espèces dominantes qu'aux rares (Jost, 2006). Ainsi, une communauté de trois espèces dont les abondances relatives sont 49%,49% et 2% a un indice Inverse Simpson de ${}^2D = 2.08$, indiquant la dominance de deux espèces. Enfin, une autre mesure de la diversité communément utilisée est l'équitabilité qui indique la répartition de l'abondance des espèces dans la communauté (Pielou, 1966). Une équitabilité de 0 et de 1 indique, respectivement, une communauté avec une unique espèce dominante et une communauté avec des abondances homogènes entre espèces (Pielou, 1966). Richesse, diversité et équitabilité sont des indices reliés, une communauté avec une équitabilité totale aura ainsi une diversité égale à la richesse (Hill, 1973 ; Jost, 2006).

La diversité β correspond à la variation de la composition des espèces entre échantillons (Legendre et Legendre, 2012, p. 258). La variation de composition est quantifiée pour chaque paire d'échantillons à l'aide d'indices de dissimilarité calculés à partir du nombre d'espèces communes, noté a, à la paire d'échantillons et du nombre d'espèces uniques à chacun des deux échantillons notés b et c (Chao et al., 2006). Par exemple, la dissimilarité binaire de Jaccard est définie comme $(b + c)/(a + b + c)$ (Jaccard, 1900 ; Legendre et De Caceres, 2013). C'est une dissimilarité utile en raison de sa capacité à détecter les groupes d'échantillon similaires (Kuczynski et al., 2010) et parce qu'elle est invariante au nombre total d'espèces par échantillon, contrairement à l'indice de Bray-Curtis (Legendre et De Caceres, 2013). Une version quantitative de la dissimilarité de Jaccard est définie comme $(U \times V)/(U + V - U \times V)$ où U et V indiquent respectivement la somme des abondances relatives dans chacun des deux échantillons pour les espèces communes (Chao et al., 2006). L'utilisation combinée des versions quantitative et binaire de cette dissimilarité donne des informations complémentaires puisque la version binaire donnent une importance égale aux espèces rares (Chao et al., 2006 ; Jaccard, 1900). La dissimilarité appliquée à toutes les paires d'échantillons forme une matrice de distance entre échantillons (Legendre et De Caceres, 2013). Elle est classiquement analysée avec un test statistique, la *Permutational Multivariate ANalysis Of VAriance* (PERMANOVA ; Anderson, 2001) qui permet d'évaluer l'effet des variables sur les différence de compositions entre échantillon. De plus, une analyse de dispersion à partir de la même matrice de distance permet d'évaluer l'hétérogénéité au sein d'un groupe d'échantillons en comparant les distances de chaque échantillon au centroïde de son groupe, ce qui permet de vérifier également l'homogénéité de variance des groupes qui est postulée dans la PERMANOVA (Anderson, 2006).

Il est possible d'utiliser ces analyses tout en prenant en compte la compositionnalité des données en ayant recours à la transformation des données (Li et al., 2016). En effet, la transformation des données de comptage avec le log ratio centré puis l'utilisation d'une distance euclidienne permet d'analyser les différences entre échantillons en prenant en compte cette caractéristique (Gloor et al., 2017). Cependant, cette transformation n'admet pas de comptages nuls à cause de la transformation au log. Différentes méthodes de remplacement des zéros ont été suggérées par Gloor et al. (2017) et sont proposées dans les librairies R zCompositions (Martín-Fernández et al., 2015) ou ALDEx2 (Fernandes et al., 2014). Elles considèrent que les comptages nuls sont dus à un échantillonnage limité et non à des absences (Martín-Fernández et al., 2015). Un

pseudo-comptage – de 0.5 – est ajouté aux comptages puis de multiples ré-échantillonnages de ces nouvelles abondances relatives sont réalisés par échantillon pour estimer la probabilité de détecter ces comptages nuls avec l'effort de séquençage de l'échantillon (Fernandes et al., 2014 ; Martín-Fernández et al., 2015).

Si une différence de composition entre échantillons est mise en évidence, il est naturel d'identifier les micro-organismes différentiellement abondants entre échantillons. Il s'agit plus exactement d'identifier les ASV (ou OTU) dont les comptages sont changeants entre conditions et plusieurs approches ont été développées à cette fin. Les comparaisons de ces approches indiquent qu'il n'existe pas d'outil universel, utilisable dans toutes les situations, car la performance des approches dépend du nombre d'échantillons par condition et de la différence d'effort de séquençage (Thorsen et al., 2016 ; Weiss et al., 2017). Ainsi, la librairie R DESeq2 (Love et al., 2014) est recommandée lorsque moins de 20 échantillons sont disponibles par condition car elle possède une bonne capacité de détection des différences entres conditions, qui correspondent à des estimations du log ratio des comptages entre les conditions (Weiss et al., 2017). La librairie ALDEx2 (Fernandes et al., 2014) permet de prendre en compte la compositionnalité et fournit des résultats robustes et peu de faux-positifs (Fernandes et al., 2018 ; Thorsen et al., 2016).

2.2 Comment identifier les interactions ?

2.2.1 *Principe de reconstruction des réseaux microbiens*

Les interactions biotiques et les exigences abiotiques de l'environnement causent des fluctuations spatio-temporelles dans l'abondance des espèces (Hawkes et Connor, 2017). Ces fluctuations peuvent être mesurées par des relevés quantitatifs des espèces présentes à différents temps et différents sites (Figure 6) grâce au métabarcoding. L'inférence de réseau consiste donc à établir des associations statistiques entre les comptages des séquences des OTU (ou ASV) obtenues avec le métabarcoding (Faust et Raes, 2012 ; Vacher et al., 2016b). L'identification correcte des nœuds (OTU ou ASV) est donc un prérequis à l'inférence des associations qui doivent se baser sur des comptages fiables et sur une résolution taxonomique précise des OTU ou ASV (Berry et Widder, 2014 ; Cordero et Datta, 2016 ; McLaren et Callahan, 2018). Une large gamme d'outils d'inférence de réseaux dédiés aux données de séquençage ont été développés (revus dans Faust et Raes, 2012 ; Layeghifard et al., 2017 ; Röttjers et Faust, 2018) bien qu'un nombre plus restreint d'outils soient utilisés dans les études des communautés microbiennes en raison de leur rapidité, précision ou encore leur facilité d'utilisation.

2.2.2 *Barrières à la reconstruction*

Les réseaux d'associations inférés possèdent souvent deux types de liens : positifs ou négatifs (voir Agler et al., 2016 ; Durán et al., 2018 ; Faust et al., 2015 ; Jakuschkin et al., 2016). Certaines interactions biotiques directes, comme la compétition par interférence ou le mutualisme, induisentdes associations négatives ou positives (Figure 7). Mais des interactions biotiques indirectes causent également des associations (Figure 7). Par exemple, dans la compétition apparente entre deux espèces qui partagent un prédateur peut induire des associations négatives (Holt et Bonsall, 2017 ; Li et al., 2016). Ainsi cette classification binaire des liens – positif et négatif– ne rend pas compte pleinement des différentes interactions biotiques (Figure 7 ; Weiss et al., 2016). De plus, si deux micro-organismes partagent les mêmes exigences abiotiques,

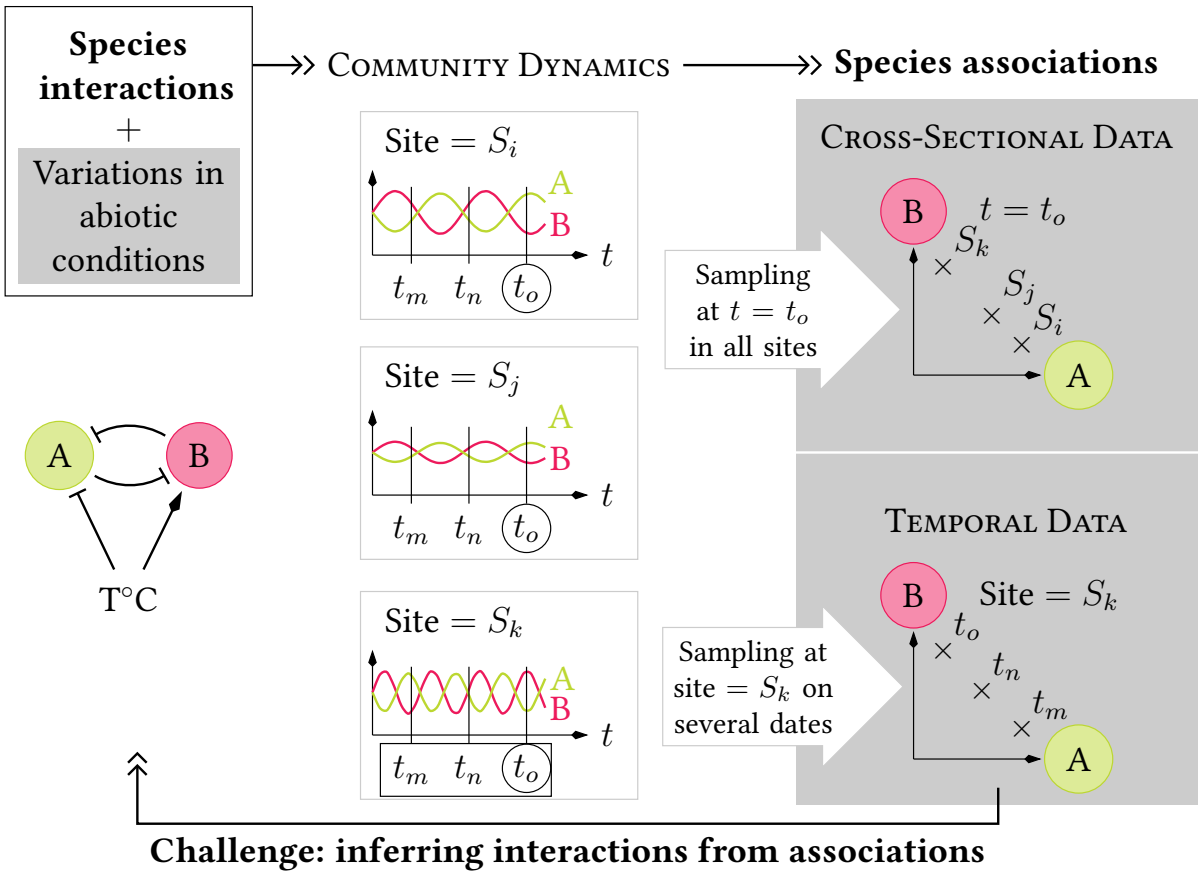


Figure 6 – Principes de reconstruction des réseaux d'interactions entre espèces à partir des profils d'abondance des espèces. La structure des réseaux d'interaction entre les espèces et les variations de l'environnement abiotique causent des variations spatiales et temporelles dans l'abondance des espèces. Ces variations peuvent être détectées en échantillonnant plusieurs sites géographiques à une date (encerclé), en échantillonnant un site à plusieurs dates (encadré) ou en combinant les deux plans d'échantillonnage (non montré). Le défi consiste à reconstituer le réseau d'interaction des espèces (totalement ou partiellement inconnu) sur la base des variables et des relations qui peuvent être mesurées (en gris), comme les associations statistiques entre les abondances des espèces et les variables environnementales abiotiques. Source : Derocles et al. (2018) et Annexe 1, p. 167.

des associations positives ou négatives peuvent être générées selon que ces exigences sont similaires ou opposées (Figures 6, 7 ; Armitage et Jones, 2019). Au-delà de la capture de signaux biotiques et abiotiques, des associations peuvent être générées par des biais méthodologiques ce qui indique que plusieurs barrières méthodologiques demeurent et renforcent le décalage entre réseaux inférés et réseaux d'interactions (Vacher et al., 2016b).

La compositionnalité des données de séquençage induit des associations statistiques fallacieuses qui se retrouvent dans les réseaux inférés (Chiquet et al., 2019 ; Fang et al., 2015 ; Friedman et Alm, 2012 ; Li et al., 2016 ; Weiss et al., 2016). Dans une communauté simulée de 5 espèces par exemple, où deux espèces n'interagissent pas entre elles, leurs abondances sont associées négativement par la corrélation de Pearson une fois converties en abondances relatives (Li et al., 2016). La présence de nombreux zéros dans les comptages des tables d'ASV (ou d'OTU) induit des corrélations positives fallacieuses entre certaines paires d'ASV (ou OTU) à cause de nombreux zéros assortis (Li et al., 2016 ; Röttjers et Faust, 2018). Les simulations indiquent que la précision des méthodes de reconstruction de réseaux est très faible lorsque la table contient plus de 50% de zéros (Weiss et al., 2016). Pouvoir tester ou non les corrélations dépendrait de la prévalence des deux ASV (ou OTU) d'après des travaux théoriques indiquant que les valeurs extrêmes des statistiques de ces tests dépendent des prévalences (Cougoul et al., 2019a).

L'inférence est difficile puisque les données de séquençage sont de grandes dimensions, c'est-à-dire que le nombre d'ASV (ou d'OTU) est bien plus élevé que le nombre d'échantillons : des miliers ou plus d'ASV (ou OTU) détectées dans des centaines (parfois moins) d'échantillons (Bálint et al., 2016 ; Kurtz et al., 2015). Toutefois les simulations indiquent que les performances des méthodes d'inférence se stabilisent au-delà de 200 échantillons mais en ne considérant que 50 espèces (Hirano et Takemoto, 2019). Une sélection des ASV (ou OTU) à inclure est donc nécessaire avant l'inférence de réseau mais il n'existe pas de recommandations claires entre conserver les ASV (ou OTU) abondantes, prévalentes ou les deux (Röttjers et Faust, 2018). Souvent, les ASV (ou OTU) sont conservées pour l'inférence si elles sont : présentes dans au moins 20% ou 25% des échantillons (Berry et Widder, 2014 ; Durand et al., 2017), ou encore dans au moins dans 2 (Durán et al., 2018), 5 (Poudel et al., 2016) ou 10 échantillons (Agler et al., 2016). L'absence d'un seuil consensuel de prévalence à partir duquel les espèces doivent être conservées peut provenir d'une réticence à filtrer les ASV (ou OTU) peu fréquentes ou rares qui pourraient être des spécialistes ou avoir un rôle important (Harrison et al., 2019 ; Lynch et Neufeld, 2015). Ainsi une approche alternative conserve les ASV (ou OTU) dans la table mais certaines associations entre paires d'ASV (ou d'OTU) sont interdites (et non inférées) basées sur leur testabilité calculé sur la prévalence des ASV (et d'OTU) de chaque paire Cougoul et al. (2019a).

Des associations indirectes entre ASV (ou OTU) peuvent se manifester lorsque les réseaux sont denses et créer des associations potentiellement fallacieuses dues à l'interdépendance entre ASV (ou OTU) (Berry et Widder, 2014 ; Li et al., 2016 ; Röttjers et Faust, 2018). Si deux espèces A et B sont chacune en compétition avec une troisième espèce C, les abondances de A et B seront associées (Holt et Bonsall, 2017) mais identifier cette association comme indirecte nécessiterait de connaître l'abondance de l'espèce C. Ce même principe opère si des ASV (ou OTU) ne sont pas sélectionnés pour l'inférence de réseaux ou bien lorsque qu'ils ne sont pas observés. L'absence d'observation d'une ASV (ou OTU) peut résulter (1) d'un pipeline de métabarcoding inadéquat à au moins une étape (voir p. 12) tel qu'un filtre trop drastique,

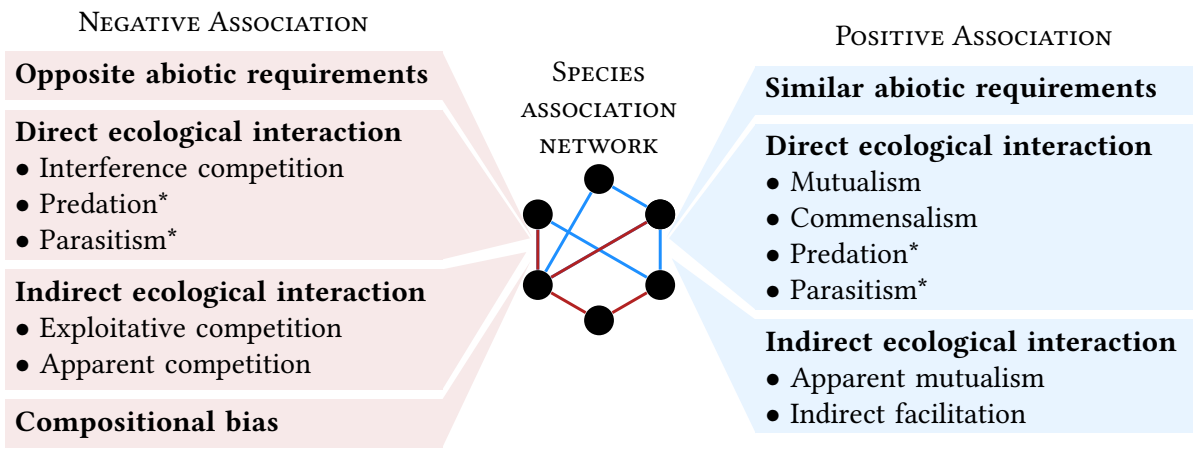


Figure 7 – Les réseaux d'associations d'espèces ne sont pas des réseaux d'interactions d'espèces. Des associations significatives entre les abondances des espèces peuvent être engendrées par des interactions écologiques mais aussi par les exigences abiotiques des espèces et les biais méthodologiques. De plus, il n'est pas facile de relier le signe d'association statistique au type d'interaction écologique (*les cas de prédation et de parasitisme sont discutés dans le texte). Ces obstacles doivent être surmontés pour reconstituer les réseaux d'interactions entre espèces à partir des données sur l'abondance des espèces, comme celles obtenues à l'aide d'approches de métabarcoding. Adapté de Derocles et al. (2018) et Annexe 1, p. 167.

(2) d'un faible effort de séquençage empêchant sa détection, (3) d'une absence d'amplification, (4) ou encore d'un règne microbien non inclus dans l'approche de métabarcoding. Le règne manquant peut même être non microbien tel les acariens qui consomment le mycélium des champignons foliaires (English-Loeb et al., 2007).

2.2.3 *Méthodes de reconstruction des réseaux*

Méthodes basées sur les corrélations

Les méthodes de co-occurrences ou de corrélations ont été les approches initialement utilisées pour l'inférence de réseaux à partir des comptages de séquences issues du métabarcoding (Barberán et al., 2012 ; Chaffron et al., 2010). L'inclusion de liens dans le réseau dépend par exemple de la significativité du test exact de Fisher ou du test de la corrélation des rangs de Spearman (Barberán et al., 2012 ; Chaffron et al., 2010). Les associations peuvent aussi être incluses si leur score de corrélation est supérieur aux scores de corrélations obtenus sur des tables de comptages permutées aléatoirement (Connor et al., 2017). Cependant, la nature compositionnelle des données de séquençage induit des corrélations fallacieuses (Friedman et Alm, 2012 ; Li et al., 2016 ; Weiss et al., 2016). Une autre approche, robuste à la compositionnalité, *SparCC*, contourne ce problème en prenant pour chaque paire d'ASV le logarithme du rapport des comptages des ASV de la paire puisque ce rapport est égal au rapport des "vraies" abondances inconnues (Aitchison, 1982 ; Friedman et Alm, 2012). Les corrélations entre paires d'ASV (ou d'OTU) peuvent être estimées à partir de la variance de ces rapports sous réserve que de nombreux ASV (ou OTU) soient inclus dans l'inférence et que le réseau microbien est *sparse* c'est-à-dire qu'il possède peu de liens (Friedman et Alm, 2012). D'autres méthodes d'inférence postulent également que les réseaux microbiens sont *sparse* et donc que les matrices obtenues possèdent peu de valeurs non-nulles (e.g., Chiquet et al., 2019 ; Kurtz et al., 2015). CoNet ne présume rien de tel sur le réseau et cette méthode implémente les mesures de corrélation –

comme Pearson ou Spearman –, et ajoute des mesures de dissimilarités, telles que Jaccard, non pas entre échantillons (voir p. 19) mais entre les comptages des ASV (ou OTU) pour réaliser un consensus des méthodes par un système de vote (Faust et Raes, 2016). La pertinence des liens dans cette approche est ensuite établie par permutation des données. Les simulations de communautés microbiennes ont permis de comparer les performances de ces approches corrélatives (Weiss et al., 2016). Les résultats indiquent qu'il n'existe pas de méthode universelle et que les caractéristiques du jeu de données, telles que le nombre effectif d'espèces ou le type de données – multi-sites ou temporelles (Figure 6) – doivent guider le choix de la méthode, voire d'un ensemble de méthodes à combiner (Weiss et al., 2016). Les enseignements tirés de cette étude complètent des recommandations sur le nombre d'échantillons nécessaires – au moins 25– ou leur similarité –20% d'ASV communes– pour réaliser des inférences de réseaux fidèles aux simulations (Berry et Widder, 2014).

Méthodes basées sur les corrélations partielles

D'autres approches ont pris le parti de démêler les associations indirectes qui pouvaient potentiellement fausser les réseaux inférés. Ces approches sont basées sur des modèles graphiques gaussiens qui modélisent p variables aléatoires gaussiennes notées $X_{1,i,\ldots,p}$ définies comme $X \sim \mathcal{N}_p(\mu, \Sigma)$ et où la matrice de variance-covariance Σ, de taille $p \times p$ décrit les dépendances entres ces variables aléatoires (Chiquet et al., 2019 ; Lauritzen, 1996 ; Layeghifard et al., 2017). Deux propriétés des modèles graphiques gaussiens sont intéressantes pour l'inférence de réseau. Premièrement, une fois la matrice Σ inversée telle que $\Omega = \Sigma^{-1}$, une valeur nulle dans la matrice Ω indique l'indépendance entre les variables (Chiquet et al., 2019). Deuxièmement, une corrélation partielle ρ_{ij} peut être calculée entre les variables i et j et est définie comme $\rho_{ij} = -\Omega_{ij}/\sqrt{\Omega_{ii}\Omega_{jj}}$ (Chiquet et al., 2019). Cette corrélation partielle indique la corrélation entre deux variables toutes choses égales par ailleurs, et donc élimine les associations indirectes (Kurtz et al., 2015 ; Layeghifard et al., 2017). Mais ces propriétés ne sont valides que lorsque les variables considérées sont gaussiennes. Or ce n'est pas le cas des comptages de séquences utilisés pour l'inférence de réseaux microbiens.

Plusieurs méthodes se basent sur ces propriétés de la matrice de covariance et contournent le problème des variables non gaussiennes en transformant les données (Vacher et al., 2016b). Par exemple, *SpiecEasi* applique la transformation du log-ratio centré à la table d'ASV pour se rapprocher du cas gaussien (Kurtz et al., 2015). Cette transformation divise chacun le nombre de séquences par ASV par la moyenne géométrique des comptages de l'échantillon et applique le logarithme à chacun de ces ratios (Gloor et al., 2017 ; Li et al., 2016). Ensuite, la sélection des paires d'ASV (ou OTU) non indépendantes est réalisée grâce à une procédure nommée *Least Absolute Shrinkage and Selection Operator* (LASSO) qui vise à forcer les coefficients de la matrice à tendre vers 0 en appliquant une gamme de pénalités de plus en plus forte (Tibshirani, 1996). Dans le cas de SpiecEasi, c'est la matrice de covariance inverse Ω qui est pénalisée suite au postulat que les réseaux microbiens possèdent peu de liens et donc que la matrice Ω est *sparse* (Kurtz et al., 2015). Le LASSO peut être appliqué à Ω en entier pour reconstruire le réseau dans son ensemble (Friedman et al., 2008), ou à chaque entrée p de Ω pour identifier les liens de chaque nœuds pour finalement les réconcilier et produire un réseau complet (Kurtz et al., 2015). Toute la difficulté de cette stratégie réside dans le choix de la pénalité optimale λ comprise entre une pénalité maximale λ_{max} résultant en un réseau sans lien et une pénalité minimale λ_{min} résultant en un réseau dense. SpiecEasi identifie la valeur de la pénalité optimale à l'aide de l'approche *Stability Approach to Regularization Selection* (StARS) qui consiste à sous-échantillonner aléatoirement l'ensemble des données, reconstruire les réseaux et identifier

la valeur de λ générant un réseau avec une faible variabilité (Liu et al., 2010).

De façon similaire à SpiecEasi, les méthodes d'inférence suivantes sont basées sur la transformation des comptages et l'inversion de la matrice de covariance : CCLasso (Fang et al., 2015), REBACCA (Ban et al., 2015) ou encore MPLasso (Lo et Marculescu, 2017). Néanmoins dans le cas de MPLasso, l'identification des associations les plus probables peut être facilitée par l'incorporation d'informations *a priori* sur les associations entres micro-organismes provenant par exemple de la littérature (Lo et Marculescu, 2017). Cet apport d'informations est inclus en augmentant la pénalité λ des associations pour lesquels aucune co-occurrence significative dans la littérature n'est démontrée. A l'inverse, la pénalité est diminuée si la littérature supporte une telle co-occurrence voire décrit une interaction connue. L'ajustement de ces pénalités pour guider l'inférence améliore les réseaux reconstruits en comparaison avec d'autres méthodes (Lo et Marculescu, 2017).

D'autres approches considèrent que les comptages de séquences proviennent de variables dites latentes ou cachées comme alternative à la transformation des comptages. Les associations directes, comme la corrélation partielle, sont définies entres les variables latentes (non observées) et les comptages de séquences (observés) sont générés conditionnellement à ces variables latentes (Chiquet et al., 2019 ; Warton et al., 2015a). C'est le cas de MInt (Biswas et al., 2016) qui est un modèle hiérarchique Poisson-Normal multivarié où les variables latentes, une par ASV (ou OTU) suivent des distributions gaussiennes et les comptages des séquences des ASV (ou OTU) suivent des distributions de Poisson en sachant ces variables latentes. La matrice de covariance entre les variables latentes capture donc les associations entre les ASV (ou OTU) et l'estimation des liens est ensuite réalisée via la pénalisation de la matrice de covariance inverse (Biswas et al., 2016). Une approche similaire modélise conjointement les comptages de la table d'ASV à l'aide du modèle Poisson Log-Normal (PLN ; Aitchison et Ho, 1989). PLN considère que les comptages sont issus de distributions de Poisson dont la moyenne est fonction de l'effort de séquençage, des covariables et des variables latentes gaussiennes (Chiquet et al., 2017 ; Chiquet et al., 2019). Comme pour MInt, la matrice de covariance entre les variables latentes gaussiennes informe sur les associations entre les ASV (ou OTU) (Chiquet et al., 2017 ; Chiquet et al., 2019). Le choix de la pénalité optimale dans PLN peut se faire grâce à l'approche StARS comme dans SpiecEasi ou bien à l'aide du critère *Bayesian Information Criterion* (BIC) courant en statistiques mais adapté aux données en grandes dimensions (Chen et Chen, 2008). gCoda est similaire à CCLasso mais suppose que les abondances absolues des micro-organismes sont latentes, donc non observées, et suivent une distribution logistique normale, qu'il s'agira d'estimer afin d'informer sur les associations entre OTU (ou ASV) retrouvées grâce à l'inversion de la matrice de covariance (Fang et al., 2017). MAGMA se base sur des variables latentes gaussiennes mais considère les comptages générés par des distributions négatives binomiales avec inflation de zéros (Cougoul et al., 2019b).

Une approche alternative, FlashWeave, fournit des corrélations partielles mais n'est pas basée sur l'inversion de matrice de covariance mais plutôt sur l'identification pour chaque nœud de prédicteurs suffisants et nécessaires de ses comptages transformés tels que les comptages d'autres ASV (ou OTU) ou les variables environnementales (Tackmann et al., 2019). Une fois les prédicteurs de tous les nœuds identifiés, le réseau est reconstruit en indiquant un lien entre deux nœuds si au moins un est prédicteur de l'autre (Tackmann et al., 2019). Ovaskainen et al. (2017) a défini un ensemble de modèles à variables latentes, *Hierarchical Modelling of Species Communities* (HMSC), applicable à plusieurs types des données tels que des comptages

de séquences ou des données de présence-absence. Cette approche permet notamment de prendre en compte les relations phylogénétiques entre espèces et les traits associés à chacune d'elles en plus de leurs occurrences et des variables environnementales (Ovaskainen et al., 2017). D'autres méthodes développées récemment modélisent les abondances absolues des micro-organismes (Xiao et al., 2017 ; Yoon et al., 2019) qui ne sont malheureusement pas disponibles à partir des abondances relatives obtenues par le séquençage. Pour estimer les abondances absolues, les cellules microbiennes peuvent être dénombrées par la cytométrie en flux. La quantité d'ADN peut être estimée par fluorescence avec la *quantitative PCR* (qPCR) ou par des méthodes de micro-fluidiques plus récentes comme la *digital droplet PCR* (ddPCR) (Hindson et al., 2013 ; Porter et Hajibabaei, 2018). Mais ces abondances absolues concernent les cellules ou les ADN et ne sont pas directement superposables aux abondances relatives des ASV (ou OTU) suite au nombre de copies variables du marqueur ciblé dans les cellules et dans le génome (Ganley et Kobayashi, 2007 ; Rebollar et al., 2017). Ces deux types de données peuvent être combinées soit en multipliant les abondances relatives des ASV (ou OTU) par les abondances absolues (Dannemiller et al., 2014 ; Props et al., 2017 ; Venturelli et al., 2018) ou soit en ré-échantillonnant conjointement les deux (Vandeputte et al., 2017).

Méthodes non corrélatives

L'inférence de réseaux microbiens n'est pas limitée aux approches corrélatives entre comptages de séquences et d'autres approches proposent d'identifier autrement les interactions entre ASV (ou OTU). Une approche de machine-learning basée sur la logique considère les données d'entrée comme une série d'exemples, ou de programmes logiques, qui doivent être prouvés en utilisant les connaissances préalables (Bohan et al., 2011 ; Ma et al., 2019 ; Tamaddoni-Nezhad et al., 2013). En introduisant les traits nécessaires à la prédation comme connaissances préalables, l'approche a pu identifier des liens trophiques connus et inconnus à partir des abondances (Bohan et al., 2011 ; Davey et al., 2013), mais elle reste à être utilisée sur des comptages de séquences. De plus, les connaissances préalables des traits d'interactions dans le cas des interactions microbiennes peuvent faire défaut. Ainsi le *meta-interpretive learning* (MIL), propose de déduire les règles d'interactions à partir des données d'entrée et des connaissances préalables (s'il y en a) et éventuellement en inventant de nouvelles règles (Tamaddoni-Nezhad et al., 2015). Harris (2016) suggère d'utiliser des réseaux de Markov construits à partir de données de co-occurrence, présence-absence, afin d'estimer efficacement les interactions entre quelques espèces sous la forme de probabilités de co-présence. Cependant, cette approche reste à être évaluée dans le contexte de haute dimension des données de séquençage puisque cette inférence de réseaux repose sur de nombreux échantillons et se limite à 25 espèces au maximum (Harris, 2016). Certaines méthodes issues de domaines éloignés de l'écologie sont également utilisées pour tenter d'identifier des interactions. Ainsi, des auteurs ont tenté de reconstruire des réseaux proie-prédateur via l'approche des k plus proches voisins, mais en se basant sur des données de présence-absence et les traits des prédateurs, malheureusement avec un succès minime (Desjardins-Proulx et al., 2017). Verny et al. (2017) proposent de reconstruire des réseaux basés sur la théorie de l'information en ôtant sélectivement au regard de l'information que le lien apporte. Cette approche a été appliquée à plusieurs types de données biologiques binaires pour générer des réseaux de causalité, parfois même dirigés, mais cette approche n'a pas encore été appliquée à des données de séquençage.

2.2.4 *Surmonter les barrières*

La plupart des méthodes présentées ci-dessus peuvent gérer la compositionnalité des données de séquençage, souvent par la transformation des comptages – SparCC, SpiecEasi, MPLasso, gCoda, CCLasso, REBACCA ou FlashWeave – ou en incluant l'effort de séquençage en tant qu'offset dans le modèle – HMSC, MAGMA ou PLN. Une comparaison récente de méthodes d'inférence de réseaux a néanmoins révélé que les méthodes robustes à la compositionalité n'étaient pas plus performantes que les corrélations de Pearson ou Spearman sur des simulations de communautés (Hirano et Takemoto, 2019). Les auteurs avancent que ces résultats contradictoires à la théorie (e.g., Ban et al., 2015 ; Fang et al., 2015 ; Kurtz et al., 2015) pourraient être dus à la nature des simulations (Hirano et Takemoto, 2019).

Surmonter les associations induites par les réponses conjointes ou disjointes des micro-organismes à leurs habitats peut se faire en introduisant des variables environnementales qui décrivent ces habitats : les covariables (Vacher et al., 2016b). Ne pas inclure les covariables lorsqu'elles ont un effet sur les comptages entraîne des inférences de réseaux inadéquates par rapport aux simulations (Biswas et al., 2016 ; Chiquet et al., 2019 ; Cougoul et al., 2019b ; Tackmann et al., 2019). Plusieurs méthodes d'inférences de réseaux –CoNet, MInt, HMSC, FlashWeave, MAGMA et PLN– sont capables d'inclure de telles covariables. CoNet ou FlashWeave les incluent explicitement comme des nœuds du réseau tandis que MInt, PLN et MAGMA prennent en compte leurs effets au moment de générer les comptages depuis les variables latentes en modifiant la moyenne de la distribution de Poisson (MInt et PLN) ou Négative Binomiale (MAGMA). Les associations indirectes liées à l'interdépendance entre les ASV (ou OTU) inclus durant l'inférence peuvent être retirées avec les approches de corrélations partielles (Li et al., 2016).

2.3 Applications potentielles

L'évaluation et la surveillance des écosystèmes sont classiquement basées sur des espèces indicatrices dont la présence et l'abondance reflètent le fonctionnement de l'écosystème (Baird et Hajibabaei, 2012 ; Cordier et al., 2018b). Les relevés des micro-organismes à partir de données de métabarcoding sont désormais effectués à l'échelle planétaire (Thompson et al., 2017) et permettent de considérer des approches de bio-surveillance à plus grande échelle et automatisées (Bohan et al., 2017 ; Cordier et al., 2018b). Certains auteurs ont donc proposé d'inclure les réseaux microbiens reconstruits à partir de ces données, notamment leurs propriétés (Delmas et al., 2019), comme des indicateurs du fonctionnement de l'écosystème (Bohan et al., 2017 ; Karimi et al., 2017). Ces approches de bio-surveillance ne peuvent s'imaginer que dans un contexte de partage de données et de protocoles établis (voir Annexe 1, p. 167 ; Derocles et al., 2018) et leur succès doit reposer sur des approches de métabarcoding fiables et des méthodes de reconstruction de réseaux robustes.

Une fois reconstruits, les réseaux microbiens indiquent des hypothèses d'interactions entre ASV ou (OTU) qu'il reste à vérifier expérimentalement (Carr et al., 2019 ; Röttjers et Faust, 2018). Les expériences de co-cultures de micro-organismes confirment qu'une partie des liens reconstruits reflètent des interactions biotiques (Das et al., 2018 ; Tipton et al., 2018 ; Wang et al., 2017). Par exemple, une association positive dans un réseau de co-occurrence prédite comme une symbiose potentielle entre eucaryotes microscopiques a été confirmé ensuite par microscopie (Lima-Mendez et al., 2015). De même, le réseau de corrélations obtenu entre

différents règnes microbiens colonisant les racines – bactéries, champignons et oomycètes– a permis de guider des expériences futures et de mettre en évidence un antagonisme inter-règne, des bactéries vers les champignons et oomycètes, nécessaire à la survie de la plante hôte (Durán et al., 2018). Ainsi, la capacité potentielle des réseaux microbiens reconstruits à détecter des interactions biotiques pourrait être mise à profit pour identifier des agents de biocontrôle, c'est-à-dire des micro-organismes capables d'antagonisme envers des pathogènes de plantes (Poudel et al., 2016). En 2019, la France autorise l'utilisation de moins d'une centaine de produits de biocontrôle basés sur des micro-organismes (Anonyme, 2019a). L'identification de nouveaux candidats potentiels est fastidieuse et obtenir des autorisations de mise sur le marché peut être long. Les approches basées sur les réseaux microbiens pourraient permettre, à terme, d'obtenir une liste réduite d'agents de lutte biologique potentiels afin que la mise en évidence de leurs propriétés et les tests *in planta* puissent être au cœur des recherches et non leur identification.

Si les réseaux microbiens peuvent être effectivement reconstruits de manière exhaustive et robuste, les micro-organismes clés de voûte, avec un effet structurant fort sur la communauté, pourraient être identifiés sur la base de l'analyse des réseaux (Agler et al., 2016 ; Banerjee et al., 2018). Ces micro-organismes pourraient ensuite être inoculés pour modifier les communautés microbiennes associées aux plantes cultivées et favoriser des communautés qui limitent les maladies de l'hôte ou stimulent sa croissance (Toju et al., 2018).

3 Objectifs de la thèse

L'objectif de la thèse est d'identifier les approches bioinformatiques et statistiques les plus pertinentes pour reconstruire les réseaux microbiens associés aux plantes à partir de données de métabarcoding, afin de pouvoir utiliser ces réseaux dans les domaines du biocontrôle et de la bio-surveillance. Pour atteindre cet objectif, trois tâches ont été identifiées. Elles constituent la division des travaux présentés dans ce manuscrit. Premièrement, il est nécessaire de caractériser le plus précisément possible la communauté microbienne qui colonise la plante hôte. J'ai donc comparé la capacité de plusieurs centaines de pipelines bioinformatiques à identifier les espèces fongiques associées aux plantes à partir de données de métabarcoding (Chapitre II). Deuxièmement, il est nécessaire d'évaluer la réplicabilité des réseaux inférés et leur aptitude à détecter des changements environnementaux. J'ai donc reconstruit, en utilisant une méthode classique d'inférence (Friedman et Alm, 2012), des réseaux microbiens répliqués et j'ai évalué leur réponse à un changement de pratique culturale (Chapitre III). Enfin, il est nécessaire d'évaluer l'aptitude des réseaux microbiens reconstruits à générer des hypothèses plausibles d'interactions entre micro-organismes, et plus particulièrement entre les pathogènes et le reste de la communauté microbienne. En utilisant une méthode récente d'inférence de réseaux (Chiquet et al., 2017 ; Chiquet et al., 2018), j'ai donc généré des hypothèses d'interactions entre micro-organismes et je les ai confrontées à des expériences de co-cultures (Chapitre IV). Le modèle biologique au centre de la thèse est le pathosystème vigne (*Vitis vinifera*) - oïdium de la vigne (*Erysiphe necator*), mais les approches bioinformatiques et statistiques évaluées sont génériques et peuvent être appliquées à d'autres écosystèmes microbiens.

4 Modèle d'étude : la vigne

Le modèle d'étude au coeur de cette thèse est la vigne, *Vitis vinifera*, une plante grimpante pérenne du genre *Vitis*. La vigne et sa culture sont mentionnées depuis les premiers écrits et alimentent les mythologies (Ayala, 2011). Des régions viticoles telles que Saint-Émilion, sont même désormais inscrites au patrimoine culturel mondial reconnu par l'UNESCO [1]. La domestication de la vigne pour la production de vin est estimée entre 5000 et 9000 ans avant notre ère (Ayala, 2011) et des milliers de variétés appelées cépages, existent désormais (Myles et al., 2011). La vigne est aujourd'hui cultivée mondialement pour la consommation de ses fruits mais surtout pour leur transformation en vin suite à la fermentation microbienne des baies. La viticulture occupait une surface de 792 533 ha en 2018, soit 79% des cultures permanentes en France métropolitaine (Anonyme, 2019b, p. 40).

Tous les organes de la vigne sont peuplés de micro-organismes : les feuilles, la surface des baies de raisin, les bourgeons, les fleurs, les graines, les rameaux, l'écorce, le bois et les racines (par exemple Compant et al., 2011 ; Cortesi et al., 2008 ; Dissanayake et al., 2018 ; Martins et al., 2013 ; Morrison-Whittle et al., 2017 ; Zarraonaindia et al., 2015). Ces communautés fongiques et bactériennes sont détectées tout au long de la saison végétative (Fort et al., 2016 ; Pinto et al., 2014). Certains micro-organismes persistent sur cette plante pérenne jusqu'à la saison suivante dans les baies non vendangées qui se momifient sur les ceps (Sipiczki, 2016), dans les bourgeons et également sur l'écorce et les feuilles mortes (Cortesi et al., 2008). Dans le cas de la vigne, l'importance des micro-organismes associés va au-delà de la plante et de ses fruits puisque la vaste majorité des raisins collectés en France sont destinés à la production viticole (Anonyme, 2019b, p. 39). Dans ce contexte, la nature des micro-organismes détectés sur les baies est importante. Les pathogènes peuvent endommager les baies et provoquer des pertes de récolte et une altération du goût des vins (Kassemeyer, 2017). Des levures non pathogènes dites "indigènes" sont également présentes à la surface des baies (Sun et al., 2009 ; Zott et al., 2010). Celles-ci peuvent se retrouver dans les moûts et participer à la fermentation spontanée (Sun et al., 2009) si elles résistent aux conditions abiotiques : produits phytosanitaires appliqués à la vigne (Kosel et al., 2019), les fortes concentrations en sucres, puis en alcool suite à la fermentation et à l'ajout de sulfites (Dupont et al., 2017 ; Grangeteau et al., 2017). Les communautés fongiques du moût sont initialement plus similaires aux communautés de la surface des baies, mais une fois la fermentation effectuée, elles ressemblent aux communautés de l'écorce (Morrison-Whittle et Goddard, 2018). Il reste à démontrer si les espèces de ces communautés sont ubiquistes ou si elles sont transférées des compartiments jusqu'au produit fini (Morrison-Whittle et Goddard, 2018).

Les micro-organismes les mieux caractérisés parmi les colonisateurs de la vigne sont ceux qui causent des maladies (Armijo et al., 2016). La vigne, comme beaucoup d'autres plantes cultivées, est en effet soumise à des pressions importantes de maladies parasitaires liées à des insectes ou des micro-organismes (Blancard, 2019). Trois maladies majeures de la vigne sont causées par des organismes fongiques et oomycètes : les champignons *Botrytis cinerea* (pourriture grise) et *Erysiphe necator* (oïdium) et l'oomycète *Plasmopara viticola* (mildiou) (Armijo et al., 2016). Ces maladies affectent principalement les feuilles puis les baies causant leur éclatement et une

[1] https://whc.unesco.org/fr/list/932

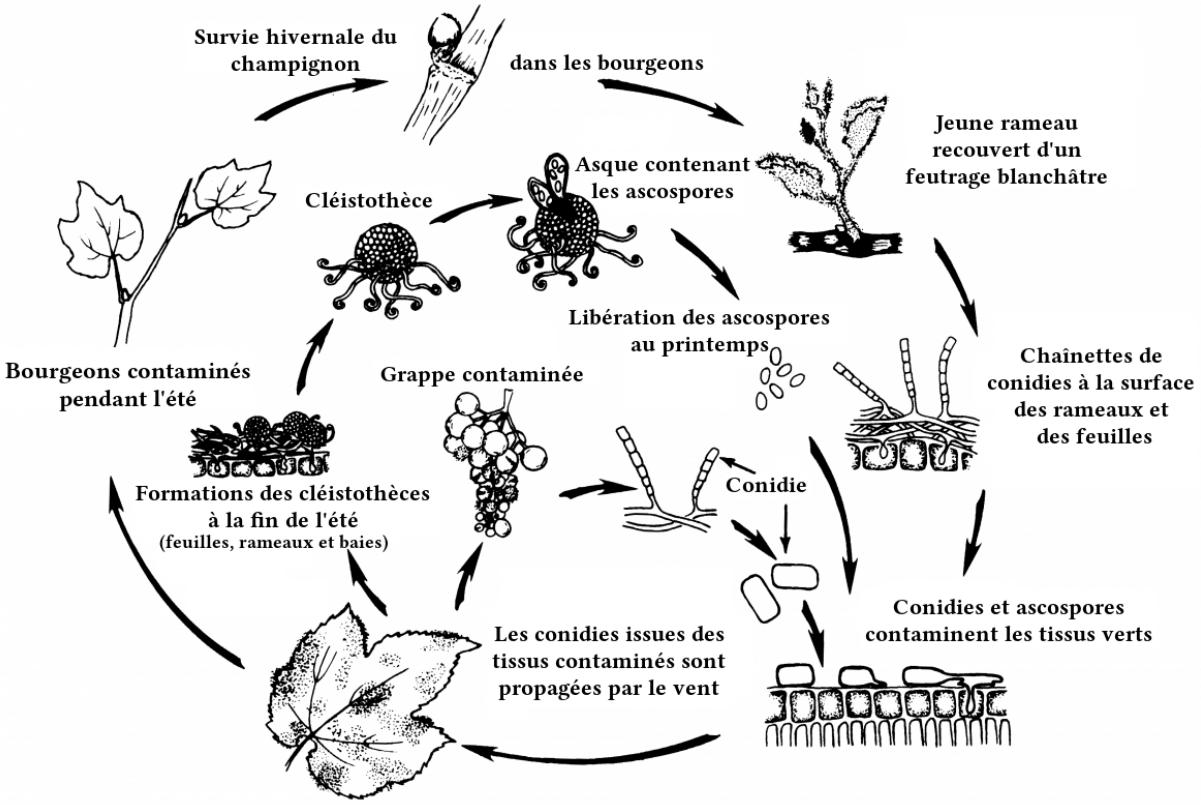


Figure 8 – Cycle biologique de l'agent causal de l'oïdium de la vigne (*Erysiphe necator*). À partir d'une feuille contaminée (en bas à gauche), le cycle de reproduction asexuée via la survie dans les bourgeons mène à la formation de conidies (cercle extérieur). Autrement, le pathogène peut réaliser une reproduction sexuée et avoir hiverné sous forme de cléistothèces menant à la formation d'ascospores (cercle intermédiaire). Ces deux sources d'inoculum primaire induisent une contamination des tissus verts (en bas à droite) et le développement des symptômes. Les conidies sur ces tissus servent de source d'inoculum secondaire tout le long de la saison végétative (cercle interieur) via leur dispersion aérienne favorisant la propagation de la maladie. Source : Pearson et Goheen (1988) et adaptée d'après les sites https://ohioline.osu.edu et https://www.syngenta.fr.

sensibilité accrue à d'autres maladies (Armijo et al., 2016). De plus, la flavescence dorée est une maladie fatale pour la vigne due à une infection incurable causée par une bactérie sans paroi, un phytoplasme, qui est transportée par un insecte vecteur du nom de cicadelle. La déclaration de cette maladie est obligatoire afin de mitiger sa propagation sur le territoire et seul des traitements contre l'insecte vecteur sont disponibles [2]. De façon similaire, la maladie de Pierce causée par la bactérie *Xylella fastidiosa* est devenue une priorité épidémiologique au vu de sa récente détection dans le sud de la France et en Corse et des dégâts qu'elle a déjà causés aux vignobles californiens (Almeida et al., 2019). La vaste majorité des traitements phytosanitaires en viticulture française concerne la lutte contre le mildiou et l'oïdium (Simonovici, 2019, p. 10).

Au cours de cette thèse, je me suis plus particulièrement focalisé sur l'agent causal de l'oïdium de la vigne (*Erysiphe necator*) et les communautés associées à ce champignon. Nous avons mis l'accent sur les communautés fongiques afin de pouvoir caractériser conjointement le pathogène et la communauté associée à l'aide du même marqueur moléculaire lors de

[2]http://ephytia.inra.fr/fr/C/6070/Vigne-Phytoplasme-de-la-flavescence-doree

l'approche de métabarcoding. *Erysiphe necator* est un champignon biotrophe obligatoire, c'est-à-dire qu'il ne peut survivre sans un hôte vivant duquel il tire ses nutriments (Armijo et al., 2016 ; Gadoury et al., 2012 ; Kemen et Jones, 2012). À la suite d'une infection de la plante hôte durant la saison végétative, le pathogène survit à l'hiver sous deux formes possibles : de mycélium dans les bourgeons ou sous forme de cléistothèces dans l'écorce, les rameaux et les feuilles au sol (Figure 8 ; Blancard, 2019). Le mycélium d'*E. necator* (à droite au milieu de la Figure 8) se forme à partir de la forme de reproduction asexuée du champignon : les conidies. Ce mycélium s'étend sur la surface des tissus végétaux puis développe une structure dénommée appressorium permettant l'entrée mécanique dans les cellules de l'hôte en quelques heures (p. 66 Corio-Costet, 2007). Une fois la cuticule perforée et l'appressorium dans les cellules, une autre structure spécialisée se développe, l'haustorium, qui permet à *E. necator* d'absorber les nutriments tels que sucres, acides aminés et vitamines (Armijo et al., 2016 ; Blancard, 2019). Une fois sa nutrition assurée, le pathogène développe d'autres structures telles que les conidiophores qui portent les conidies (Figure 8) et dont la densité importante forme une poudre grise visible sur les feuilles fortements contaminées (Corio-Costet, 2007 ; Gadoury et al., 2012). Suite à l'infection, les feuilles peuvent se déformer et même tomber, et les conidiophores sur les feuilles servent de source d'inoculum secondaire pour les autres organes de la vigne (Figure 8), comme les rameaux ou bien les baies qui seront plus acides et moins sucrées suite à l'infection (Calonnec et al., 2004 ; Corio-Costet, 2007 ; Gadoury et al., 2012). La reproduction sexuée est possible chez *E. necator* via les cléistothèces (au centre de la Figure 8) qui comportent plusieurs asques (de 4 à 6). Ces structures contiennent les ascospores (de 2 à 8) qui une fois libérées permettent de contaminer les tissus verts de la vigne (Figure 8 ; Corio-Costet, 2007). Les cléistothèces sont de plus des structures très résistantes qui permettent à *E. necator* d'hiverner dans les rameaux, le sol ou l'écorce afin de coloniser la vigne à la saison suivante (Blancard, 2019 ; Corio-Costet, 2007).

L'aire d'origine de ce pathogène est les États-Unis, et son introduction en France en 1848 a causé d'importants dégâts dans les vignobles. Cette crise sanitaire a causé une chute massive de la production viticole qui n'a pu être rétablie à un niveau pré-infestation que dix années après, grâce à l'introduction du soufre comme moyen de lutte (Corio-Costet, 2007, p. 57). Aujourd'hui encore, le soufre est un moyen de lutte largement utilisé sous différentes formulations puisqu'il constitue avec le cuivre, les rares produits phytosanitaires autorisés par les directives européennes pour la viticulture biologique (European Commision, 2007). La viticulture conventionnelle peut utiliser ces mêmes produits et peut les compléter avec un éventail plus large de produits phytosanitaires (European Commision, 2009). Ces fongicides possèdent des mécanismes plus ciblés, tels que l'inhibition de la synthèse des stérols, mais *Erysiphe necator* développe des résistances à l'encontre de ces stratégies (Corio-Costet, 2007 ; Dufour et al., 2011). D'autres approches ont donc été développées sur la base de la résistance de certains cépages à l'oïdium (Armijo et al., 2016 ; Corio-Costet, 2007 ; Doster et Schnathorst, 1985) ou sur la lutte biologique qui vise à contrôler les niveaux d'abondance du pathogène grâce à d'autres organismes (Compant et Mathieu, 2016 ; Corio-Costet, 2007). Par exemple, le champignon *Ampelomyces quisqualis* est capable de parasiter *E. necator* et de limiter sa croissance, une propriété qui a rendu possible sa commercialisation (Angeli et al., 2009 ; Kiss, 1998 ; Németh et al., 2019). D'autres micro-organismes ont été mis en évidence comme des potentiels agents de lutte contre ce pathogène : des bactéries du genre *Bacillus* (*B. subtilis* ou *B. pumilus*[3]), des champignons (*Penicillium chrysogenum*, *Verticillium lecani* ou *Trichoderma afroharzianum*), et des acariens se nourrissant de mycélium (Corio-Costet, 2007 ; English-Loeb et al., 2007 ;

[3]https://ephy.anses.fr/ppp/sonata

Sawant et al., 2017). L'usage des produits de biocontrôle est modeste et représente en moyenne 14% de l'indice de fréquence de traitement (IFT), c'est-à-dire le nombre de doses de référence utilisées par hectare au cours d'une campagne culturale (Simonovici, 2019, p. 9). Mais cet usage varie en fonction de la région viticole (Simonovici, 2019, p. 13).

Les viticultures conventionnelle et biologique visent à produire des vins de qualité, et la comparaison de la qualité des baies collectées entre ces pratiques culturales n'indique pas de différence majeure (Döring et al., 2019 ; Provost et Pedneault, 2016). La pratique culturale influence néanmoins la vigueur et le rendement des plantes, qui sont souvent diminuées en viticulture biologique par rapport à la viticulture conventionnelle (Döring et al., 2019). Plusieurs méta-analyses indiquent que la pratique culturale a des effets sur la biodiversité. Celle-ci est en effet moins altérée par l'agriculture biologique que par l'agriculture conventionnelle (Muneret et al., 2018 ; Seufert et Ramankutty, 2017 ; Tuck et al., 2014). La pratique culturale doit donc être prise en compte dans l'étude des pathobiomes des plantes cultivées, puisque la biodiversité microbienne (Campisano et al., 2014 ; Kernaghan et al., 2017 ; Pancher et al., 2012) et les pathogènes peuvent être influencés (Linder et al., 2006).

II

L'exactitude des données sur les communautés fongiques des plantes et du sol dépend fortement du pipeline de métabarcoding

1 Abstract

Bioinformatics matters : The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline

Charlie Pauvert[1], Marc Buée[2], Valérie Laval[3], Véronique Edel-Hermann[4], Laure Fauchery[2], Angélique Gautier[3], Isabelle Lesur[1,5], Jessica Vallance[6], Corinne Vacher[1*]

1– BIOGECO, INRA, Univ. Bordeaux, 33615 Pessac, France. 2– INRA, UMR 1136 Interactions Arbres/Micro-Organismes, F-54280, Champenoux, France. 3– BIOGER, INRA, 78850, Thiverval Grignon, France 4– Agroécologie, AgroSup Dijon, CNRS, INRA, Univ. Bourgogne Franche-Comté, 21000, Dijon, France 5– HelixVenture, 33700, Mérignac, France 6– SAVE, Bordeaux Sciences Agro, INRA, ISVV, Univ. Bordeaux, 33882 Villenave d'Ornon, France.

* Correspondence : Dr. Corinne Vacher (corinne.vacher@inra.fr)

Fungal communities associated with plants and soil influence plant fitness and ecosystem functioning. They are frequently studied by metabarcoding approaches targeting the ribosomal internal transcribed spacer (ITS), but there is no consensus concerning the most appropriate bioinformatic approach for the analysis of these data. We sequenced an artificial fungal community composed of 189 strains covering a wide range of Ascomycota and Basidiomycota, to compare the performance of 360 software and parameter combinations. The most sensitive approaches, based on the USEARCH and VSEARCH clustering algorithms, detected almost all fungal strains but greatly overestimated the total number of strains. By contrast, approaches using DADA2 to detect amplicon sequence variants were the most effective for recovering the richness and composition of the fungal community. Our results suggest that analyzing single forward (R1) sequences with DADA2 and no filter other than the removal of low-quality and chimeric sequences is a good option for fungal community characterization.

Keywords : Bioinformatics, Environmental DNA, Fungi, Illumina MiSeq, Internal transcribed spacer (ITS), Metabarcoding, DADA2, USEARCH, VSEARCH, LULU.

Available at https://doi.org/10.1016/j.funeco.2019.03.005

2 Introduction

Fungal communities associated with soil and plant tissues have a significant impact on plant fitness and ecosystem function (Baldrian, 2017 ; Buée et al., 2009a ; Dighton et al., 2005 ; Rodriguez et al., 2009 ; Vacher et al., 2016a ; Vandenkoornhuyse et al., 2015). Identification of the fungal species present is a prerequisite for understanding these complex communities, but this task is challenging, due to the cryptic nature, microscopic characters and morphological variability of many fungal species (Hibbett et Taylor, 2013 ; Yahr et al., 2016). Sequence-based taxonomic identification of fungal community members, or metabarcoding, has been the standard technique for the last 10 years (Buée et al., 2009b ; Cordier et al., 2012a ; Hibbett et al., 2009 ; Hibbett et Taylor, 2013 ; Hibbett et al., 2016 ; Jumpponen et Jones, 2009 ; Öpik et al., 2009 ; Schmidt et al., 2013) The internal transcribed spacer (ITS) region is now recognized as the universal barcode for fungi (Schoch et al., 2012), and is conventionally used to sequence fungal communities (Bálint et al., 2014 ; Lindahl et al., 2013), in combination with large taxonomic reference databases, such as UNITE (Abarenkov et al., 2010 ; Kõljalg et al., 2013).

Despite their widespread use, metabarcoding approaches suffer from various biases due to the sampling process, molecular biology steps and bioinformatic analyses used (Bálint et al., 2016 ; Lindahl et al., 2013 ; Palmer et al., 2018 ; Schmidt et al., 2013 ; Sommeria-Klein et al., 2016). These biases can prevent accurate recovery of the fungal community. For instance, the fungi identified may differ according to the barcode region chosen and the primers used for its amplification (Tedersoo et al., 2015), the sequencing platform (Motooka et al., 2017), the method used to assemble reads (Nguyen et al., 2015), the sequence clustering method (Cline et al., 2017 ; Halwachs et al., 2017) and the filters subsequently applied to the operational taxonomic unit (OTU) table (Bokulich et al., 2013 ; Brown et al., 2015). Fungal ecologists thus face difficult decisions at every stage in metabarcoding studies (Alberdi et al., 2018).

Many pipelines have been developed that can be used for processing fungal ITS sequence data. These pipelines include MOTHUR (Schloss et al., 2009), QIIME (Caporaso et al., 2010), SCATA (Durling et al., 2011), CLOVR-ITS (White et al., 2013), VSEARCH (Rognes et al., 2016), FROGS (Escudié et al., 2018), PIPITS (Gweon et al., 2015) and DADA2 (Callahan et al., 2016). However, the variable length of fungal ITS sequences between taxa and high levels of sequence variability render analysis and interpretation of the data particularly difficult (Halwachs et al., 2017 ; Palmer et al., 2018 ; Tedersoo et al., 2015). Fortunately, comprehensive guidelines have been developed, to help fungal community ecologists to make the most appropriate choices (Bálint et al., 2016 ; Bálint et al., 2014 ; Lindahl et al., 2013). These guidelines suggest, for example, that sequence clustering yields the best results with ITS extraction tools such as ITSx (Bálint et al., 2014 ; Bengtsson-Palme et al., 2013 ; Lindahl et al., 2013). The removal of rare OTUs, which may be artifacts, is also generally recommended (Bálint et al., 2016), but there is no consensus concerning the threshold number of sequences below which an OTU can be considered rare. The proposed thresholds range from 1 to 10 sequences (Brown et al., 2015) or depend on the relative abundance of OTUs (Bokulich et al., 2013). Fungal mock communities have also recently been used for the development of guidelines (Bakker, 2018 ; Cline et al., 2017 ; Nguyen et al., 2015). Nguyen et al. (2015) showed, for example, that single forward reads could be used to recover all of the 25 well-amplified species of their mock

community, whereas only 23 of these species were recovered with assembled paired-end reads. The clustering algorithm of USEARCH (Edgar, 2010) has also been recommended, based on the demonstration that it recovered the expected number of mock species (Cline et al., 2017).

There is currently no clear consensus in the scientific community concerning the most appropriate bioinformatic approach for analysis of the fungal ITS regions sequenced on Illumina MiSeq platforms. We aimed to fill this gap, by creating and sequencing a mock community of 189 Dikarya strains commonly found in agricultural and forest soils and in plant tissues. As advised by (Nguyen et al., 2015), this mock community had a large taxonomic breadth and some genera were represented by several closely related strains (Figure 1). We compared the ability of 360 combinations of bioinformatic softwares and parameters to recover the fungal strains present in this mock community, in the expected proportions. In particular, we investigated whether clustering-free software packages that identify exact sequence variants of amplicons (ASVs) rather than clustering similar sequences into OTUs (Callahan et al., 2016) outperformed conventional clustering approaches, by fully exploiting molecular barcode resolution (Callahan et al., 2017).

We also tested novel post-clustering curation tools (Frøslev et al., 2017). We provide new guidelines, based on our results, for researchers using metabarcoding approaches for the analysis of fungal community richness and composition.

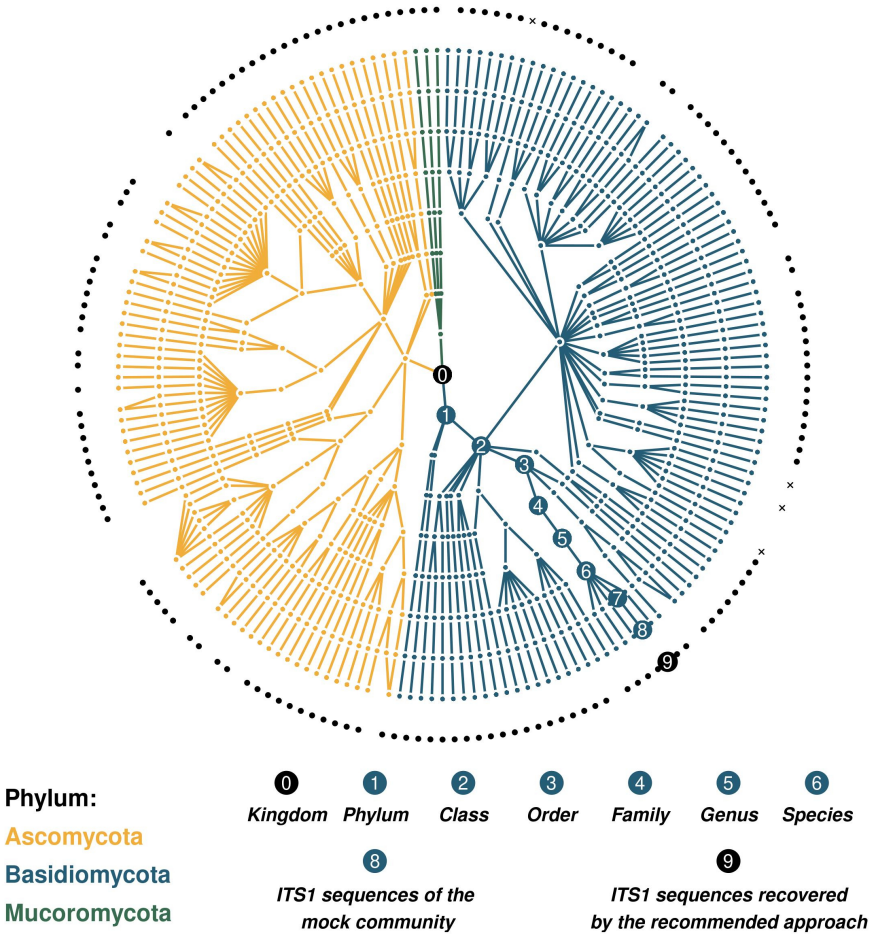


FIGURE 1 – Taxonomic composition of the artificial fungal community. The terminal nodes of the tree are the ITS1 sequences (n = 175) of the fungal strains (n = 189) that constitute the mock community. The ITS1 sequences that were not present in the raw Illumina dataset are indicated with a cross and those that were recovered by the recommended bioinformatic approach (Si5, Table 3) are indicated with a black circle.

3 Materials and methods

3.1 Fungal mock community

The mock community consisted of an equimolar mixture of DNA extracted from 189 pure fungal strains isolated from soils, sporocarps or plant tissues. All the strains belonged to the superkingdom Holomycota (Tedersoo et al., 2018) : 87 Ascomycota strains, 99 Basidiomycota strains and 3 Mucoromycota strains, corresponding to 181 different species, 97 genera, 67 families, 30 orders and 11 classes. Altogether, 30 genera were represented by several species and 4 species were represented by several strains (Figure 1 and Table S4).

Fungal DNA was obtained from the inner flesh of sporocarps, or from aerial mycelium scraped aseptically from the surface of pure cultures grown on PDA (Potato Dextrose Agar), MA (Malt Agar) or Pachlewski's medium (Martin et al., 1983). The mycelium was lyophilized for 24 h in an Edwards Modulyo 4K lyophilizer (Edwards, United Kingdom) and 100 mg of lyophilized mycelium was then placed in a Fast-Prep tube (2 mL) containing 130 mg glass beads (4.5 mm in diameter ; Dutscher, France) and ground with a FastPrep®machine (MP Biomedicals, France) for 30 s at maximum shaking frequency. DNA was extracted with a DNeasy Plant Minikit (Qiagen, France), in accordance with the manufacturer's instructions, except that the incubation time was extended to 1 h at 65°C, and the volumes of buffers AP1 and P3 were doubled. DNA from all strains was quantified with a Qubit®2.0 Fluorometer (Life Technologies, USA) and pooled in an equimolar mixture. Pooling was performed three times (replicates A, B and C). The fungal ITS1 region was amplified from each replicate with the ITS1F (5'-CTTGGTCATTTAGAGGAAGTAA-3', Gardes et Bruns, 1993) and ITS2 (5'-GCTGCGTTCTTCATCGATGC-3', White et al., 1990) primers. These primers are considered as universal fungal primers and are commonly used in fungal community analyses (Buée et al., 2009b ; Cordier et al., 2012a ; Nguyen et al., 2015 ; Palmer et al., 2018). They are known to amplify Ascomycota, Basidiomycota and Mucoromycota (Bellemain et al., 2010 ; Schoch et al., 2012). PCR was performed with a GeneAmp PCR System 2700 (Applied Biosystems, USA). The reaction mixture (20 µL final volume) consisted of 1× of PCR buffer, 0.56 mg mL^{-1} of bovine serum albumin (A2153-10G, Sigma, USA), 0.2 mM of each dNTP, 0.2 μM of each primer, 0.05 U μL^{-1} Taq DNA polymerase (D1806, Sigma-Aldrich) and 5 ng of DNA template. The following cycling parameters were then used for amplification : enzyme activation at 94°C for 3 min ; 35 cycles of denaturation at 94°C for 30 s, annealing at 53°C for 30 s, extension at 72°C for 45 s, and a final extension at 72°C for 10 min. The quality of the PCR products was checked by electrophoresis on 2% agarose gels. PCR products were purified (CleanPCR, MokaScience), multiplex identifiers and sequencing adapters were added, and library sequencing on an Illumina MiSeq platform (v3 chemistry, 2×250 bp) and sequence demultiplexing (with exact index search) were performed at the Get-PlaGe sequencing facility (Toulouse, France).

Full-length ITS sequences were also obtained by Sanger sequencing for the 189 fungal strains. PCR was performed with the ITS1F and ITS4 (5'-TCCTCCGCTTATTGATATGC-3'„ White et al., 1990) primers, with the same PCR mixture as described above. The PCR program consisted of an initial denaturation at 95°C for 3 min, followed by 30 cycles of denaturation at 95°C for

30 s, annealing at 55°C for 30 s, and elongation at 72°C for 45 s. Sequencing reactions were performed by Genewiz (Takeley Essex, UK) and sequences from both strands were assembled with MultAlin (Corpet, 1988) and manually curated.

3.2 Bioinformatics approaches

We analyzed the MiSeq sequences with 360 combinations of bioinformatic softwares and parameters (referred to hereafter as bioinformatic approaches). These approaches differed in (1) the paired-end read assembly algorithm used (*Assembly*), (2) the fungal ITS1 extraction method (*Extraction*), (3) the method of sequence variation analysis (*Variation*), (4) the treatment of chimeric sequences (*Chimeras*) and (5) the final filtering of the community (*Filtering*). The various steps of the bioinformatic analyses are described below and in Figure 2.

3.2.1 *Paired-end read assembly algorithm (Assembly)*

Two paired-end read assembly algorithms were compared (Figure 2). Paired-end sequences were joined with the FASTQ-JOIN function of QIIME v1.8.0 (*Assembly*=FASTQ-JOIN_OL) or with PEAR v0.9.10 (*Assembly*=PEAR_OV) (Caporaso et al., 2010 ; Zhang et al., 2014). For each algorithm, three minimum overlapping lengths (OLs) between forward and reverse sequences were tested (50 bp, 100 bp and 150 bp). For the FASTQ-JOIN algorithm, no mismatch was allowed in the overlap region. We also considered the use of single forward (R1) sequences (*Assembly*=QUALITY_R1) (Figure 2).

3.2.2 *Fungal ITS1 extraction (Extraction)*

The ITS1 region was either extracted (*Extraction*=YES) from the high-quality sequences with ITSx v1.0.10 (Bengtsson-Palme et al., 2013), or not extracted (*Extraction*=NO) (Figure 2). ITSx uses an alignment of conserved ribosomal genes to identify and delineate highly variable regions, such as the ITS1 region, accurately. The minimum length of the region between the binding sites for the ITS1F-ITS2 primers is about 100 bp (Motooka et al., 2017), but the ITS1 region sensu stricto (as defined by ITSx) is shorter, as it does not include portions of the 18S and 5.8S flanking regions. We thus discarded sequences of less than 100 bp in length in cases in which the ITS1 region was not extracted, or 50 bp in cases in which it was extracted.

3.2.3 *Sequence variation analysis (Variation)*

Two clustering algorithms were compared (Figure 2). Fungal ITS1 sequences displaying more than 97% similarity were clustered into OTUs with the popular USEARCH v7.0 program (*Variation*=USEARCH) (Edgar, 2010) or with the open-source alternative VSEARCH v2.5.2 (*Variation*=VSEARCH) (Rognes et al., 2016). A similarity threshold of 97% was chosen as this threshold is commonly used in fungal metabarcoding studies (e.g. Bakker, 2018 ; Durand et al., 2017) and has been shown to perform well on a mock community (Tedersoo et al., 2015), despite its tendency to aggregate closely related species (Bálint et al., 2016 ; Ryberg, 2015). All other settings were left to default. Sequences with a Phred score greater than 30 over 75% of the read length were included in the clustering process. Quality filtering was performed with the QIIME script *split_libraries_fastq.py* (Figure 2).

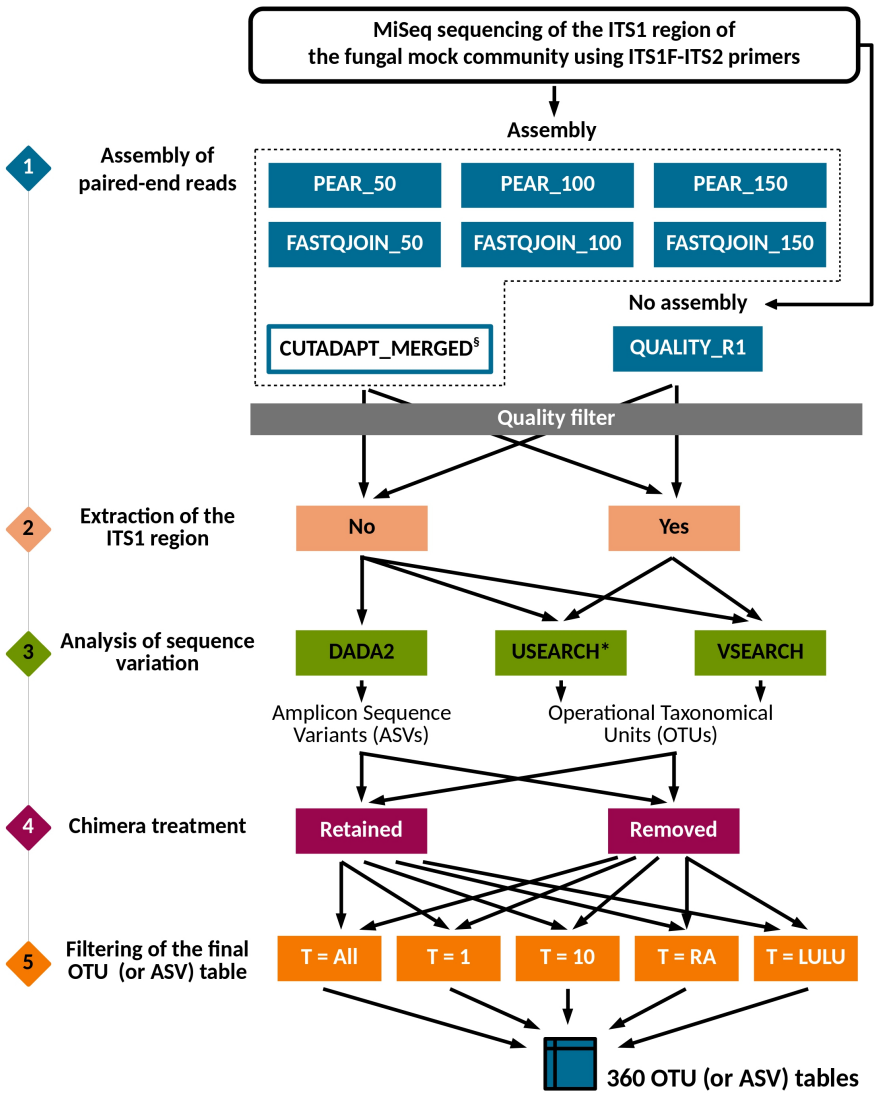


Figure 2 – Overview of the 360 bioinformatic approaches compared in this study. The Illumina MiSeq sequences were (1) assembled with FASTQ-JOIN (Caporaso et al., 2010) or PEAR (Zhang et al., 2014) with three minimum overlapping lengths (50 bp, 100 bp or 150 bp), or not assembled. In this latter case, single forward (R1) reads were used. After quality filtering, (2) the ITS1 region was extracted from the reads with ITSx (Bengtsson-Palme et al., 2013), or not extracted. (3) Sequence variations were then analyzed with DADA2 (Callahan et al., 2016), USEARCH (Edgar, 2010) or VSEARCH (Rognes et al., 2016) and (4) chimeras were either retained or removed. (* : for USEARCH, chimera detection was performed before clustering). Finally, (5) the datasets were either filtered by removing rare or erroneous OTUs (or ASVs), or left unfiltered. Filtering thresholds (T) were based on the number of sequences per OTU, or on their relative abundance (RA), or OTU curation was performed using the LULU algorithm (Frøslev et al., 2017). § : When DADA2 was used, an alternative method of read processing (CUTADAPT_MERGED) was included.

We also used the R package DADA2 (Callahan et al., 2016) to correct sequencing errors and to infer exact amplicon sequence variants (*Variation*=DADA2) (Figure 2). We retained only reads with less than one expected error (given the quality scores ; Edgar et Flyvbjerg, 2015). Quality filtering was performed with the *fastqFilter* function. Quality data were lost during the extraction step. We therefore applied DADA2 only to the fungal ITS1 sequences not extracted with ITSx. This analysis strategy is different from that recommended in the DADA2 tutorial (http://benjjneb.github.io/dada2/tutorial.html). We, therefore, also included an approach adhering to the strategy described in the tutorial (*Assembly*=CUTADAPT_MERGED) (Figure 2). Primers were removed from both forward and reverse reads, with Cutadapt v1.13 (Martin, 2011). The forward and reverse reads were then truncated (at 200 bp and 180 bp, respectively) and we retained only reads with fewer than two expected errors (as in the default parameters of the *filterAndTrim* function). Reads were merged after the inference of sequence variation as described in the tutorial.

3.2.4 *Treatment of chimeric sequences (Chimeras)*

The chimeric sequences identified were either removed (Chimeras=Removed) or retained in the dataset (*Chimeras*=Retained) (Figure 2). Detection with QIIME script *identify_chimeric_seqs.py* was performed on the demultiplexed reads before USEARCH clustering (Figure 2), as recommended in the QIIME tutorial (http ://qiime.org/scripts/identify_chimeric_seqs.html). Following VSEARCH clustering, we combined *de novo* and reference-based strategies for chimera detection. The *de novo* strategy used the UCHIME (Edgar et al., 2011) algorithm implemented in VSEARCH. The reference-based strategy used the ITS1-only UNITE-UCHIME dataset v7.2 (as of 2017-10-10) as a reference (Nilsson et al., 2015). Following DADA2 sequence variation analysis, chimeric sequences were removed with the *removeBimeraDeNovo* function, using the consensus option.

3.2.5 *Final filtering of the community (Filtering)*

Finally, we filtered the OTU and ASV tables (Figure 2). Five filtering methods were compared : *Filtering*=1 involved removing OTUs (or ASVs) composed of a single sequence ; *Filtering*=10 involved removing OTUs (or ASVs) for which less than 10 sequences were obtained when all three replicates were considered ; *Filtering*=RA involved removing OTUs (or ASVs) with a relative abundance lower than 0.005% of the total number of sequences ; *Filtering*=LULU used the LULU curation algorithm (Frøslev et al., 2017) to collapse erroneous OTUs (or ASVs) into their parent OTUs ; *Filtering*=All involved keeping all OTUs or ASVs, regardless of the number of sequences obtained. Representative sequences were assigned to taxa with the QIIME script assign_taxonomy.py, with BLAST v2.2.22 (Altschul et al., 1990) and default QIIME parameters (e-value < 0.001 ; identity $\geq 90\%$) against the local database of Sanger sequences for the fungal strains of the mock community. The LULU curation algorithm was applied with both default settings and a set of parameters adjusted to the features of the mock community. Three parameters can be tuned in LULU : the minimum sequence similarity between a 'potential daughter' and its 'potential parent' (default 84%), the minimum ratio of parent OTU abundance to daughter OTU abundance in all samples (default 1) and their minimum co-occurrence rate across samples (default 95%). Increasing the first parameter is only advised when the barcode region has little variation or when few PCR and sequencing errors are expected, and changing the second parameter is generally not recommended (Frøslev et al., 2017). Therefore we tuned the third parameter. We lowered its value to 66.6% to account for the small number of samples in our study (3 replicate samples per bioinformatic approach).

3.2.6 *Comparison criteria*

We defined three criteria for comparisons of the ability of the bioinformatic approaches to recover the mock community : sensitivity, precision and compositional similarity. Sensitivity and precision were defined as the true positive rate $TP/(TP + FN)$ and the positive predictive value $TP/(TP + FP)$, respectively, where TP is the number of true-positive OTUs (or ASVs), FN is the number of false-negative OTUs (or ASVs) and FP is the number of false-positive OTUs (or ASVs). True-positive OTUs corresponded to fungal strains present in the mock community and identified by the bioinformatic approach considered. False-negative OTUs corresponded to fungal strains present in the mock community but not detected by the bioinformatic approach considered. False-positive OTUs corresponded to all other OTUs. If several OTUs were assigned to the same fungal strain of the mock community (i.e. 'split' OTUs), only the most abundant was considered to be a true-positive OTU, the others being considered false-positive OTUs. Compositional similarity was defined as the Bray-Curtis similarity (Odum, 1950) between the community recovered and the mock community. It was calculated as $1 - BC$, where BC is the Bray-Curtis dissimilarity obtained from the *vegdist* function of the R *vegan* package (Oksanen et al., 2018), assuming a uniform distribution of sequences between the fungal strains in the mock community. The expected number of sequences per fungal strain in the mock community was calculated for each replicate and each bioinformatic approach as the total number of high-quality sequences (obtained after the *Filtering* step) divided by the total number of fungal strains in the mock community. Ribosomal RNA gene copy number information was not available for each strain and was not used to adjust the expected number of sequences.

All three criteria theoretically range between 0 and 1. They equal 1 when the algorithm successfully identifies all members of the mock community. However, maximum sensitivity may be below 1 if the sequences of some fungal strains are absent from the raw Illumina dataset. We, therefore, estimated the total number of strains present in the raw dataset, by aligning the forward and reverse MiSeq sequences with the ITS1 Sanger sequences, with a similarity threshold of 100% and an alignment length threshold of 90% of the length of the shorter sequence. Alignments were performed with VSEARCH (*--usearch_global*) (Rognes et al., 2016), using the following parameters : *--id 1 –userout --userfields query+target+qcov+tcov+id --maxaccepts 20 –top_hits_only.*

4 Results

4.1 Assessment of maximum sensitivity

The manually curated Sanger database contained the ITS1 sequences of the 189 fungal strains of the mock community (Figure 1 and Table S 4). Several strains had identical Sanger sequences for ITS1 : two strains from the genus *Alternaria*, six from the genus *Botrytis* (*B. calthae*, *B. pseudocinerea*, *B. ranunculi* and three strains of *B. cinerea*), two strains from two different species of *Colletotrichum* (*C. destructivum* and *C. higginsianum*), two strains of *Craterellus cornucopioides*, four pairs of strains from the genus *Fusarium* (*F. acuminatum* and *F. avenaceum*, *F. langsethiae* and *F. sporotrichioides*, *F. oxysporum* and *F. commune*, *F. verticillioides* and another species of the *F. fujikuroi* species complex), two strains from two different species of *Lepista* (*L. irina* and *L. nuda*) and two strains from the genus *Zymoseptoria*. The Sanger database, therefore, contained 175 unique ITS1 sequences.

Only 160 of these 175 unique ITS1 sequences were detected in the raw Illumina MiSeq dataset (Table S 4), suggesting that the other 15 strains were either not amplified by the ITS1F-ITS2 primer pair, not sequenced or were sequenced with errors. Eleven of these strains were detected in the Illumina data when the similarity threshold between Sanger and Illumina sequences was lowered to 93.5% (Table S 4), suggesting that their apparent absence was caused by mismatches between the Sanger sequence and the Illumina sequences. Four of these eleven strains had ambiguous bases in the Sanger sequence, preventing a perfect match with Illumina sequences. These ambiguous bases might be due to the within-strain polymorphism of the ITS region that exists for some fungi (Fiers et al., 2011). The other four ITS1 sequences absent from the raw Illumina dataset came from the following species : *Lepiota clypeolaria*, *Mycena abramsii, M. galopus and Panellus stipticus*. The first species was successfully amplified with the ITS1F-ITS4 primer pair before Sanger sequencing and possessed the exact sequence of the ITS2 primer, suggesting that its absence from the Illumina dataset was caused by DNA pooling biases rather than lack of amplification. In contrast, the sequence of the ITS2 primer was detected with some mismatches for the three last species. Their absence could be due to a lack of amplification by the ITS1F-ITS2 primer pair (Table S 4). The maximum sensitivity attainable by any bioinformatic approach (that is, the maximum proportion of fungal strains that could actually be found) therefore ranged from 84.7% to 90.1%. The lower bound was obtained by considering that the raw Illumina dataset contained the ITS sequences of 160 strains (of 189), while the upper bound also took into account the nine strains sequenced with some errors.

4.2 Influence of Assembly, Extraction, Variation, Chimera and Filtering on sequencing data

4.2.1 *Influence of Assembly*

In total, we obtained 143873 paired-end Illumina reads of 250 bp each (Table S5). We obtained 43352, 56406 and 44115 sequences for the three replicates. The mean quality of the forward (R1)

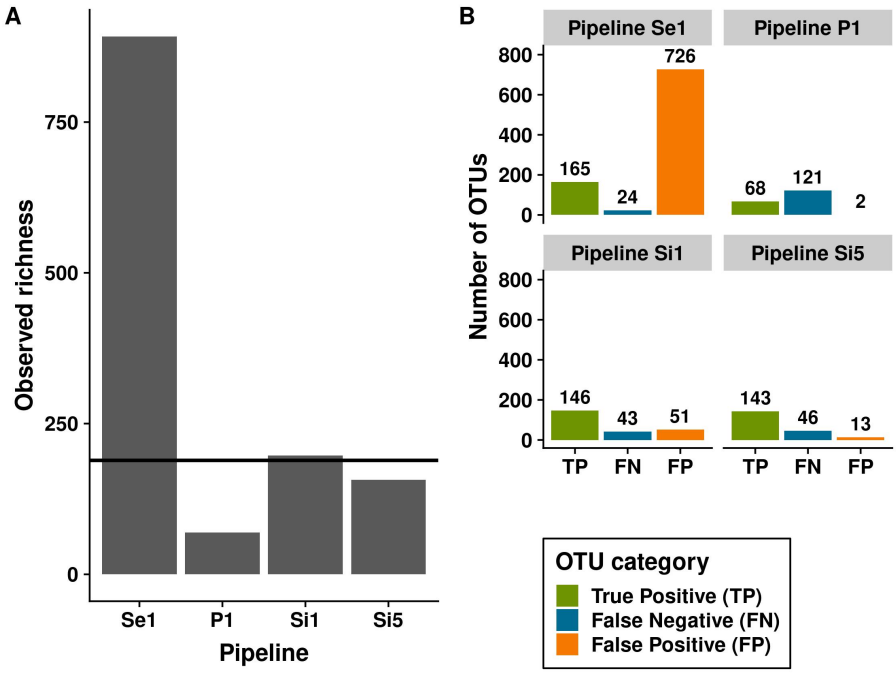


FIGURE 3 – Richness estimates for the top four approaches. (A) Total number of OTUs (or ASVs) retrieved by the most sensitive approach (Se1 ; Table 1), the most precise approach (P1 ; Table 2), the approach with the best performance in terms of compositional similarity to the mock community (Si1 ; Table 3) and the bioinformatic approach recommended in this study (Si5 ; Table 3). The black horizontal line indicates the expected richness. (B) Total number of OTUs (or ASVs) per bioinformatic approach depending on OTU (or ASV) category. TP = true positives, FN = false negatives and FP = false positives. Results were averaged over the three replicates and rounded for clarity.

reads was slightly higher than that of the reverse (R2) reads (35.59 *versus* 34.31, respectively) (Table S5).

The choice of paired-end read assembly algorithm strongly influenced the number, length and quality of the consensus sequences (Table S5 and Figure S1). FASTQ-JOIN retained on average 53.1% of the raw reads (whatever the minimum overlapping length), whereas PEAR retained on average 96.3% of the raw reads (Table S5). Mean sequence quality was also higher for PEAR than for FASTQ-JOIN (Table S5 and Figure S5). More than 90% of assembled reads passed the quality filter, whatever the assembly algorithm used (Table S6). Thus, PEAR generated twice as many high-quality assembled reads as FASTQ-JOIN (Table S6).

4.2.2 *Influence of Extraction*

The extraction of the ITS region with ITSx (BENGTSSON-PALME et al., 2013) retained 97% to 99% of the assembled reads, but only 59% of the forward reads (Table S7). Reads were 118 nucleotides shorter, on average, after extraction. Most reads were between 200 and 300 bp long before extraction (Table S5), versus 100 to 200 bp after extraction (Table S7).

Table 1 – List of the 10 most sensitive approaches. Sensitivity, precision and similarity values were averaged over the three replicates for each bioinformatic approach. Richness is defined as the mean number of OTUs identified by the bioinformatic approach.

Approach	*Assembly*	*Extraction*	*Variation*	*Chimeras*	*Filtering*	*Richness*	*Sensitivity*	*Precision*	*Similarity*
Se1	QUALITY_R1	No	VSEARCH	Retained	All	892	0.875	0.186	0.362
Se2	QUALITY_R1	No	USEARCH	Removed	All	878	0.873	0.188	0.36
Se3	QUALITY_R1	No	USEARCH	Retained	All	949	0.871	0.174	0.3608
Se4	QUALITY_R1	No	VSEARCH	Removed	All	577	0.871	0.286	0.366
Se5	PEAR_50	No	USEARCH	Retained	All	1410	0.869	0.117	0.354
Se6	PEAR_100	No	USEARCH	Retained	All	1413	0.866	0.116	0.355
Se7	PEAR_50	No	VSEARCH	Retained	All	1257	0.866	0.131	0.357
Se8	PEAR_100	No	VSEARCH	Retained	All	1246	0.862	0.131	0.357
Se9	PEAR_50	No	USEARCH	Removed	All	1287	0.861	0.127	0.354
Se10	QUALITY_R1	No	VSEARCH	Retained	1	612	0.861	0.266	0.364

4.2.3 *Influence of Variation and Chimera*

The total number of OTUs (or ASVs) varied by several orders of magnitude, depending on the method used to analyze sequence variation. For example, USEARCH identified 878 non-chimeric OTUs and 71 chimeric OTUs on average with the following parameters, *Assembly*=QUALITY_R1 and *Extraction*=No. VSEARCH identified 577 non-chimeric OTUs and 315 chimeric OTUs with the same parameters. DADA2 identified 157 non-chimeric ASVs and 40 chimeric ASVs. These striking differences were found for all mock replicates (Figure S6).

4.2.4 *Influence of Filtering*

The final filtering step also strongly influenced the number of OTUs. For instance, removing OTUs with less than 10 sequences (*Filtering*=10) reduced the number of non-chimeric OTUs identified by USEARCH from 878 to 329, and the number of non-chimeric OTUs identified by VSEARCH from 577 to 257, for *Assembly*=QUALITY_R1 and *Extraction*=No. The LULU curation algorithm reduced even more the number of non-chimeric OTUs but, in contrast to other filtering methods, it did not lose any sequence (Table S8). ASV tables were more robust than OTU tables to variations in the filtering methods (Table S8).

4.3 Comparison of the bioinformatic approaches on the basis of sensitivity, precision and compositional similarity criteria

Bioinformatic analyses generated 360 matrices containing the number of sequences per OTU (or ASV) for the three replicates (Figure 2). The matrices differed considerably. For example, the mean number of OTUs (or ASVs) per replicate ranged from 57 to 1562, depending on the bioinformatic approach used (Table 9). Sensitivity, precision and compositional similarity values were calculated at replicate level. The ranking of approaches according to these criteria differed between the three mock replicates (Table S9), but the approaches that performed very well according to a given criterion for one replicate generally also performed well for the other two replicates. We, therefore, used the mean value of the criteria over the three replicates to rank the bioinformatic approaches.

The sensitivity of the 360 bioinformatic approaches ranged from 22% to 87% (Table S9). The 10 most sensitive approaches are listed in Table 1. All used USEARCH or VSEARCH to cluster sequences into OTUs and did not extract the fungal ITS1 region with ITSx. All these approaches produced very large numbers of OTUs, up to seven times more than the actual number of fungal strains in the mock community (Table 1 and Figure 3A). They, thus, recovered most of the fungal strains of the mock community but also generated many false-positive OTUs (Figure 3B). The precision of these approaches was, therefore, very low (Figure 4). However, they displayed a high degree of compositional similarity to the mock community despite the large number of false-positive OTUs (Table 1).

Precision (the proportion of OTUs (or ASVs) corresponding to true strains), ranged from 9% to 98% (Table S9). The 10 most precise approaches are listed in Table 2. All used DADA2 to identify amplicon sequence variants (Figure 4), did not extract the ITS region with ITSx and used the LULU curation algorithm. The 8 first approaches used assembled reads as input data (Table 2). Removing chimeras with DADA2 before LULU curation appeared to be unfavourable, as it slightly reduced the sensitivity of all these top-ranking approaches (Table 2). Adjusting the minimum co-occurrence threshold of the LULU algorithm did not influence the results (data not shown). Unlike the most sensitive approaches, the most precise approaches yielded fewer ASVs than there were fungal strains in the mock community (Table 2) and produced very few false-positive ASVs (Figure 3B). However, they did not recover all mock strains. The most precise approach, P1, recovered only 36% of the fungal strains of the mock community (Table 2 and Figure 3B).

Compositional similarity to the mock community ranged from 0.15 to 0.396 (Table S9). The 10 best approaches according to this criterion are listed in Table 3. Like the most precise approaches, these 10 approaches used DADA2 to identify amplicon sequence variants (Figure 2) and did not extract the ITS region with ITSx. However, unlike the most precise approaches, most used non-assembled reads as input data and they did not use the LULU curation algorithm. The approach with the best performance according to the similarity criterion (Si1) used R1 reads as input data, retained chimeras and applied no filters to the final ASV table. This approach recovered 77.4% of the fungal strains from the mock community but had a relatively low precision (Table 3). In contrast, the Si5 approach, which used the same options but with the removal of chimeras, had a precision increased by 17%. This is because chimera removal efficiently discarded false positive ASVs, lowering their number from 51 to 13 (Figure 3B). As a side effect, chimera removal triggered the loss of 3 true positive ASVs, slightly reducing the sensitivity of the Si5 approach (Figure 3B). The removal of primers (as recommended in the DADA2 tutorial) did not

Table 2 – List of the 10 most precise approaches. Sensitivity, precision and similarity values were averaged over the three replicates for each bioinformatic approach. Richness is defined as the mean number of ASVs. LULU was applied with default settings.

Approach	*Assembly*	*Extraction*	*Variation*	*Chimeras*	*Filtering*	*Richness*	*Sensitivity*	*Precision*	*Similarity*
P1	PEAR_150	No	DADA2	Retained	LULU	69	0.358	0.976	0.215
P2	PEAR_150	No	DADA2	Removed	LULU	67	0.347	0.976	0.212
P3	CUTADAPT_MERGED	No	DADA2	Retained	LULU	100	0.515	0.973	0.271
P4	CUTADAPT_MERGED	No	DADA2	Removed	LULU	98	0.504	0.973	0.268
P5	FASTQJOIN_150	No	DADA2	Retained	LULU	61	0.312	0.973	0.187
P6	FASTQJOIN_150	No	DADA2	Removed	LULU	59	0.302	0.972	0.182
P7	PEAR_100	No	DADA2	Retained	LULU	96	0.49	0.969	0.251
P8	PEAR_100	No	DADA2	Removed	LULU	94	0.48	0.968	0.249
P9	QUALITY_R1	No	DADA2	Retained	LULU	107	0.547	0.966	0.278
P10	QUALITY_R1	No	DADA2	Removed	LULU	105	0.536	0.965	0.275

Table 3 – List of the 10 approaches with the best performances in terms of compositional similarity to the mock community. Sensitivity, precision and similarity values were averaged over the three replicates for each bioinformatic approach. Richness is defined as the mean number of ASVs. The bioinformatic approach recommended in this study (Si5) is shown in bold.

Approach	*Assembly*	*Extraction*	*Variation*	*Chimeras*	*Filtering*	*Richness*	*Sensitivity*	*Precision*	*Similarity*
Si1	QUALITY_R1	No	DADA2	Retained	All	197	0.774	0.743	0.396
Si2	QUALITY_R1	No	DADA2	Retained	1	197	0.774	0.743	0.396
Si3	QUALITY_R1	No	DADA2	Retained	RA	192	0.758	0.75	0.396
Si4	QUALITY_R1	No	DADA2	Retained	10	187	0.739	0.751	0.396
Si5	**QUALITY_R1**	**No**	**DADA2**	**Removed**	**All**	**157**	**0.758**	**0.915**	**0.395**
Si6	QUALITY_R1	No	DADA2	Removed	1	157	0.758	0.915	0.395
Si7	QUALITY_R1	No	DADA2	Removed	RA	152	0.743	0.921	0.395
Si8	QUALITY_R1	No	DADA2	Removed	10	148	0.723	0.924	0.394
Si9	PEAR_50	No	DADA2	Retained	All	212	0.765	0.684	0.391
Si10	PEAR_50	No	DADA2	Retained	1	212	0.765	0.684	0.391

improve the performance of these two top-ranking approaches. It slightly lowered the precision of the Si1 approach (Table S7). The compositional similarities of the Si1 and Si5 approaches were 0.396 and 0.393, respectively (Table 3). These values were among the highest obtained, but were far from the maximal value of 1 indicating an exact match between the observed and expected community. This difference resulted from the huge variability in the number of sequences per ASV, contrasting with the expected uniform distribution of reads between fungal strains (Figure S7). The expected number of reads for each fungal strain was then multiplied by the number of fungal strains with an identical ITS1 sequence (Figure S7), which increased compositional similarity values (Table S9) but did not change the ranking of the bioinformatic approaches (Spearman $\rho = 0.99$; p <2.2e-16).

Finally, comparison of the bioinformatic approaches revealed that some steps that are commonly recommended, such as ITS extraction and chimera removal, can have positive effects but also negative ones. For instance, the extraction of the ITS1 region with ITSx before USEARCH and VSEARCH clustering increased significantly precision but it decreased sensitivity (Figure S8), suggesting that ITS extraction discarded some false-positive OTUs but also some true-positive OTUs. Similarly, bioinformatic approaches that kept chimeras after DADA2 sequence correction (Si1-Si4 in Table 3) were slightly more sensitive than approaches that remove chimeras (Si5-Si8 in Table 3), indicating that chimera removal discarded some true-positive OTUs. This negative effect of chimera removal also occurred in the USEARCH and VSEARCH pipelines, but to a lower extent (Figure S6).

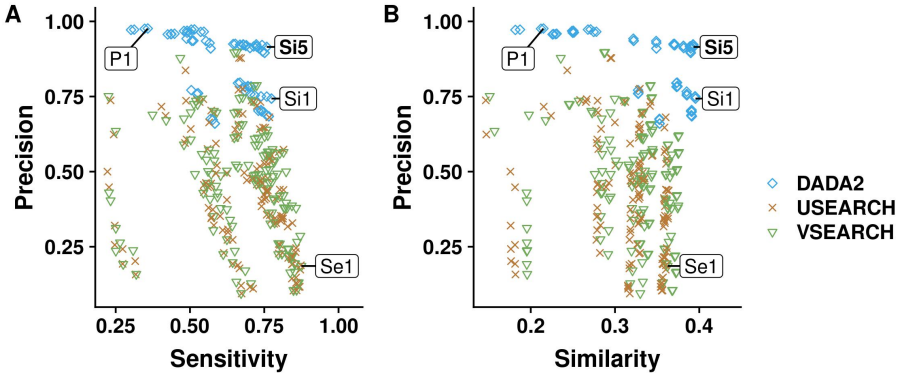


FIGURE 4 – Values of precision and (A) sensitivity or (B) compositional similarity to the mock fungal community, for all 360 bioinformatic approaches. Each dot corresponds to the mean value obtained for an approach over the three replicates. The methods used to analyze sequence variation (DADA2, USEARCH or VSEARCH) are highlighted with different colors and symbols. Se1 (Table 1), P1 (Table 2) and Si1 (Table 3) correspond to the most sensitive approach, the most precise approach and the approach with the best performance in terms of compositional similarity to the mock community, respectively. The bioinformatic approach recommended in this study is Si5 (Table 3).

5 Discussion

Metabarcoding approaches have revolutionized fungal ecology over the last decade (Hibbett et al., 2009) and have become the gold standard for describing the richness and composition of communities and the networks of associations between community members (Bálint et al., 2016). They have been so successful that fungal ecologists are struggling to cope with the boom in sequencing platforms, bioinformatic pipelines, taxonomic databases and community analysis tools. Benchmark studies and methodological reviews are required to help them make the most appropriate choices (e.g. Bálint et al., 2016 ; Lindahl et al., 2013 ; Pollock et al., 2018 ; Weiss et al., 2016). In this study, we focused on one aspect of the metabarcoding approach, bioinformatic analysis, assessing its effect on the recovery of community richness and composition. We compared the ability of 360 bioinformatic approaches to recover a mock community of fungal strains commonly found in soils and plants and including 97 genera from subkingdom Dikarya. This mock community was much larger than the fungal mock communities analyzed in previous studies (Amend et al., 2010 ; Bakker, 2018 ; Cline et al., 2017 ; Ihrmark et al., 2012 ; Nguyen et al., 2015 ; Taylor et al., 2016) and covered both the Ascomycota and Basidiomycota clades (Figure 1).

We selected three criteria for comparing bioinformatic approaches : sensitivity, precision and compositional similarity to the mock community. The first two criteria are related to the number of OTUs (or ASVs) recovered and are commonly used in benchmark studies (see Weiss et al., 2016). The third takes relative abundance into account and has been used by Bakker (2018). We believe that this third criterion is very important, particularly if the fungal metabarcoding data are to be used to reconstruct fungal association or interaction networks for biocontrol (Hassani et al., 2018 ; Poudel et al., 2016 ; Vacher et al., 2016b) or biomonitoring applications (Bohan et al., 2017 ; Derocles et al., 2018 ; Karimi et al., 2017). Indeed, network inference requires the most accurate possible recovery of microbial species and their abundances (Berry et Widder, 2014 ; Faust et Raes, 2012 ; Friedman et Alm, 2012 ; Weiss et al., 2016).

Our comparison revealed huge discrepancies between bioinformatic approaches, thereby confirming the importance of carefully selecting the most appropriate method for the analysis of fungal metabarcoding data (Anslan et al., 2018 ; Cline et al., 2017 ; Nguyen et al., 2015). The number of operational taxonomic units (OTUs) or amplicon sequence variants (ASVs) identified by the bioinformatic approaches compared ranged from 57 to 1562, even though there were only 189 strains in the mock community. These results confirm that fungal community analyses should not focus on absolute values of richness estimated from metabarcoding data, but rather on the relative changes in richness between samples (Cline et al., 2017). The percentage of fungal strains recovered by the bioinformatic approaches ranged from 22% to 87.5%. This second value may be considered a very good result, because we estimated the maximum sensitivity attainable by a bioinformatic approach, given our data, at 90.1%. Indeed, not all the strains in the mock community could be distinguished on the basis of their ITS1 sequences, and several strains were either not amplified at all or not accurately amplified. Our analyses revealed that four fungal species (*Lepiota clypeolaria*, *Mycena abramsii, M. galopus,* and *Panellus stipticus*) were absent from the sequence dataset (Figure 1), suggesting a lack of amplification by the so-called "universal" primers (Bellemain et al., 2010 ; Tedersoo et Lindahl, 2016) or a sequencing

failure (Nguyen et al., 2015 ; Palmer et al., 2018).

DADA2, a clustering-free software package (Callahan et al., 2016), effectively recovered the composition of the mock community. The top ten bioinformatic approaches in terms of performance for the compositional similarity criterion all used DADA2 to identify amplicon sequence variants. The total number of ASVs generated by these 10 approaches ranged from 148 to 197, which was therefore of the same order of magnitude as the total number of strains in the mock fungal community (i.e. 189). We highlighted several options for increasing the efficiency of DADA2 for fungal metabarcoding datasets. Firstly, our results confirm that the use of single forward (R1) reads as input data is a good option (Nguyen et al., 2015). This made it possible to ensure that strains with longer ITS regions (such as those of the genus *Cantharellus*, for instance ; Feibelman et al., 1994) were not excluded. Based on our results, we also recommend retaining the primers for fungal communities amplified with the ITS1F-ITS2 primer pair. Indeed, we found that primer removal did not improve the recovery of mock community composition. These findings may be accounted for by the absence of degenerate nucleotides in the ITS1F-ITS2 primer pair. Primer retention may be relevant in this case, because non-degenerate primers have no impact on the denoising step of DADA2. The merging of reads after sequence variation inference, as recommended in the DADA2 tutorial (http://benjjneb.github.io/dada2/tutorial.html), did not improve the recovery of the mock community either.

The Si5 approach represented one of the best trade-offs between the three selection criteria among the 360 bioinformatic approaches compared. The Si5 approach used single forward (R1) reads as input. Quality filtering, sequence variation analysis and chimera removal were performed with DADA2 (Callahan et al., 2016). ITS1 extraction (Bengtsson-Palme et al., 2013) and downstream OTU table filtering were not required. The Si5 approach recovered the ITS1 regions of 80 out of 87 Ascomycota strains, 83 out of 99 Basidiomycota strains and all 3 Mucoromycota strains (Figure 1), suggesting that there was no detection bias against Ascomycota strains despite the intron insert downstream of the ITS1F primer binding site that might impair their amplification (see Taylor et al., 2016). We recommend the use of this simple bioinformatic approach in ecological studies of fungal communities, for the following reasons : (i) it did not overestimate the number of fungal strains, (ii) it was among the ten best bioinformatic approaches in terms of recovery of the composition of the mock community and (iii) it performed very well according to the two other criteria used for comparison (precision and sensitivity). Based on these results, the Si5 approach appears to be an appropriate bioinformatic approach for studies involving whole-community profiling and network inference.

By contrast, the clustering algorithms of USEARCH (Edgar, 2010) and VSEARCH (Rognes et al., 2016) should be favored in studies in which species detection is the main goal. These clustering algorithms generally overestimated the actual number of fungal strains, but were able to retrieve almost all detectable strains. Their sensitivity was close to the maximum value. The most sensitive approach, Se1, used single forward (R1) reads as input and clustered them with the VSEARCH algorithm. ITS1 extraction (Bengtsson-Palme et al., 2013), chimera removal and downstream OTU table filtering were not required. In general, our comparison revealed that the steps of ITS extraction and chimera removal can eliminate fungal strains that are actually present in the community and should not be systematically used. The second most sensitive approach, Se2, used the USEARCH clustering algorithm. These two highly sensitive bioinformatic approaches are potentially useful for the early detection of invasive species (Comtet et al., 2015), including fungal pathogens (Munck et Bonello, 2018), for the detection

of emerging pathogens accounting for the decline or death of host populations (Ricciardi et al., 2017), and for exploring environmental reservoirs of pathogens (Agtmaal et al., 2017). On the other hand, if the purpose of a study is to focus only on fungal species present with high certainty (i.e. on a precise but incomplete community), then DADA2 and LULU (Frøslev et al., 2017) should be combined and applied to assembled sequences. The Pi3 approach, that merges reads after sequence variation inference as recommended in the DADA2 tutorial, seems to be a good compromise in this case.

Overall, our study highlights the importance of carefully selecting the bioinformatic approach to be used according to the objective of the metabarcoding study. Indeed, the ability of bioinformatic approaches to recover fungal strains and the relative abundances of the strains recovered varied greatly. Some approaches detected almost all strains of the mock community but overestimated community richness, whereas others retrieved the actual richness and composition of the mock community more accurately. The former are more appropriate for the detection of target species, whereas the latter are more appropriate for community ecology studies. However, none of the bioinformatic approaches compared recovered the mock community perfectly. In particular, none of the approaches found the expected distribution of sequences between fungal strains. This may be due to differences in the number of ribosomal RNA gene repeats between fungal species (Ganley et Kobayashi, 2007), and imperfections in equimolar pooling of DNA samples, together with biased amplification for pooled species (Palmer et al., 2018). Because of these biases, current fungal community analyses should not focus on the within-sample distribution of taxa abundance, but rather on the changes in taxa abundance between samples. Future methodological developments should focus on reducing biases caused by molecular biology steps (Nichols et al., 2018 ; Porter et Hajibabaei, 2018) and on improving the bioinformatics pipelines to better recover the abundances of fungal strains. Our comparison of bioinformatics approaches could be extended, since the 360 bioinformatic approaches compared here constitute only a small fraction of the approaches that could be used to analyze fungal metabarcoding data. Other approaches may give better results, and their ranking may vary with sequence data quality (Nguyen et al., 2015). Future bioinformatic approach comparisons should therefore be based on multiple mock communities sequenced independently. They could also include error-correction methods alternative to that of DADA2, such as UNOISE2 (Edgar, 2016), or recent clustering approaches, such as OptiClust (Westcott et Schloss, 2017) or SeekDeep (Hathaway et al., 2018), or consider reference-based clustering approaches (Cline et al., 2017 ; Halwachs et al., 2017) but see Westcott et Schloss (2015). All the data required for the extension of our methodological comparison are provided.

Data Availability

The raw sequence data were deposited in Dataverse and are available in the FASTQ format at https://doi.org/10.15454/8CVWRR.
The code is available as an archive at https://doi.org/10.15454/VKTWKR.

Acknowledgments

We thank Matthieu Barret, Martial Briand, Lucas Auer, Gregory Gambetta, Guilherme Martins, Frédéric Barraquand and Tania Fort for useful discussions on the first draft of the manuscript. We also thank the editor, Peter Kennedy and one anonymous reviewer for helpful comments of the submitted version. We are grateful to the INRA MEM metaprogramme (Meta-Omics of Microbial Ecosystems) for financial and scientific support. The mock community sequencing

was funded by the INRA MEM MetaBAR project (PI : MB) and the bioinformatic analyses were performed as part of the INRA MEM Learn-biocontrol project (PI : CV). Additional funding was received from the LABEX COTE (ANR-10-LABX-45) and the LABEX CEBA (ANR-10-LABX-25-01). CP's PhD grant was funded by the INRA and Bordeaux Sciences Agro (BSA). We thank the Genotoul sequencing facility (Get-PlaGe) for sequencing the mock community and the Genotoul bioinformatics facility (Bioinfo Genotoul) for providing computing and storage resources. We also thank Julie Sappa from Alex Edelman & Associates for English language revision.

Author contributions

CP performed the bioinformatic work, analyzed the results in accordance with the recommendations of IL and CV, and wrote the first draft of the article in collaboration with JV and CV. MB coordinated the design and sequencing of the mock community. MB, VL, VEH and LF provided fungal DNA for the mock community and performed the molecular biology work. AG and VL provided the Sanger sequence database. CV conceived the study in collaboration with MB, supervised the work and made a major contribution to the writing of the manuscript. All authors revised the manuscript.

6 Supplementary information

Table 4 – List of the fungal strains of the mock community and their ITS Sanger sequences. The following properties of the mock strains are also indicated : the uniqueness of their ITS1 sequence (Unique_ITS1_Sanger_Sequence), whether the exact ITS1 sequence was present or not in the raw MiSeq data (Present_in_Raw_MiSeq_Data) and the putative cause of absence (Putative_Cause_Absence). Absences were attributed either to the presence of ambiguous bases in the Sanger sequence preventing a perfect match with the MiSeq sequence (AB), or to sequencing errors preventing a perfect match between MiSeq and Sanger sequences (SE), or to lack of amplification due to an imperfect match with the ITS2 primer sequence (LA) or to the loss of the strain during DNA pooling (DP). In the case of AB and SE, the percentage of similarity between the Sanger sequence and the closest MiSeq sequence is given into brackets. To assign the strain to the LA category, the sequence of the ITS2 primer was searched for in both the Sanger sequence and the closest NCBI sequence (given into brackets). Strains that were absent but did not fall into the three first categories were classified into the DP category.

Attached as a .csv file
https://ars.els-cdn.com/content/image/1-s2.0-S1754504818302800-mmc2.csv

Table 5 – Sequence summary statistics for the raw data and the assembled reads, before quality filtering. Values were averaged over the 3 replicates for each pipeline.

Assembly	Mean number of reads	Mean quality	Median length	Min/Max length	% reads [0,100 nt[	% reads [100, 200 nt[	% reads [200, 300 nt[	% reads ≥ 300 nt
RAW_R1	47958	35.59	250	250/250	0	0	100	0
RAW_R2	47958	34.31	250	250/250	0	0	100	0
FASTQJOIN_50	26209	37.77	267	250/444	0	0	97.85	2.15
FASTQJOIN_100	26132	37.77	267	250/400	0	0	97.85	2.15
FASTQJOIN_150	23972	37.78	266	250/350	0	0	97.85	2.15
PEAR_50	47439	39.01	267	50/447	0.04	0.1	97.78	2.08
PEAR_100	47283	39.02	267	100/400	0	0.1	97.82	2.09
PEAR_150	43835	39.13	265	150/350	0	0.06	97.85	2.09

Table 6 – Sequence summary statistics for the raw data and the assembled reads, after quality filtering. The first two columns correspond to sequence with a Phred score higher than 30 on 75% of length (Q30). The last two columns correspond to sequences with less than one expected error according to error probabilities obtained from quality scores (EE). For the CUTADAPT_MERGED assembly method, reads with less than two expected errors were retained. Percentages of reads passing quality filter were computed over the total number of raw reads.

Assembly	Mean number of reads (Q30)	% of reads (Q30)	Mean number of reads (EE)	% of reads (EE)
QUALITY_R1	38356	80%	37441	78.1%
FASTQJOIN_50	24499	93.5%	25857	98.7%
FASTQJOIN_100	24425	93.5%	25821	98.8%
FASTQJOIN_150	22277	92.9%	23832	99.4%
PEAR_50	45283	95.5%	44707	94.2%
PEAR_100	45162	95.5%	44657	94.4%
PEAR_150	42080	96%	41970	95.7%
CUTADAPT_MERGED	NA	NA	38058	79.4%

Table 7 – Number and length of sequences after extraction of the ITS1 region with ITSx. Percentages of reads extracted were computed over the total number of quality reads (Q30) (Table S6).

Assembly	Mean number of reads	Mean quality	Median length	Min/Max length	% reads [0,100 nt[	% reads [100, 200 nt[	% reads [200, 300 nt[	% reads ≥ 300 nt
QUALITY_R1	22622	59 %	0.84	99.12	0.04	0		
FASTQJOIN_50	24155	98.6%	0	82.91	17.09	0		
FASTQJOIN_150	22144	99.4%	0	90.36	9.64	0		
FASTQJOIN_100	24121	98.8%	0	82.95	17.05	0		
PEAR_50	44109	97.4%	0.49	86.47	13.04	0		
PEAR_100	44065	97.6%	0.49	86.5	13.01	0		
PEAR_150	41443	98.5%	0.52	91.97	7.51	0		

Table 8 – Number of OTUs (or ASVs) and sequences in the final OTU (ASV) table according to the method of sequence variation analysis and the filtering option, when QUALITY_R1 reads without ITS1 extraction and without chimeric sequences were used. Values were averaged over the 3 replicates for each bioinformatic approach. LULU was applied with default settings.

Variation	*Filtering*	Mean number of OTUs (or ASVs)	Mean number of sequences
DADA2	All	157	34991
DADA2	1	157	34991
DADA2	10	148	34950
DADA2	RA	152	34978
DADA2	LULU	105	34991
USEARCH	All	878	38245
USEARCH	1	660	38027
USEARCH	10	329	37408
USEARCH	RA	437	37688
USEARCH	LULU	298	38245
VSEARCH	All	577	37760
VSEARCH	1	428	37611
VSEARCH	10	257	37293
VSEARCH	RA	311	37444
VSEARCH	LULU	253	37760

Table 9 – Sensitivity, precision and similarity values for all bioinformatic approaches.

Attached as a .csv file
https://ars.els-cdn.com/content/image/1-s2.0-S1754504818302800-mmc3.csv

Table 10 – Effect of primer removal on the sensitivity, precision and similarity values for the Si1 and Si5 pipelines.

Pipeline	Primers	Richness	Sensitivity	Precision	Similarity
Si1	Retained	197	0.774	0.743	0.396
Si1	Removed	200	0.774	0.735	0.396
Si5	Retained	157	0.758	0.915	0.393
Si5	Removed	157	0.758	0.913	0.393

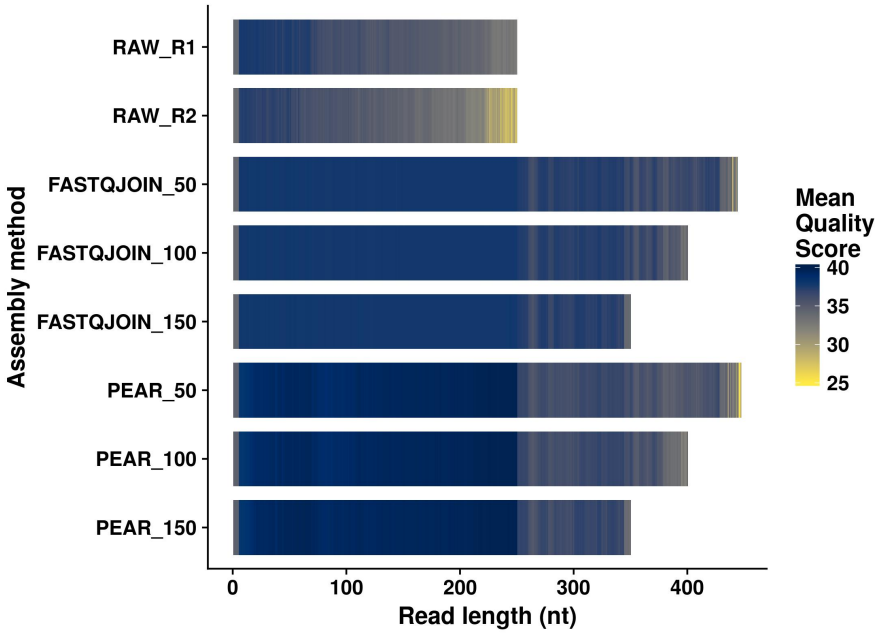


FIGURE 5 – Mean quality score along raw reads and assembled reads before quality filtering. The total number of sequences (N) is indicated for each category of reads.

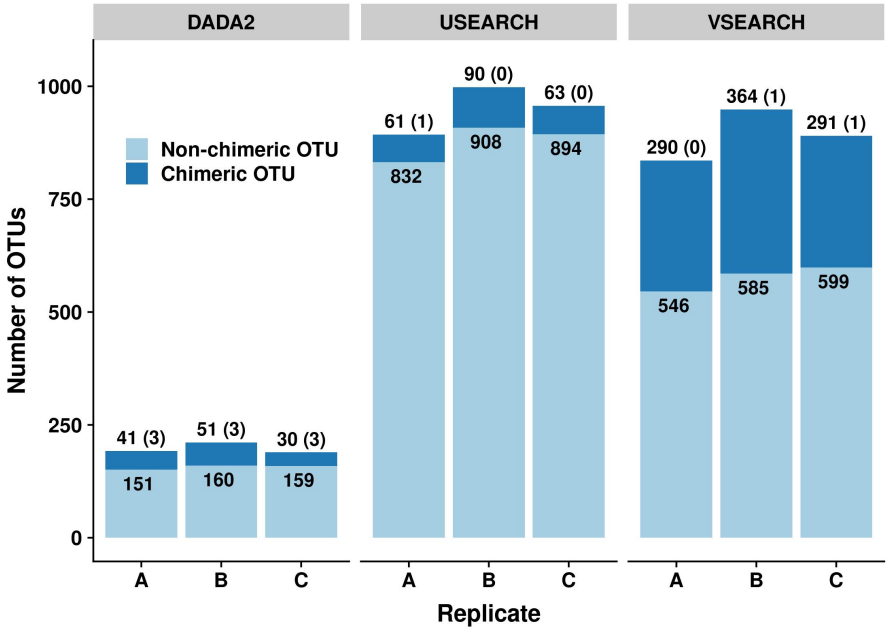


FIGURE 6 – Total number of OTUs (or ASVs) per replicate as a function of the method for analyzing sequence variation, when using QUALITY_R1 reads without ITS1 extraction. The number of True Positives OTUs present in the Chimeric OTUs is indicated in brackets.

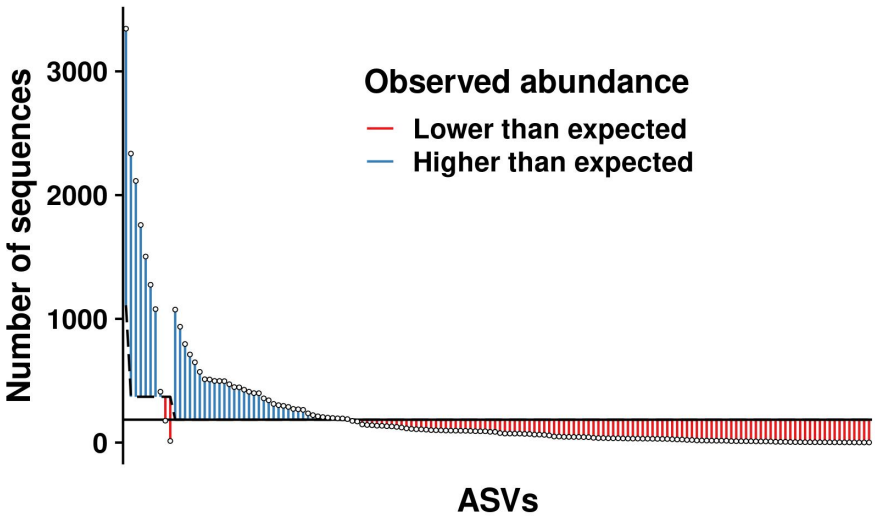


FIGURE 7 – Variability of sequence counts among the true positive ASVs generated by the Si5 approach. Dots correspond to the observed number of sequences per ASV (averaged over the 3 replicates). The horizontal black plain line indicates the expected number of reads per ASV, under the assumption of a uniform distribution of sequences between fungal strains of the mock community. Under this assumption, the expected number of reads per ASV is defined as the final number of sequences obtained with the bioinformatic approach Si5 divided by the total number of strains in the mock community. The black dashed line indicates the expected number of reads per ASV multiplied by the total number of fungal strains (with identical ITS1 sequences) corresponding to an ASV. Deviations from this expectation are indicated with blue and red bars.

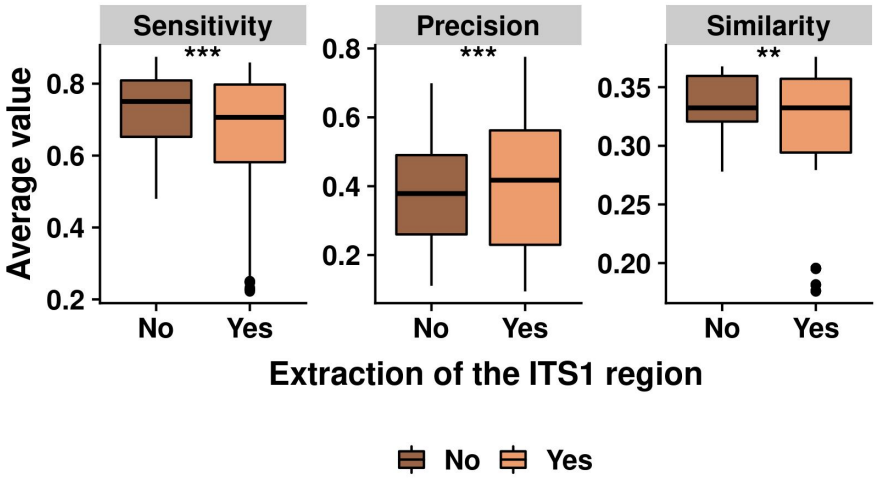


FIGURE 8 – Sensitivity, precision and compositional similarity obtained for all the bioinformatic approaches using USEARCH and VSEARCH, as a function of the extraction of the ITS1 region with ITSx. Values were averaged over the 3 replicates for each bioinformatic approach and compared using paired *t*-tests.

III

Les réseaux microbiens inférés manquent de réplicabilité : implications pour la biosurveillance de nouvelle génération

1 Abstract

Inferred microbial networks lack replicability : consequences for next-generation biomonitoring

Charlie Pauvert[1], Jessica Vallance[2,3], Laurent Delière[2,4], David A. Bohan[5], Marc Buée[6], Corinne Vacher[1*]

1– BIOGECO, INRA, Univ. Bordeaux, 33615 Pessac, France. 2 – INRA, UMR 1065 Santé et Agroécologie du Vignoble, ISVV, F-33882 Villenave d'Ornon, France. 3– Université de Bordeaux, Bordeaux Sciences Agro, UMR 1065 SAVE, F-33175 Gradignan, France. 4– INRA, UE 1442 Vigne Bordeaux, F-33883 Villenave d'Ornon. 5– Agroécologie, AgroSup Dijon, INRA, Université Bourgogne, Université Bourgogne Franche-Comté, Dijon, France. 6– INRA, UMR 1136 Interactions Arbres/Micro-Organismes, F-54280, Champenoux, France.

* Correspondence : Dr. Corinne Vacher (corinne.vacher@inra.fr)

Plant-associated microbial interaction networks protect plants against disease. There is, therefore, a need to monitor in real time their responses to environmental changes to predict disease risk and adjust crop protection strategies. Next-Generation Biomonitoring (NGB) proposes to reconstruct automatically these networks from metabarcoding data, to complement ecological community properties commonly used for ecosystem health assessment. This study aimed to evaluate the benefits and shortcomings of community-level and network-level properties for biomonitoring. We specifically investigated whether microbial networks inferred from metabarcoding data show robust responses to agricultural practices, using the grapevine microbiota as a study system. Our results demonstrate a strong footprint of the agricultural practice on the metabarcoding data, when analyzed at the community level. The richness, diversity and evenness of fungal communities were significantly higher in organic than conventional plots. The cropping system also affected the composition of grapevine foliar fungal communities significantly. Contrary to our expectations, microbial networks were less sensitive to changes in agricultural practices than microbial communities, confirming that NGB should not only consider network-level properties but also community-level properties. Moreover, we found that microbial networks lacked replicability within a cropping system but that consensus networks, built from several network replicates, could generate relevant hypotheses of microbial interactions. As things stand, community-level properties appear to be a more reliable and statistically powerful monitoring option than network-level properties. Future developments, especially in network inference methods, are likely to challenge our findings and help to improve the monitoring of the ecosystem services provided by the plant microbiota.

Keywords : Agricultural practices, Community ecology, Ecosystem services, Microbial networks, Network inference, Pathobiome.

Available as a preprint at https://doi.org/10.1101/642199

2 Introduction

Interactions among organisms and with their abiotic environment regulate the ecological processes underlying ecosystem services (Mace et al., 2012). Ecological interactions among organisms (e.g. predation, mutualism, parasitism) at a single point in space and time are usually represented as a network, with the organisms as nodes and the interactions as links (Pocock et al., 2012). Current challenges focus on understanding how and why these networks vary in space and time (Pellissier et al., 2018 ; Pilosof et al., 2017), and which network properties should be conserved or enhanced to sustain ecosystem services (Montoya et al., 2006 ; Raimundo et al., 2018 ; Tylianakis et al., 2010). Next-Generation Biomonitoring (NGB) proposes to reconstruct ecological networks on top of community structures from Next-Generation Sequencing (NGS) data, and to analyze network and community variations in space and time for detecting and explaining changes in ecosystem functions and services (Baird et Hajibabaei, 2012 ; Bohan et al., 2017 ; Derocles et al., 2018).

NGB requires the reconstruction of replicated networks of ecological interactions as (Delmas et al., 2019 ; Pellissier et al., 2018 ; Poisot et al., 2012 ; Tylianakis et Morris, 2017). By analogy with the α- and β-diversity of ecological communities, these frameworks define α- and β-properties for ecological networks as whole-network metrics (e.g. connectance) and dissimilarities between pairs of networks, respectively (Pellissier et al., 2018). These metrics can be used to assess the impact of global environmental changes on the identity and abundance of the species forming ecological communities and on the type and strength of their interactions (Pecl et al., 2017 ; Scheffers et al., 2016). They can for instance be used to evaluate the impact of agricultural practices, which are a key driver of global change (Tilman et al., 2002), on species diversity (Tuck et al., 2014) and on pest and disease regulation services supported by species interactions (Ma et al., 2019 ; Macfadyen et al., 2009 ; Tylianakis et al., 2007). Organic farming provides indeed lower levels of pests (Muneret et al., 2018) but others outcomes of such agriculture remains uncertain (Seufert et Ramankutty, 2017). Network metrics (Pellissier et al., 2018) could provide further insights into the footprint of agricultural practices, like organic farming, on the functioning of ecological communities and associated ecosystem services.

Networks of interactions among microorganisms are a natural target for NGB, because NGS techniques are the rule for studying microbial communities (Bálint et al., 2016), and because microbial networks are crucial to human life and well-being (Gilbert et Neufeld, 2014). Network ecology, which originates from the study of trophic links between macroorganisms (Ings et al., 2009), initially ignored interactions with and among smaller organisms (Lafferty et al., 2006) despite evidence of the contribution of microbial interactions to biogeochemical cycles (Falkowski et al., 2008), plant diversity and productivity (van der Heijden et al., 2008) and disease regulation (Berendsen et al., 2012). It is now recognized that microbial networks are particularly important for the health of animals and plants (i.e. holobionts ; Zilber-Rosenberg et Rosenberg, 2008) because resistance to pathogens is mediated by direct antagonistic interactions between the residential microbiota and the pathogen species (i.e. the barrier effect ; Arnold et al., 2003 ; Kamada et al., 2013 ; Kemen, 2014 ; Koch et Schmid-Hempel, 2011 ; Laur et al., 2018). and by indirect interactions due to the activation of the host immune system by the residential microbiota (i.e. the priming effect ; Hacquard et al., 2017 ; Kamada et al.,

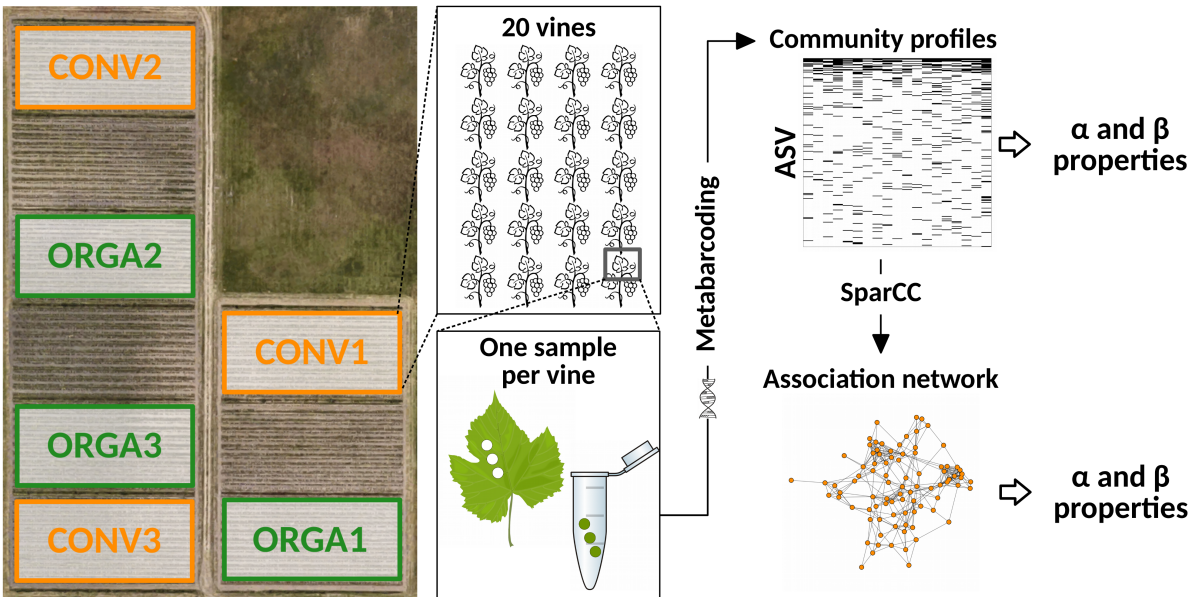


Figure 1 – Experimental design. Foliar fungal communities were characterized in three conventional (CONV) and three organic (ORGA) vineyard plots by a metabarcoding approach. We analyzed 20 foliar samples per plot. For each plot, we thus obtained 20 community profiles (described in terms of amplicon sequence variants (ASV)) and one association network (inferred with the SparCC software developed by Friedman et Alm (2012)). More networks were then obtained by varying network reconstruction parameters (Figure 5). The effects of cropping system (CONV versus ORGA) on the grapevine foliar microbiota were assessed with both community and network α- and β-properties.

2013 ; Perazzolli et al., 2012 ; Ritpitakphong et al., 2016 ; Vogel et al., 2016). The subset of a host-associated microbial network consisting of a pathogen and its interacting partners has been termed the pathobiome (Brader et al., 2017 ; Vayssier-Taussat et al., 2014). NGB of plant and animal health will require elucidation of the microbial interactions forming pathobiomes (Durán et al., 2018), identification of the intrinsic network properties that hinders invasion by pathogens (Agler et al., 2016 ; Murall et al., 2017 ; Poudel et al., 2016) and assessment of the changes in these properties (Creamer et al., 2016 ; Karimi et al., 2017 ; Morriën et al., 2017).

The success of NGB approaches will depend on our ability to automatically reconstruct networks that are similar to real ecological networks (Bohan et al., 2011). Classically, ecological networks have been built on the basis of observations of ecological interactions. Metabarcoding approaches are now providing additional information (Evans et al., 2016), and making it possible to generate hypotheses about cryptic interactions such as those of microorganisms (Berry et Widder, 2014 ; Faust et Raes, 2012). Hypotheses for microbial interactions can be formulated on the basis of co-abundance data derived from environmental DNA metabarcoding data (i.e. the sample × taxa matrix) via statistical or machine-learning approaches (Vacher et al., 2016b). The representation of all positive and negative statistical associations between abundances for microbial taxa typically gives rise to a network resembling a hairball (Röttjers et Faust, 2018) and the challenge is to go beyond networks of this type, by removing spurious associations not due to microbial interactions and linking the remaining positive and negative associations to possible interaction mechanisms (Derocles et al., 2018). The first source of spurious associations is the compositional nature of metabarcoding data. In a metabarcoding dataset, the total number of sequences per sample is arbitrary, imposed by the sequencer.

Sequence counts contain only relative abundance information for species. Comparisons that do not take this feature into account can result in the identification of artifactual associations (Gloor et al., 2017). Early methods of microbial network inference, such as SparCC (Friedman et Alm, 2012) attempted to overcome this bias using log ratios of counts. A second source of spurious associations is the joint response of microbial taxa to abiotic or biotic factors, creating indirect associations that reflect taxon-specific environmental requirements (Armitage et Jones, 2019 ; Röttjers et Faust, 2018). Early studies dealt with this issue by first using regression to eliminate environmental effects from the sequence counts and then inferring networks from the residuals (Biswas et al., 2016 ; Jakuschkin et al., 2016) and several methods integrating environmental covariates, such as HMSC (Ovaskainen et al., 2017), PLN (Chiquet et al., 2017 ; Chiquet et al., 2019), FlashWeave (Tackmann et al., 2019) and MAGMA (Cougoul et al., 2019b), have since been developed. Finally, the taxonomic resolution of the nodes should be fine enough to discern the variation in ecological interactions between microbial strains (Röttjers et Faust, 2018). This has led to the generation of novel bioinformatics approaches that more fully exploit the resolution of molecular barcodes, such as DADA2 (Callahan et al., 2016). Despite these advances, however, the inference of microbial networks from metabarcoding data remains nascent (Layeghifard et al., 2017), and inferred interactions should be interpreted with care (Barner et al., 2018 ; Freilich et al., 2018 ; Röttjers et Faust, 2018 ; Weiss et al., 2016 ; Zurell et al., 2018), because few validation experiments have been conducted to date (e.g. Das et al., 2018 ; Durán et al., 2018 ; Wang et al., 2017).

In this study, we investigated whether the impact of agricultural practices on microbial networks might be detected by combining current metabarcoding and network inference approaches. We first assessed the influence of conventional versus organic agriculture on plant-associated microbial communities using community-level properties and secondly using network-level properties. Furthermore, we checked whether microbial networks can generate robust hypotheses concerning microbial interactions. We inferred microbial networks from environmental DNA sampled from replicated agricultural plots and processed in DADA2 (Callahan et al., 2016) and SparCC (Friedman et Alm, 2012). Grapevine was the model plant species because effects of organic farming were detected on associated microbial communities (Castañeda et al., 2018 ; Kernaghan et al., 2017 ; Pancher et al., 2012 ; Schmid et al., 2011 ; Varanda et al., 2016) but remain to be assessed on networks. We focused on the fungal component of the foliar microbiota as it contains major pathogens (Armijo et al., 2016).

3 Materials and methods

3.1 Study site and sampling design

Samples were collected in 2015, from an experimental vineyard (Figure 1) located near Bordeaux (INRA, Villenave d'Ornon, France ; 44°47'32.2"N 0°34'36.9"W). The experimental vineyard was planted in 2011 and was designed to compare three cropping systems : sustainable conventional agriculture (CONV), organic farming (ORGA) and pesticide-free farming (RESI) (Delière et al., 2014). The *Vitis vinifera* L. cultivar Merlot noir grafted onto a 3309 C rootstock was used in both the CONV and ORGA cropping systems. RESI used a disease-resistant cultivar and was not included in this study. The experiment had a randomized block design (Schielzeth et Nakagawa, 2013) consisting of three blocks, each composed of three plots, one for each of the cropping systems tested. Each plot covered an area of 2100 m² and was composed of 20 rows of 68 vines each, with a $1.60m^2$ between rows and $0.95m^2$ between vines in a single row.

CONV plots were managed according to the general principles of integrated pest management (IPM), as listed in Appendix III of the 2009/128/EC Directive (European Commision, 2009). ORGA plots were managed according to European Council Regulation (EC) No 834/2007 (European Commision, 2007). ORGA plots were treated with copper and sulfur-based products, whereas additional phytosanitary products were allowed in CONV plots (Table S4). The cropping systems differed in terms of the types of pesticides applied and the timing of applications, but not in terms of doses (Table S4). All products and active ingredients were applied between the end of April and mid-August of 2015. Grapes were harvested on September 10, 2015. The disease incidence and severity at harvest were higher in CONV plots than in ORGA plots for both powdery mildew (caused by the fungal pathogen *Erysiphe necator*) and black rot (caused by the fungal pathogen *Guignardia bidwellii*). Downy mildew symptoms (caused by the oomycete pathogen *Plasmopara viticola*) did not differ significantly between the cropping systems (Table S5).

Grapevine leaves were collected a couple hours before grape harvest, from 20 vines per plot in the CONV and ORGA plots (Figure 1). We attempted to avoid edge effects by selecting the 20 vines from the center of each plot. The third leaf above the grapes was collected from each vine, placed in an individual bag and immediately transported to the laboratory. In total, 120 leaves, corresponding to 1 leaf × 20 vines × 3 plots × 2 cropping systems, were collected. Leaves were processed on the day of collection, with sterilized tools in the sterile field of a MICROBIO electric burner (MSEI, France). Three contiguous discs of 6 mm diameter were cut from the center of each leaf, approximately 2 cm from the midrib. They were placed in the well of a sterile DNA extraction plate. The leaf disks were then freeze-dried overnight (Alpha 1-4 DA Plus, Bioblock Scientific).

3.1.1 *DNA extraction and sequencing*

Leaf disks (Figure 1) were ground with a single-glass ball mill (TissueLyser II, Qiagen) and DNA was then extracted with a CTAB chloroform/isoamyl alcohol (24 :1) protocol. A dozen

"empty" wells (i.e. containing nothing but extraction reagents) were included on each plate as negative control samples for DNA extraction. Three of these negative control samples were randomly selected and pooled before sequencing. Three replicates of a fungal mock community, each consisting of an equimolar pool of DNA from 189 pure fungal strains, were also included as positive control samples (Pauvert et al., 2019a).

The nuclear ribosomal internal transcribed spacer (ITS) region, which is considered to be the universal barcode region for fungi (Schoch et al., 2012), was then amplified with the ITS1F (5'-CTTGGTCATTTAGAGGAAGTAA-3', Gardes et Bruns, 1993) and ITS2 (5'-GCTGCGTTCTT-CATCGATGC-3', White et al., 1990) primers primer pair, which targets the ITS1 region. PCR was performed in an Eppendorf thermocycler (Eppendorf), with a reaction mixture (25 μl final volume) consisting of 0.04 U Taq polymerase (SilverStar DNA polymerase, Eurogentec), 1X buffer, 2 mM MgCl2, 200 μM of each dNTP, 0.2 μM of each primer, 1 ng.μl^{-1} bovine serum albumin (New England BioLabs) and 2 μl DNA template. A pseudo-nested PCR protocol was used, with the following cycling parameters : enzyme activation at 95°C for 2 min ; 20 (1st PCR with regular primers ; Table S6) and then 15 (2nd nested PCR with pre-tagged primers ; Table S6) cycles of denaturation at 95°C for 30 s, 53°C for 30 s, 72°C for 45 s ; and a final extension phase at 72°C for 10 min. "Empty" wells (i.e. containing nothing but PCR reagents) were included on each plate as a negative control for PCR. Three negative control samples were randomly selected and pooled before sequencing. In addition, the six PCR products corresponding to the 24th leaf of the six plots under study (three CONV plots, three ORGA plots) were split in two, with each half of the sample sequenced independently to serve as technical replicates for sequencing.

We checked the quality of all the PCR products by electrophoresis in 2% agarose gels. PCR products were purified (CleanPCR, MokaScience), multiplex identifiers and sequencing adapters were added, and library sequencing on an Illumina MiSeq platform (v3 chemistry, 2×250 bp) and sequence demultiplexing (with exact index search) were performed at the Get-PlaGe sequencing facility (Toulouse, France).

3.2 Bioinformatic analysis

Based on the mock community included in the sequencing run, we found that analyzing single forward (R1) sequences with DADA2 (Callahan et al., 2016) was a good option for fungal community characterization (Pauvert et al., 2019a). Using DADA2 v1.6, we retained only R1 reads with less than one expected error (based on quality scores ; Edgar et Flyvbjerg, 2015) that were longer than 100 bp, and we then inferred amplicon sequence variants (ASV) for each sample. Chimeric sequences were identified by the consensus method of the *removeBimeras* function. Taxonomic assignments were performed with RDP classifier (Wang et al., 2007), implemented in DADA2 and trained with the UNITE database v. 7.2 (UNITE Community, 2017). Only ASVs assigned to a fungal phylum were retained. The ASV table was then filtered as described by Galan et al. (2016) with a script (https://gist.github.com/cpauvert/1ba6a97b01ea6cde4398a8d531fa62f9) that removed ASVs from all samples for which the number of sequences was below the cross-contamination threshold, defined as their maximum number in negative control samples. Finally, we checked the compositional similarity of the technical replicates (Figure S7), retaining the sample with the largest number of sequences for each replicate. The final ASV table contained 1116 ASVs, 112 samples and 4,760,068 high-quality sequences.

Table 1 – Most abundant amplicon sequence variants (ASVs) in grapevine foliar fungal communities according to cropping system. The relative abundances (RA, in %) and ranks of ASVs were calculated for all leaf samples (TOTAL ; $n = 112$) and for samples collected from organic (ORGA ; $n = 55$) and conventional plots (CONV ; $n = 57$).

ASV taxonomic assignment	TOTAL		ORGA		CONV	
	Rank	RA	Rank	RA	Rank	RA
Aureobasidium sp.	1	61.4	1	55.8	1	66.7
Cladosporium delicatulum	2	6.3	4	6.9	2	5.8
Filobasidium sp.	3	5.1	2	9.7	9	0.7
Alternaria sp.	4	4.4	5	3.9	4	5.0
Epicoccum nigrum	5	4.1	7	2.7	3	5.4
Cladosporium ramotenellum	6	3.5	3	7	46	<0.1
Mycosphaerella tassiana	7	3.3	8	1.8	5	4.8
Didymella sp.	8	1.4	6	2.7	33	0.1
Erysiphe necator	9	1.1	38	<0.1	6	2
Vishniacozyma victoriae	10	0.9	9	1.6	17	0.3

3.3 Statistical analyses

Statistical analyses were performed with R software v3.4.1 (R Core Team, 2014), with the packages lme4 (Bates et al., 2015), vegan (Oksanen et al., 2018), permute (Simpson, 2016), phyloseq (McMurdie et Holmes, 2013) including the DESeq2 extension (Love et al., 2014), and igraph (Csardi et Nepusz, 2006). Data were manipulated and plots were created with reshape2, plyr and ggplot2 (Wickham, 2007 ; Wickham, 2009 ; Wickham, 2011), cowplot (Wilke, 2018), ggraph (Pedersen, 2018) and VennDiagram (Chen, 2018).

Effect of cropping system on community α-diversity Generalized linear mixed models (GLMMs) were used to test the effect of cropping system on the richness, diversity and evenness of fungal communities. The models included the cropping system as a fixed treatment effect, the block and its interaction with the cropping system as random factors, and the sampling depth (defined as the total number of raw sequences per sample) as an offset s(Bálint et al., 2015 ; McMurdie et Holmes, 2014). Community richness was defined as the number of ASVs per sample. We used a logarithmic link function to model these count data, assuming a negative binomial distribution to deal with overdispersion (Zuur et al., 2009). Community diversity was measured with the Inverse Simpson index (Simpson, 1949) and modeled with a Gaussian distribution and the logarithmic link function. Evenness was estimated with Pielou's index (Pielou, 1966) and modeled with a Gaussian distribution and the logarithmic link function. The offset was transformed according to the link function. The significance of the fixed treatment effect was finally assessed with the Wald χ^2 test (Bolker et al., 2009).

Effect of cropping system on community β-diversity Permutational analyses of variance (PERMANOVAs ; Anderson, 2001) were used to evaluate the effect of cropping system on compositional dissimilarities between fungal communities detected with the quantitative and binary versions of the Jaccard dissimilarity index (Chao et al., 2006 ; Jaccard, 1900). The models included cropping system, sampling depth (log-transformed) and their interaction as fixed effects. Permutations ($n = 999$) were constrained within blocks. ASVs differing in abundance

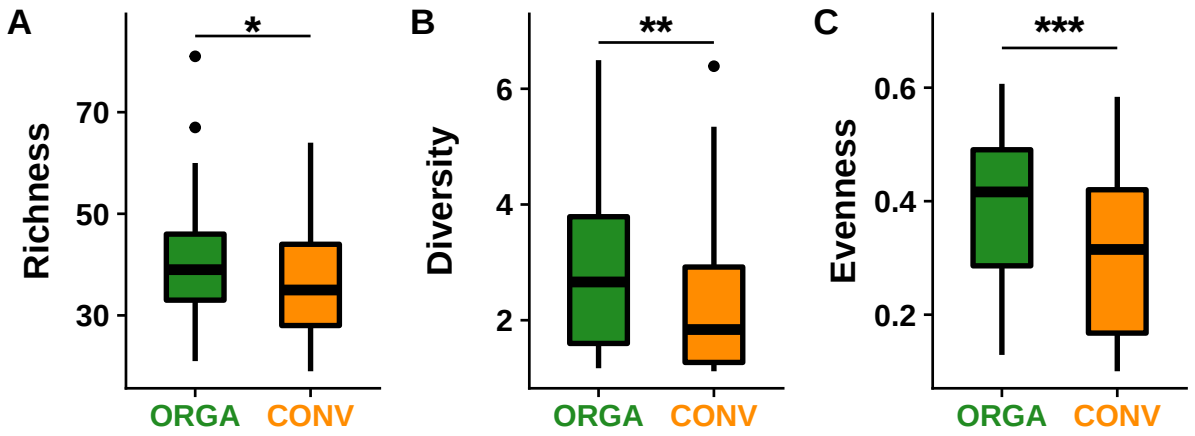


FIGURE 2 – Effect of cropping system – conventional (CONV) versus organic (ORGA) – on the α-diversity metrics of grapevine foliar fungal communities. (A) Community richness, defined as the number of ASVs. (B) Community diversity, measured with the inverse Simpson index. (C) Community evenness, measured with Pielou's index. Differences in α-diversity metrics between cropping systems were evaluated in Wald χ^2 tests (* p<0.05 ; **p<0.01 ; ***p<0.001).

between cropping systems were identified with DESeq2 (LOVE et al., 2014), by calculating the likelihood ratio between a full model including block and cropping system as fixed effects and a simplified model including only the block factor. The estimated fold-changes in abundance were considered significant if the p-value was below 0.05 after Benjamini and Hochberg adjustment.

Effect of cropping system on network α-properties Fungal association networks were inferred at plot level (Figure 1) with the SparCC algorithm (FRIEDMAN et ALM, 2012) implemented in FastSpar (WATTS et al., 2019) with default SparCC values. Ten networks per plot were constructed by varying the percentage P of ASVs included in the network (with P ranging from 10% to 100% of the most abundant ASVs in the plot). Networks had ASVs as nodes and a positive or negative link between ASVs in cases of significant associations between abundance. Six α-properties were calculated for all networks : number of links, network density, number of connected components, diameter of the largest component, mean node degree and proportion of negative links (Table S7). The effect of cropping system on these properties was investigated by performing Wilcoxon rank-sum tests for all values of P. The Benjamini-Hochberg procedure was used to correct p-values for multiple testing.

Effect of cropping system on network β-properties The topological distance between networks was calculated for all pairs of networks, with the D index defined by SCHIEBER et al. (2017) for binary networks. The dissimilarity of associations between networks, β_{WN} according to the framework described by POISOT et al. (2012), was then calculated for all pairs of networks with the binary Jaccard dissimilarity index. β_{WN} was then partitioned into two components (POISOT et al., 2012) : the dissimilarity of associations between ASVs common to both networks (β_{OS}) and the dissimilarity of associations due to species turnover (β_{ST}). PERMANOVA was used to evaluate the effect of cropping system on the topological distance between networks (D) and the dissimilarity of associations between networks (β_{WN}, β_{ST} and β_{OS}). The models included cropping system, P and their interactions as fixed effects. The permutations (n = 999) were constrained within blocks. Consensus networks were built to identify robust associations

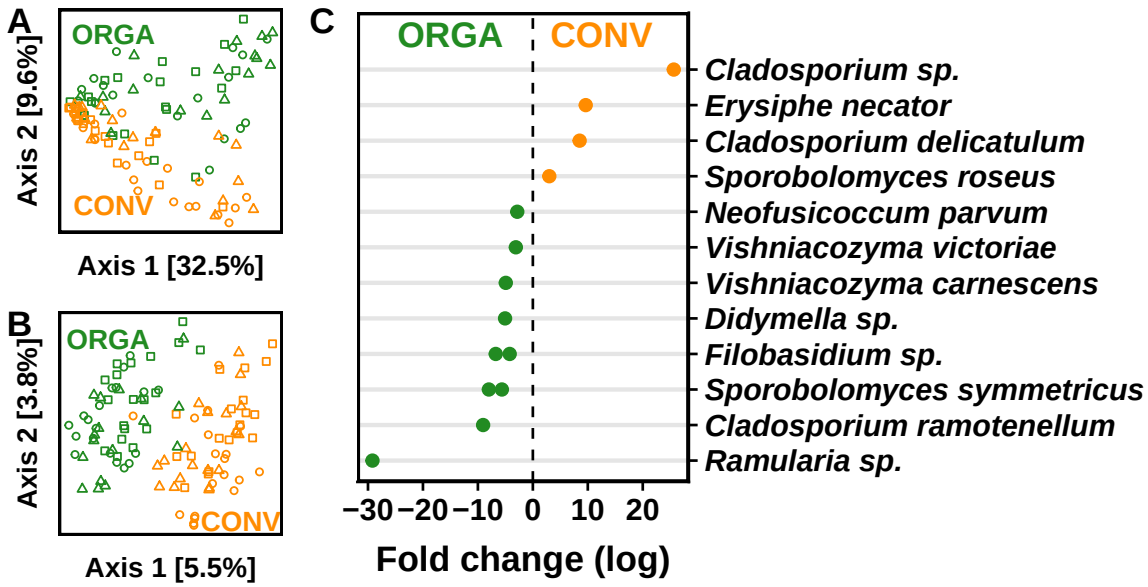


Figure 3 – Effect of cropping system – conventional (CONV) versus organic (ORGA) –on the β-diversity metrics of grapevine foliar fungal communities. Principal coordinate analyses (PCoA) were used to represent dissimilarities in composition between samples, as assessed with the (A) quantitative and (B) binary Jaccard indices. The effect of cropping system on both β-diversity metrics was significant (Table 2). Green circles, squares and triangles correspond to samples collected in the ORGA1, ORGA2 and ORGA3 plots, respectively. Orange circles, squares and triangles correspond to the CONV1, CONV2 and CONV3 plots, respectively (Figure 1). (C) Log-transformed ratio of ASV relative abundance in CONV plots over that in ORGA plots, for 14 ASVs identified as differentially abundant between cropping systems by DESeq2 analysis followed by Benjamini-Hochberg adjustment (Love et al., 2014).

that could indicate biotic interactions between fungal strains.

4 Results

The foliar fungal communities were dominated by Ascomycota in both ORGA (87.2% of sequences) and CONV (96.8%) plots. They were largely colonized by an ASV assigned to the Aureobasidium genus. More than half the sequences belonged to this ASV (Table 1), whatever the cropping system. The causal agent of grapevine powdery mildew, *Erysiphe necator*, was among the 10 most abundant fungal species. The proportion of sequences assigned to this pathogen species was higher in CONV than in ORGA plots (Table 1), consistent with the visual records of disease symptoms (Table S5).

4.1 Effect of cropping system on community α- and β-diversity

Fungal community richness, diversity and evenness were significantly higher in ORGA than CONV plots (Wald χ^2 = 4.74, p=0.029; Wald χ^2 = 8.28, p=0.004; Wald χ^2 = 12.88, p < 0.001, respectively; Figure 2). The composition of foliar fungal communities differed significantly between cropping systems (Table 2), in terms of both the relative abundance (Figure 3A) and presence-absence of ASVs (Figure 3B). DESeq2 analysis revealed that four ASVs, including the fungal pathogen *Erysiphe necator*, were significantly more abundant in CONV plots, whereas 10 other ASVs, including several yeast species (from the genera *Vishniacozyma*, *Sporobolomyces* and *Filobasidium*), were significantly more abundant in ORGA plots (Figure 3C). Principal coordinate analysis (PCoA) with the Jaccard quantitative index (Figure 3A) revealed large differences in the relative abundances of ASVs between samples within a cropping system. The first axis of the PCoA accounted for 32.5% of the variance in community composition but did not discriminate between cropping systems. It was significantly correlated with the relative abundance of the dominant ASV (assigned to the *Aureobasidium* genus) (Spearman ρ = 0.96; $p < 0.001$).

4.2 Effect of cropping system on network α- and β-properties

In total, we obtained sixty fungal association networks, corresponding to the ten values of P (Figure 4A) for each of the six plots (Figure 1). None of the six network α-properties (Table S7) differed between cropping systems (Table S8), but all were significantly correlated with the percentage P of ASVs included in the network (Table S9). The total number of links and the mean node degree increased with P, whereas the number of connected components, network diameter, connectance and the proportion of negative links decreased (Figure 4B; Table S9). Similarly, the topological distance between networks did not differ between cropping systems, but was influenced by P (Table 3). By contrast, cropping system had a significant effect on the overall dissimilarity of associations (β_{WN}) and the dissimilarity of associations between shared ASVs (β_{OS}; Table 3 and Figure 5). However, fungal association networks varied considerably between plots within a cropping system (Figure 4A and Figure 6). When all ASVs were used for network construction (P = 100%), only two associations were common to all three network

Table 2 – Effect of cropping system – conventional versus organic – on the β-diversity metrics of grapevine foliar fungal communities. Dissimilarities in community composition between samples were assessed with both the quantitative and binary Jaccard indices. The effects of sequencing depth (SD, log-transformed) and cropping system on composition dissimilarities between communities were evaluated in permutational analyses of variance (PERMANOVA). The number of permutations was set to 999 and permutations were constrained by block.

Dissimilarity Index	**PERMANOVA**				
Quantitative Jaccard	**Variable**	**Df**	**F**	**R2**	**Pr(>F)**
	Sequencing_Depth (SD)	1	4.49	0.04	<0.01
	Cropping_System (CS)	1	9.42	0.08	<0.01
	SD × CS	1	1.17	0.01	0.26
	Residuals	108		0.854	
	Total	111		1	
Binary Jaccard	**Variable**	**Df**	**F**	**R2**	**Pr(>F)**
	Sequencing_Depth (SD)	1	1.05	0.01	0.29
	Cropping_System (CS)	1	5.21	0.05	<0.01
	SD × CS	1	1.03	0.01	0.41
	Residuals	108		0.937	
	Total	111		1	

replicates of the ORGA system (Figure 6A ; Figure S8A) and only six were common to all three network replicates of the CONV system (Figure 6B ; Figure S8B). The two associations replicated in the ORGA system were negative associations between the dominant ASV (assigned to the genus *Aureobasidium*) and *Cladosporium ramotenellum*, and between *Vishniacozyma victoriae* and *Neofusicoccum parvum* (Figure S8A). Five of the six associations replicated in the CONV system were positive, the remaining negative association being that between the dominant ASV (assigned to the genus *Aureobasidium*) and *Epicoccum nigrum* (Figure S8B). No association common to all six networks was identified (Figure 6C).

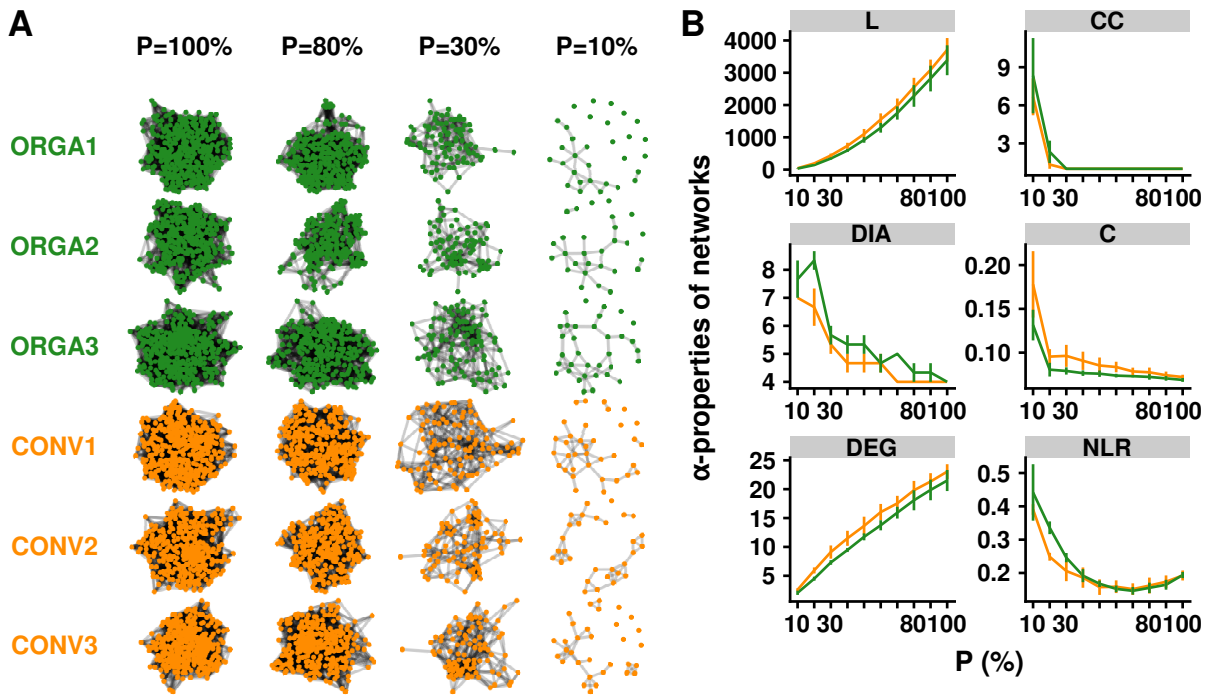


FIGURE 4 – Effect of cropping system – conventional (CONV) versus organic (ORGA) – on the α-properties of grapevine foliar fungal networks. (A) Association networks inferred from fungal metabarcoding data with SparCC (FRIEDMAN et ALM, 2012). A total of 60 networks were inferred, corresponding to 2 cropping systems × 3 replicates (blocks) ×10P values, with P the percentage of most abundant ASVs used for network inference. Only four values of P are shown on the figure. (B) Variations in network α-properties. The following properties (Table S7) were calculated for each network : the number of links (L) and connected components (CC), the network diameter (DIA) and connectance (C) and the mean degree (DEG) and negative link ratio (NLR). The percentage P of ASVs used for network reconstruction had a significant influence on all properties (Table S9), whereas cropping system did not (Table S8).

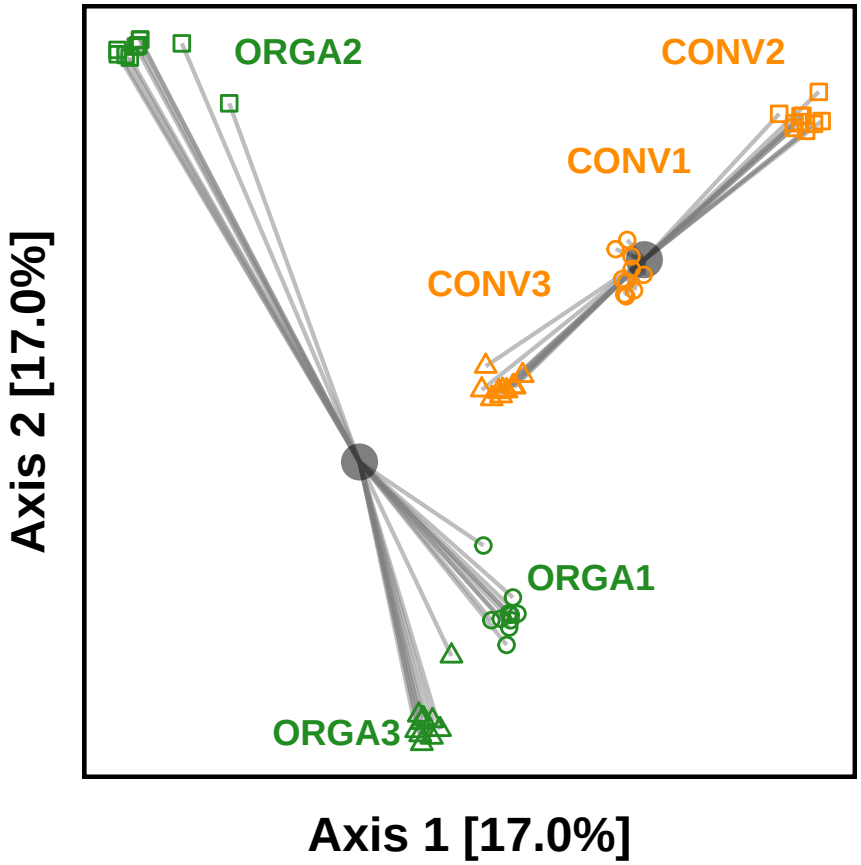


FIGURE 5 – Effect of cropping system – conventional (CONV) versus organic (ORGA) – on the β-properties of grapevine foliar fungal networks. Principal coordinate analysis (PCoA) representing dissimilarities between networks, measured with the β_{OS} index (POISOT et al., 2012) calculated with the binary Jaccard index. β_{OS} measures the dissimilarity between two networks in terms of the presence-absence of associations between shared ASVs. The centroids for each cropping system are represented by gray circles. The effect of cropping system on β_{OS} was significant (Table 3).

5 Discussion

Plant-associated microbial interaction networks protect plants against disease (Hassani et al., 2018; Kemen, 2014). Their responses to environmental changes, such as changes in agricultural practices, must be monitored in real time to better forecast disease risk. The concept of Next-Generation Biomonitoring (NGB) proposes that this could be done via the automatic reconstruction of ecological networks from metabarcoding data (Bohan et al., 2017). Therefore, we investigated whether automatically-reconstructed microbial networks have robust responses to agricultural practices (conventional versus organic agriculture). Using grapevine foliar fungal communities as study system, we found a strong footprint of the agricultural practice on the metabarcoding data, when analyzed at the community level. The richness, diversity and evenness of fungal communities were significantly higher in organic than conventional vineyards, consistent with the recent findings of Kernaghan et al. (2017) (but see Castañeda et al., 2018). The cropping system also significantly affected the composition of grapevine foliar fungal communities, as reported in previous studies (Castañeda et al., 2018; Kernaghan et al., 2017; Pancher et al., 2012; Schmid et al., 2011; Varanda et al., 2016). For instance, *Erysiphe necator*, the causal agent of grapevine powdery mildew, was significantly more abundant in conventional than in organic plots. The cause for such contrast in the pathogen abundance remains unknown given the absence of difference in dose and number of treatments. These results are, however, consistent with visual assessments of disease symptoms, indicating that, despite their numerous biases, metabarcoding data do contain some quantitative information useful for monitoring plant disease development (Jakuschkin et al., 2016; Makiola et al., 2018; Sapkota et al., 2015). Several yeast strains, assigned to the genera *Vishniacozyma*, *Sporobolomyces* and *Filobasidium*, were significantly more abundant in organic plots. These yeast genera are frequently detected on leaf surfaces due to their tolerance of irradiation and they might influence plant growth by producing plant hormone-like metabolites (Kemler et al., 2017). In addition, *Vishniacozyma victoriae* (ex *Cryptococcus victoriae*) was reported as a biocontrol agent on postharvest diseases (Lutz et al., 2013). Others yeasts possess wanted features of such agents like killer activities for some *Sporobolomyces* yeasts (Klassen et al., 2017). The yeasts *Vishniacozyma victoriae* and *Filobasidium wieringae* (ex *Cryptococcus wieringae*) were also reported as moderate antagonists of several filamentous fungi (Hilber-Bodmer et al., 2017). Future research should investigate the interactions between these yeast species and grapevine foliar pathogens, including powdery mildew.

Contrary to our expectations, microbial networks were less sensitive to changes in agricultural practices than microbial communities, suggesting that NGB should not only consider network-level properties but also community-level properties. The α-properties of microbial networks (i.e. whole-network metrics; Pellissier et al., 2018) did not differ between cropping systems. Their variation was correlated with the network inference parameter, P, a methodological parameter corresponding to percentages of the most abundant taxa included in the network. Only two β-properties of microbial networks differed significantly between cropping systems, revealing a difference in microbial associations between organic and conventional vineyards. These differences remained significant when network pairwise comparisons were based on shared taxa only, suggesting that the differences between organic and conventional networks were due to re-associations of fungal taxa rather than a turnover of taxa. The

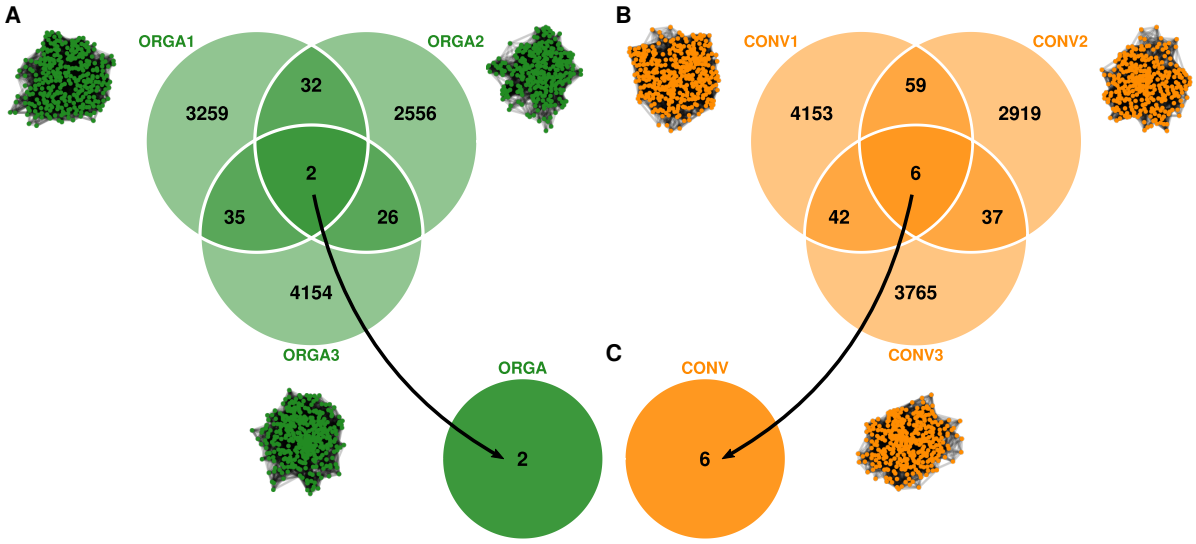


FIGURE 6 – Venn diagrams showing the number of fungal associations common to network replicates. (A) Associations common to the three network replicates inferred for the organic cropping system (ORGA1, ORGA2, ORGA3) and (B) the three network replicates inferred for the conventional cropping system (CONV1, CONV2, CONV3), regardless of the sign of the association, in the situation in which all ASVs were used for network construction (P = 100%). (C) Associations common to the six networks.

functional redundancy of fungal taxa might account for the difference between networks. Taxonomically different communities, involving different interactions between members, may have similar functions (LOUCA et al., 2016b) and protect plant health in a similar way. However, differences between networks might also be due to methodological hurdles that are discussed below and will have to be overcome in future NGB developments.

In our study, replicate microbial networks for the same cropping system had very few associations in common, revealing that the microbial networks we inferred from metabarcoding data lacked replicability. Three methodological parameters might account for this result. First, each network was built from 20 samples. Increasing the number of samples per network, to at least 25 (BERRY et WIDDER, 2014), might improve the replicability of microbial networks. Second, network nodes were amplicon sequence variants (ASVs) of the fungal ITS1 region, sequenced on an Illumina MiSeq platform. Despite their fine-scale taxonomic resolution (CALLAHAN et al., 2017), ASVs may group together fungal strains with different interaction traits (MCLAREN et CALLAHAN, 2018). A finer taxonomic resolution, achieved through the third-generation sequencing of the full ITS region (NILSSON et al., 2019), might improve the reliability of microbial networks (KENNEDY et al., 2018). Third, our method of network reconstruction did not account for spatial variations in environmental conditions (i.e. microclimate or leaf traits) that could compromise inferences (ARMITAGE et JONES, 2019). We did not measure environmental variations in our experimental system because the sampling design limited such variations. The vineyard plots were adjacent to each other and planted with grapevine clones. Moreover, we collected all leaves in less than two hours and controlled for the position of the sampled leaf on the vine. Our results suggest that this type of control might not be sufficient. Future developments of NGB approaches will have to include a higher number of samples per network, long-read sequencing and methods of network inference allowing the integration of environmental covariates (CHIQUET et al., 2017 ; CHIQUET et al., 2019 ; COUGOUL et al., 2019b ;

Ovaskainen et al., 2017 ; Tackmann et al., 2019).

Finally, our results show that consensus networks, built from several network replicates, can generate relevant hypotheses concerning microbial interactions. In our study, two negative associations were found to be common to all three replicates of the organic system. These negative associations were relevant according to current ecological knowledge. They involved two plant pathogens, *Neofusicoccum parvum* (commonly associated with grapevine trunk diseases ; Bruez et al., 2014) and *Cladosporium ramotenellum* (commonly associated with brown rot ; Swett et al., 2016), and two other species known to have antagonistic effects on plant pathogens, *Vishniacozyma victoriae* and *Aureobasidium sp.* (Lutz et al., 2013 ; Pertot et al., 2017). These results suggest that network inference is a promising tool to generate hypotheses that once tested will allow us to better understand disease regulation and perhaps discover biocontrol agent candidates (Poudel et al., 2016). Future research will however first have to determine whether it is better to build one big network with all samples or to merge several networks each based on a subset of the samples. In any case, confidence estimates of the links will have to be computed (Röttjers et Faust, 2018).

Despite these promising findings, this study indicates that the properties of NGS-based microbial networks cannot be used yet for monitoring the disease regulation services provided by the microbiota and that NGS-based microbial networks should be interpreted with caution (Carr et al., 2019). The replicated microbial association networks inferred from metabarcoding data were highly variable within each set of environmental conditions and generated only a few robust hypotheses concerning interactions between fungi, limiting for now their use to monitor the barrier effect of microbial interactions against pathogens. By contrast, community-level metrics revealed clear-cut changes in the plant microbiota in response to environmental change and reflected the disease status of the plant. Moreover, they were more statistically powerful than network-level metrics, because many samples were required to infer each microbial network replicate. Replicability of microbial networks is however likely to be improved in the near future, notably through developments in third-generation sequencing techniques and network inference methods, so microbial networks still hold promises for a reliable biomonitoring options (Karimi et al., 2017). Community-level data and network-level data, both analyzed using machine-learning approaches (e.g. Cordier et al., 2018a), could even offer complementary insights into the ecosystem services provided by the plant microbiota.

Acknowledgements

We thank Lucile Muneret, Andreas Makiola and all members of the ANR NGB Consortium (ANR-17-CE32-0011) for helpful scientific exchanges on next-generation biomonitoring and for their comments on the manuscript. We also thank Gregory Gambetta, Guilherme Martins, Frédéric Barraquand, Isabelle Lesur and Adrien Rush for useful comments on preliminary results. We thank the Genotoul sequencing facility (Get-PlaGe) for sequencing and the Genotoul bioinformatics facility (Bioinfo Genotoul) for providing computing and storage resources. We also thank Julie Sappa from Alex Edelman & Associates for English language revision of the first version of the manuscript. We thank the INRA MEM metaprogram (Meta-Omics of Microbial Ecosystems) for financial and scientific support. Sequencing was funded by the INRA MEM MetaBAR project and bioinformatic and statistical analyses were performed as part of the INRA MEM Learn-biocontrol project. Additional funding was received from the LABEX COTE (ANR-10-LABX-45), the LABEX CEBA (ANR-10-LABX-25-01) and INRA EcoServ metaprogram on ecosystem services (IBISC project) and the Aquitaine Region (Athene project,

Table 3 – Effect of cropping system – conventional versus organic – on the β-properties of grapevine foliar fungal networks. The D index quantifies the topological dissimilarity between networks (Schieber et al., 2017) whereas the other three metrics (β_{WN}, β_{OS} and β_{ST}), which were calculated with the binary Jaccard index, quantify differences in associations between networks (Poisot et al., 2012). The effect of the percentage P of the most abundant ASVs used for network inference, and the effect of cropping system on the dissimilarities between networks were evaluated in permutational analyses of variance (PERMANOVA). The number of permutations was set to 999 and permutations were constrained by block.

Dissimilarity Index	**PERMANOVA**				
Topological dissimilarity (Schieber's D)	**Variable**	**Df**	**F**	**R2**	**Pr(>F)**
	Percent_ASV (P)	1	57.75	0.50	<0.01
	Cropping_System (CS)	1	1.72	0.01	0.19
	P × CS	1	0.65	0.01	0.51
	Residuals	56		0.48	
	Total	59		1	
Overall dissimilarity of associations (β_{WN})	**Variable**	**Df**	**F**	**R2**	**Pr(>F)**
	Percent_ASV (P)	1	2.41	0.04	<0.01
	Cropping_System (CS)	1	5.0	0.08	<0.01
	P × CS	1	2.21	0.03	<0.01
	Residuals	56		0.85	
	Total	59		1	
Dissimilarity of associations between shared ASVs (β_{OS})	**Variable**	**Df**	**F**	**R2**	**Pr(>F)**
	Percent_ASV (P)	1	0.53	0.01	0.61
	Cropping_System (CS)	1	11.07	0.16	<0.01
	P × CS	1	0.56	0.01	0.57
	Residuals	56		0.798	
	Total	59		1	
Dissimilarity of associations due to ASV turnover (β_{ST})	**Variable**	**Df**	**F**	**R2**	**Pr(>F)**
	Percent_ASV (P)	1	1.30	0.02	<0.01
	Cropping_System (CS)	1	0.27	<0.01	1.00
	P × CS	1	1.30	0.02	<0.01
	Residuals	56		0.95	
	Total	59		1	

n°2016-1R20301-00007218). CP's PhD grant was funded by the INRA and Bordeaux Sciences Agro (BSA). The management of the experimental site was partly funded by the AFB (French Agency for Biodiversity) within the DEPHY network.

Data Accessibility Statement

The raw sequence data were deposited in Dataverse and are available in the FASTQ format at https://doi.org/10.15454/3DPFNJ while the filtered ASV table is available at https://doi.org/10.15454/WOICSE. The code is available as an archive at https://doi.org/10.15454/NSHUAQ.

Author Contributions

CP performed the bioinformatic and statistical analyses and wrote the first draft of the manuscript. JV and CV performed the sampling and processed the leaf samples. LD managed the sampling site and provided data on phytosanitary treatments and disease symptoms. JV performed the DNA extractions and amplifications. MB coordinated the creation of the fungal mock community and the sequencing of all samples. CV conceived the study, supervised the analyses and made a major contribution to the writing of the final draft. All authors discussed the preliminary version of the results and revised the manuscript.

6 Supplementary information

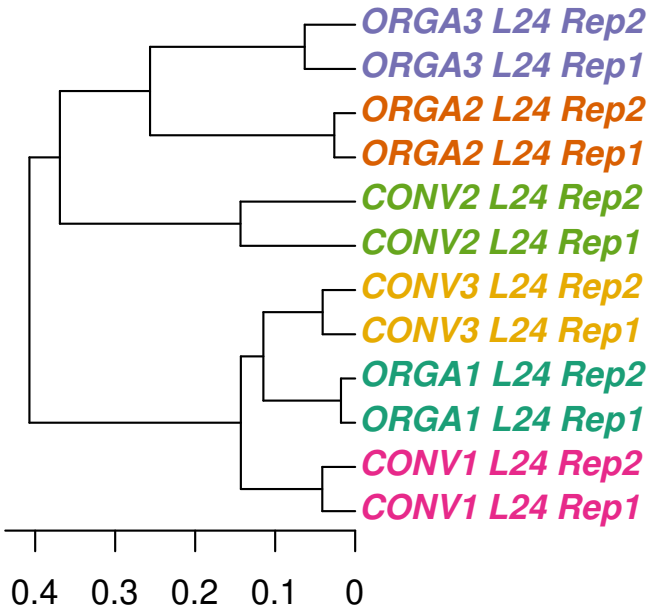


Figure 7 – Dendrogram plot of compositional dissimilarities between technical replicates for sequencing. Technical replicates were created by splitting six lots of PCR products in half and sequencing the two halves independently. The PCR products used were those corresponding to leaf 24 (L24) of the six plots studied (ORGA1, ORGA2, ORGA3, CONV1, CONV2, CONV3 ; see Figure 5). Compositional dissimilarities between samples were computed with the binary Jaccard index. The dendrogram was built using a hierarchical clustering algorithm (complete linkage method). Compositional dissimilarities between the two technical replicates of the same sample were significantly smaller than the dissimilarities among samples (PERMANOVA : $F = 39.98$; $R2 = 0.97$; $p = 0.001$).

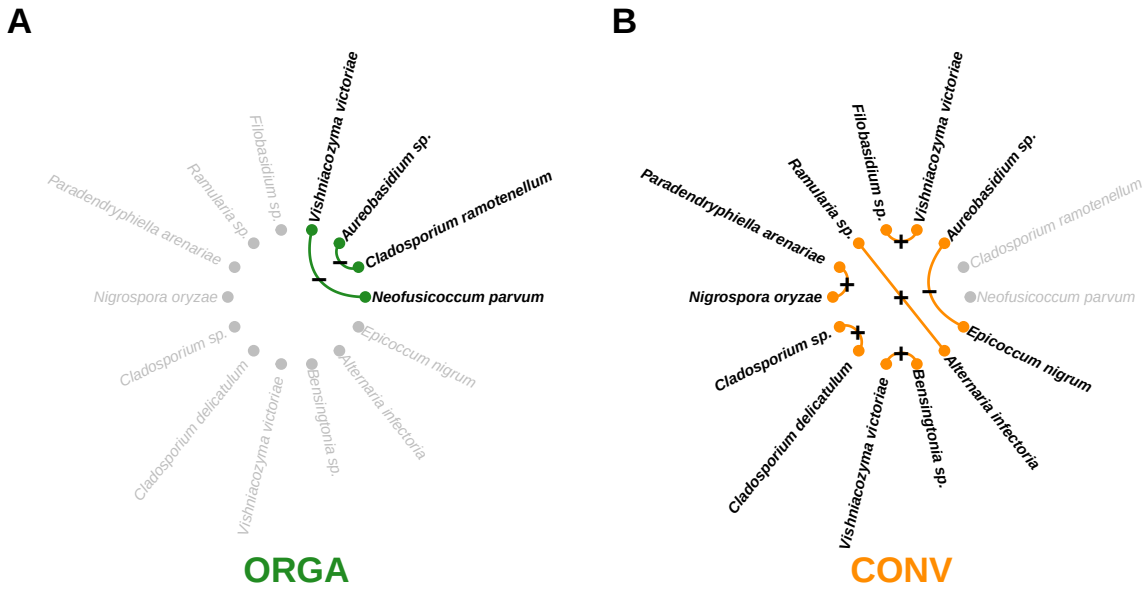


Figure 8 – Consensus fungal networks for (A) the organic (ORGA) and (B) the conventional (CONV) cropping systems. Network nodes represent fungal ASVs and links represent significant positive (+) or negative (-) associations common to the three network replicates (Figures 6A and 6B, respectively). The fungal ASVs absent from a network are indicated in gray.

Table 4 – List of phytosanitary products and active ingredients applied in the year of the sampling campaign, together with the normalized dose or the treatment frequency index. PM = powdery mildew (caused by the fungal pathogen *Erysiphe necator*) and DM = downy mildew (caused by the oomycete pathogen *Plasmopara viticola*). Leaf sampling was performed on September 10, 2015 (more than one month after the last phytosanitary treatment and a couple of hours before grape harvest). The treatment frequency index did not differ between cropping systems (ANOVA : df = 21 ; $F = 0.436$; $p = 0.516$).

Date	Cropping System	Fungicides	Active ingredients	Target disease	
				PM	DM
2015-04-30	ORGA	Heliocuivre©	Copper		0.145
2015-04-30	ORGA	Citrothiol DG©	Micronized sulfur	0.371	
2015-05-07	CONV	Chaoline©	Fosetyl aluminum + metirame		0.292
2015-05-07	CONV	Dynali©	Cyflufenamid + difenoconazole	0.289	
2015-05-13	ORGA	Heliocuivre©	Copper		0.167
2015-05-13	ORGA	Citrothiol DG©	Micronized sulfur	0.400	
2015-05-19	CONV	Cabrio Top©	Metirame-zinc + pyraclostrobin	0.500	
2015-05-28	ORGA	Citrothiol DG©	Micronized sulfur	0.800	
2015-05-28	ORGA	Bouillie Bordelaise RSR® Disperss® NC	Copper		0.533
2015-06-04	CONV	Vivando©	Metrafenone	0.833	
2015-06-04	CONV	Chaoline©	Fosetyl aluminum + metirame		0.708
2015-06-09	ORGA	Bouillie Bordelaise RSR® Disperss® NC	Copper		0.533
2015-06-09	ORGA	Citrothiol DG©	Micronized sulfur	0.600	
2015-06-25	ORGA	Citrothiol DG©	Micronized sulfur	0.600	
2015-06-25	CONV	Citrothiol DG©	Micronized sulfur	0.600	
2015-07-01	ORGA	Bouillie Bordelaise RSR® Disperss® NC	Copper		0.533
2015-07-01	CONV	Cabrio Top©	Metirame-zinc + pyraclostrobin	0.750	
2015-07-17	ORGA	Bouillie Bordelaise RSR® Disperss® NC	Copper		0.400
2015-07-17	ORGA	Heliocuivre©	Copper		0.083
2015-07-17	CONV	Bouillie Bordelaise RSR® Disperss® NC	Copper		0.400
2015-07-17	CONV	Heliocuivre©	Copper		0.083
2015-08-03	ORGA	Bouillie Bordelaise RSR® Disperss® NC	Copper		0.533
2015-08-03	CONV	Bouillie Bordelaise RSR® Disperss® NC	Copper		0.533

Table 5 – Effect of cropping system —conventional (CONV) versus organic (ORGA) — on the incidence and severity of foliar disease symptoms at harvest time (2015-09-07). Disease incidence is defined as the percentage of leaves displaying symptoms, whereas disease severity is defined as the percentage leaf damage. Symptom incidence and severity were estimated visually on 40 grapevines for each plot (40 × 3 per cropping system). The mean values are reported for each cropping system as a percentage. Wald χ^2 tests were used for comparisons after linear mixed model analysis with cropping system as a fixed effect and block as a random effect.

Disease		ORGA (%)	CONV (%)	χ^2	p-value
Downy Mildew	Incidence	0.749	0.688	0.57	0.450
	Severity	0.037	0.030	1.93	0.164
Powdery Mildew	Incidence	0.113	1.346	12.49	**<0.001**
	Severity	0.003	0.102	7.97	**0.005**
Black rot	Incidence	0.188	0.354	19.02	**<0.001**
	Severity	0.007	0.014	5.49	**0.019**

Table 6 – Primer pairs used to amplify the fungal ITS1 region

1st PCR with regular primers (bold)	
Forward	ITS1F : 5'-CTTGGTCATTTAGAGGAAGTAA-3'
Reverse	ITS2 : 5'-GCTGCGTTCTTCATCGATGC-3'
2nd nested PCR with pre-tagged primers (italics)	
Forward	ITS1F-pre-tag : 5'-CTTTCCCTACACGACGCTCTTCCGATCTCTTGGTCATTTAGAGGAAGTAA-3'
Reverse	ITS2-pre-tag : 5'-GGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCGTTCTTCATCGATGC-3'

Table 7 – List of network α-properties calculated in this study

Network metric	**Definition**	**Reference**
Number of links (L)	Total number of links	
Connectance (C)	Fraction of the total number of possible links actually realized	Coleman et Moré (1983)
Number of connected components (CC)	Number of groups of nodes connected together	Martinez (1992)
Diameter (DIA)	The longest of all the shortest paths between two nodes	Barabási et al. (2000)
Mean node degree (DEG)	Mean number of links per node	Martinez (1992)
Proportion of negative links (NLR)	Proportion of links for which the SparCC correlation is negative	Faust et al. (2015)

Table 8 – Effect of cropping system on the α–properties of fungal association networks. Properties (as defined in Table S 7) were compared between cropping systems for every value of the percentage P of the most abundant ASVs used for network inference. The U and p-values of Wilcoxon rank-sum tests are reported. The p-value is not available (NA) for situations in which property values were equal for all networks. The p-values after Benjamini-Hochberg adjustment are not reported because all were equal to one.

P (%)	L	CC	DIA	C	DEG	NLR
10	$U = 3$; $p = 0.658$	$U = 5$; $p = 1$	$U = 6$; $p = 0.505$	$U = 3$; $p = 0.663$	$U = 2$; $p = 0.383$	$U = 5$; $p = 1$
20	$U = 1$; $p = 0.19$	$U = 6.5$; $p = 0.48$	$U = 8$; $p = 0.157$	$U = 2$; $p = 0.383$	$U = 1$; $p = 0.19$	$U = 9$; $p = 0.081$
30	$U = 1$; $p = 0.19$	$U = 4.5$; $p =$ NA	$U = 6$; $p = 0.619$	$U = 1$; $p = 0.19$	$U = 1$; $p = 0.19$	$U = 6$; $p = 0.663$
40	$U = 1$; $p = 0.19$	$U = 4.5$; $p =$ NA	$U = 7$; $p = 0.302$	$U = 2$; $p = 0.383$	$U = 1$; $p = 0.19$	$U = 5$; $p = 1$
50	$U = 3$; $p = 0.663$	$U = 4.5$; $p =$ NA	$U = 7$; $p = 0.302$	$U = 3$; $p = 0.663$	$U = 3$; $p = 0.663$	$U = 6$; $p = 0.663$
60	$U = 2$; $p = 0.383$	$U = 4.5$; $p =$ NA	$U = 4.5$; $p = 1$	$U = 2$; $p = 0.383$	$U = 1$; $p = 0.19$	$U = 5$; $p = 1$
70	$U = 3$; $p = 0.663$	$U = 4.5$; $p =$ NA	$U = 9$; $p = 0.047$	$U = 2$; $p = 0.383$	$U = 2$; $p = 0.383$	$U = 5$; $p = 1$
80	$U = 3$; $p = 0.663$	$U = 4.5$; $p =$ NA	$U = 6$; $p = 0.505$	$U = 4$; $p = 1$	$U = 3$; $p = 0.663$	$U = 4$; $p = 1$
90	$U = 4$; $p = 1$	$U = 4.5$; $p =$ NA	$U = 6$; $p = 0.505$	$U = 4$; $p = 1$	$U = 3$; $p = 0.663$	$U = 3$; $p = 0.663$
100	$U = 3$; $p = 0.663$	$U = 4.5$; $p =$ NA	$U = 4.5$; $p =$ NA	$U = 2$; $p = 0.383$	$U = 4$; $p = 1$	$U = 4$; $p = 1$

Table 9 – Effect of the percentage P of the most abundant ASVs used for network inference on the α–-properties of fungal association networks. Spearman's correlation coefficient and the results of Spearman's rank correlation tests are reported for each network property. The \textit{p}-values are reported after Benjamini-Hochberg adjustment.

Property	Correlation (ρ)	S	p-value
L	0.98	862	<0.001
CC	-0.61	57987	<0.001
DIA	-0.85	66635	<0.001
C	-0.67	60144	<0.001
DEG	0.95	1635	<0.001
NLR	-0.56	56116	<0.001

IV

Les associations des réseaux microbiens offrent des indications pertinentes sur les pathobiomes des plantes

1 Abstract

Microbial association networks give relevant insights into plant pathobiomes

Charlie Pauvert[1], Tania Fort[1], Agnès Calonnec[2], Julie Faivre d'Arcier[1], Emilie Chancerel[1], Marie Massot[1], David A. Bohan[3], Julien Chiquet[4], Stéphane Robin[4], Jessica Vallance[2], Corinne Vacher[1*]

1– BIOGECO, INRA, Univ. Bordeaux, 33615 Pessac, France. 2 – INRA, UMR 1065 Santé et Agroécologie du Vignoble, ISVV, F-33882 Villenave d'Ornon, France. 2– Université de Bordeaux, Bordeaux Sciences Agro, UMR 1065 SAVE, F-33175 Gradignan, France. 3– Agroécologie, AgroSup Dijon, INRA, Université Bourgogne, Université Bourgogne Franche-Comté, Dijon, France. 4– UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005 Paris, France.

* Correspondence : Dr. Corinne Vacher (corinne.vacher@inra.fr)

Plant growth and health depend on their associations with a large number of microorganisms that interact with each other. The networks of interactions between plant pathogens and other microorganisms, termed pathobiomes, regulate disease development. Our aim was to investigate whether microbial association networks inferred form metabarcoding data and environmental covariates contain signals of ecological interactions, despite methodological biases associated to sampling and molecular biology. Grapevine (*Vitis vinifera*) and the causal agent of powdery mildew (*Erysiphe necator*) were used as a model system. We used Poisson Lognormal Models to highlight direct associations among fungal species on grapevine leaves, and tested some these associations using co-cultures and text-mining. Our results suggested that the abundance of *E. necator* covaried with that of 23 other fungal species, forming its putative pathobiome. None of these associations were known as interactions in the literature, but one of them was confirmed by our experiments. The yeast *Buckleyzyma aurantiaca* altered pathogen growth rate and morphology, in line with the inferred negative association. Another association, involving two yeast species, was also supported by co-cultures. Moreover, our results showed that the foliar fungal community was robust to invasion by *E. necator* but significantly influenced by the cropping system. This latter could hence be used to promote beneficial microbial interactions highlighted by microbial networks. Together, these findings indicate that microbial networks, when properly inferred, provide plausible hypotheses of ecological interactions that can be useful to develop microbiome-based strategies for crop protection.

Keywords : Anna Karenina principle, Biocontrol, Community ecology, Microbial networks, Network inference, Pathobiome.

2 Introduction

Plant growth and health depend on their associations with a large number of microorganisms that interact with each other (Hassani et al., 2018 ; Vannier et al., 2019). Among all these microorganisms, those that interact with pathogens and regulate diseases are particularly important and have recently been defined as the plant's pathobiome (Bass et al., 2019 ; Brader et al., 2017 ; Vayssier-Taussat et al., 2014). Some pathobiome members form a barrier that limit pathogen development through direct antagonistic interactions (Arnold et al., 2003 ; Durán et al., 2018 ; Kemen, 2014 ; Li et al., 2019), while others can prime the plant immune system (Hacquard et al., 2017 ; Lee et al., 2017 ; Vogel et al., 2016). In turn, successful pathogen invasion can disrupt the plant microbiota, triggering, in some cases, more heterogeneity (Zaneveld et al., 2017). This increase in heterogeneity among infected microbiotas is often termed the Anna Karenina principle, based on the first sentence of Leo Tolstoy's book : "Happy families are all alike ; every unhappy family is unhappy in its own way". In microbiology, the Anna Karenina principle means : "all healthy microbiomes are similar ; each dysbiotic microbiome is dysbiotic in its own way" (Zaneveld et al., 2017).

Deciphering microbial interactions within pathobiomes, and determining what environmental factors shape those interactions, is a step towards plant health protection. A better understanding of plant pathobiomes will allow us to reduce agriculture reliance on chemical pesticides, through the discovery of novel biocontrol agents (Pennington et al., 2018) and cultural practices fostering the protective microbiota (Hartman et al., 2018). To reach this aim, research at the interface between plant pathology, microbial community ecology, culturomics, metagenomics and big data analytics is needed. Metabarcoding approaches can for instance be used to describe culturable and unculturable microorganisms associated with plants (Abdelfattah et al., 2018 ; Nilsson et al., 2019), provided that rigorous methods are used from sampling to bioinformatics (Pauvert et al., 2019a ; Zinger et al., 2019). Network inference methods can then be used to reconstruct microbial association networks from microbial community data (Faust et Raes, 2012 ; Layeghifard et al., 2017 ; Vacher et al., 2016b). These networks, in which nodes correspond to microbial taxa and links to statistical associations between their sequence counts, can be interpreted as hypotheses of microbial interactions (Carr et al., 2019 ; Jakuschkin et al., 2016 ; Poudel et al., 2016). Microbiology experiments can then be performed to validate interaction hypotheses (*e.g.* Biswas et al., 2016 ; Das et al., 2018 ; Durán et al., 2018 ; Lima-Mendez et al., 2015 ; Tipton et al., 2018 ; Wang et al., 2017).

However, the application of network inference methods to pathobiome research is still in its infancy. Several methodological issues must be overcome to generate robust hypotheses of microbial interactions from metabarcoding data (Pauvert et al., 2019b), because statistical associations between sequence counts do not directly reflect ecological interactions (*e.g.* competition, parasitism) between microorganisms (Derocles et al., 2018 ; Röttjers et Faust, 2018 ; Weiss et al., 2016). For instance, the compositional nature of metabarcoding data induces statistical associations between sequence counts that are not related to any ecological process (Friedman et Alm, 2012 ; Gloor et al., 2017). Moreover, environmental filtering can generate statistical associations between microbial taxa abundances, that are not triggered by ecological interactions but environmental variations (Berry et Widder, 2014 ; Derocles

et al., 2018 ; Röttjers et Faust, 2018 ; Vacher et al., 2016b). Several recent methods of network inference deal with these two issues : HSMC (Ovaskainen et al., 2017), PLN (Chiquet et al., 2017 ; Chiquet et al., 2019), FlashWeave (Tackmann et al., 2019) and MAGMA (Cougoul et al., 2019b). Their relevance to pathobiome research remains to be demonstrated.

The aim of this study was to deepen our knowledge of plant pathobiomes by testing the following hypotheses : (H1) successful infection events destabilize plant-associated microbial communities and increase their heterogeneity (Anna Karenina principle), (H2) interactions among microorganisms within pathobiomes can be detected by inferring microbial networks from metabarcoding data and environmental covariates, and (H3) cropping systems influence the abundance of microorganisms forming pathobiomes. These hypotheses were tested using Grapevine (*Vitis vinifera*) and the causal agent of powdery mildew (*Erysiphe necator* ; Armijo et al., 2016 ; Gadoury et al., 2012), as a model system. To maximize the diversity of grapevine microbial communities and thus the possibility of discovering microbial antagonists of this fungal pathogen, the experiment was conducted in an untreated vineyard. Powdery mildew was inoculated on several vines at the beginning of the growing season, to ensure the production of abundant and evenly distributed inoculum. Under these conditions, spatial variations in the level of secondary infections were likely to be caused by variations in the barrier effect of the pathobiome, rather than variations in the abundance of the inoculum. The fungal component of the pathobiome of *Erysiphe necator* was then analyzed by combining metabarcoding approaches, network inference methods and co-culture experiments. The habitat of the fungal species belonging to the pathobiome was characterized by analyzing fungal communities of bark, ground cover and upper soil, in addition to that of vine leaves. Their response to inter-row cover cropping was also assessed.

3 Materials and methods

3.1 Study site and sampling design

The study was conducted in an experimental vineyard located near Bordeaux (INRA, Villenave d'Ornon, France ; 44°47'24.0"N 0°34'33.6"W ; Figure 1A). This experimental vineyard was planted in 1991 with *Vitis vinifera* L. cv. Merlot grafted onto 101-14 rootstock, at a density of 5350 vines ha^{-1} (1.7m×1.1m). The study was performed on a sub-plot of 5 almost untreated rows (Figure 1B) where, since 2006, only few chemical treatments had been applied in case of strong epidemics (Table S4). In this sub-plot, two cropping systems (Figure 1C) were compared : (i) perennial cover crop in the inter-rows (*i.e.* cover crop : CC) and (ii) chemical weed control with glyphosate (*i.e.* no cover crop : NCC). Eight experimental units were considered across the sub-plot : four units in NCC areas and four units in CC areas, an experimental unit being defined as a group of 5 adjacent vines in the same row (Figure 1D).

In 2016, each experimental unit was artificially inoculated with the powdery mildew agent *Erysiphe necator* to ensure a uniform intensity of primary infection. Inoculation was performed on April 22, on the central vine of each experimental unit (Figure 1E). The inoculum consisted in a monoconidial isolate collected in a greenhouse in INRA precincts (Villenave d'Ornon, Bordeaux, France) that year (strain S16). Leaves were then regularly checked for powdery mildew colonies. The first symptoms appeared on May 5 for the inoculated leaves, and around May 20 for other leaves. Leaf age was monitored in all experimental units based on biweekly records of newly grown leaves. In addition, a weather station located on the edge of the vineyard allowed us to monitor local variations in air temperature and humidity throughout the experiment.

Grapevine leaves were sampled on three dates : 40 (June 1), 62 (June 23) and 77 (July 5) days post-inoculation (Figure 1F). Four infected (I) and four non-infected (NI) leaves were collected in each experimental unit on each date (Figure 1G). A total of 192 leaves, corresponding to 8 leaves×3 dates×8 experimental units, were thus collected. Sampling dates were chosen according to current knowledge on powdery mildew cycle (Calonnec et al., 2009 ; Calonnec et al., 2006). All sampled leaves were approximately 20-day old and had received secondary inoculum before the onset of ontogenic resistance (21-28 days see Gadoury et al., 2012). The percentage of lower leaf surface covered with powdery mildew was visually assessed for each infected leaf collected. The position of each sampled leaf within the canopy was recorded as well as its distance to potential environmental sources of microorganisms. Three distances were measured : distance to the ground, distance to the cordons and distance to the artificially inoculated leaf on the central vine of the experimental unit.

Sampled leaves were placed in individual sterile plastic bags (Whirl-Pak®, USA) and taken to the lab in a cooler with ice. Foliar tissues were sampled on the day of collection in the sterile environment of a MICROBIO electric burner (MSEI, France). Tissues were collected from three tissue types of leaf blade : the healthy zones of non-infected (HNI) and infected (HI) leaves, and the center of disease spots of infected leaves (DI) (Figure 1E). For each leaf, two foliar discs

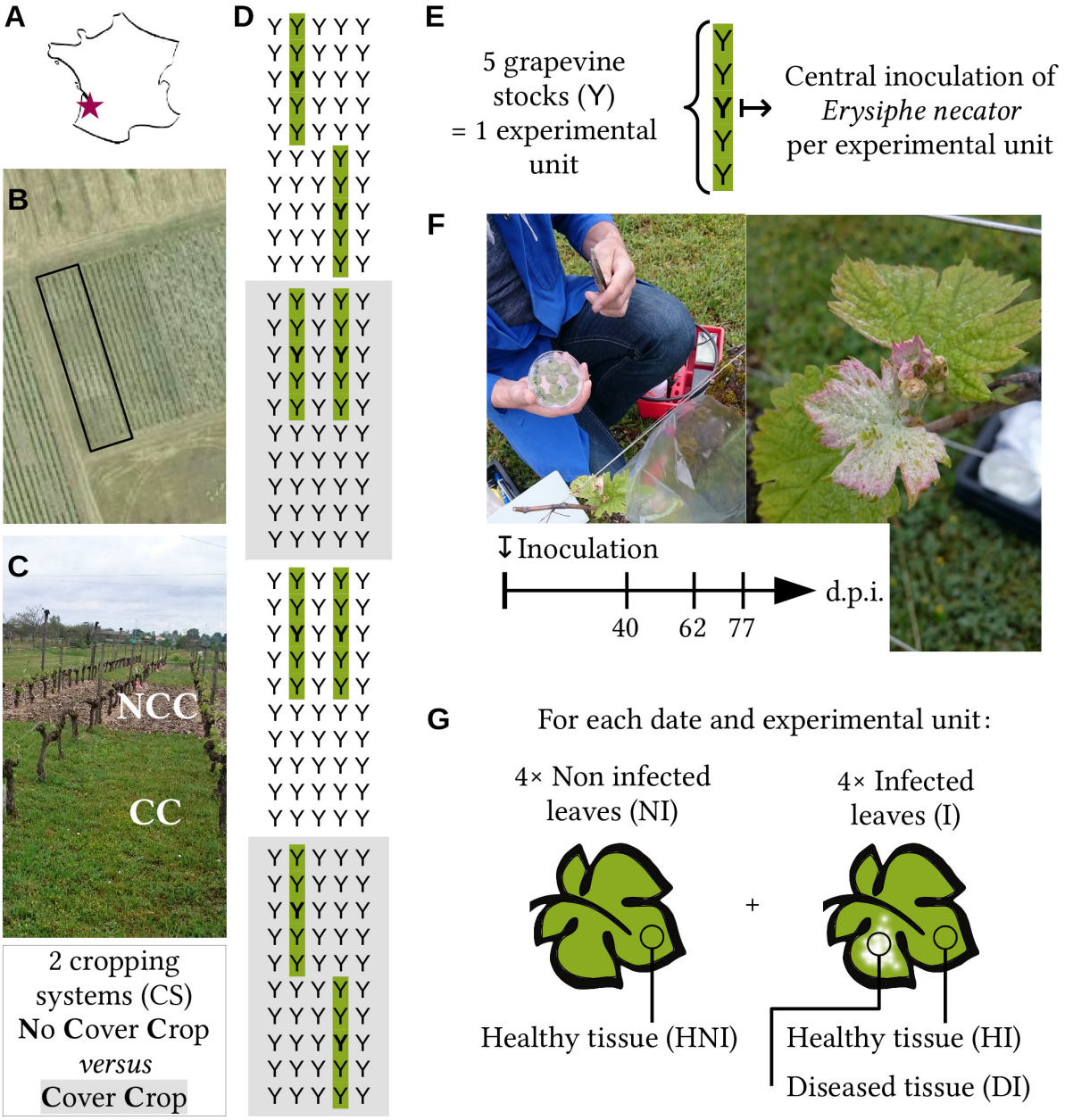


FIGURE 1 – Experimental design. The study took place in an experimental vineyard located near Bordeaux, France (A), in a sub-plot of 5 almost untreated vine rows (B) with two cropping systems differing in the presence or not of cover crop (CC *versus* NCC) in the inter-rows (C). Sampling occurred in eight experimental units each composed of five adjacent vines (underlined in green) (D). The causal agent of grape powdery mildew, *Erysiphe necato*r, was inoculated on the central vine of each experimental unit (E) and we performed three sampling campaigns 40, 62 and 77 days post inoculation (F). For each campaign, leaves with and without visible symptoms were collected to analyze healthy foliar tissues and foliar tissue from diseased spots (G).

of 6 mm diameter were collected in each zone and placed together in a collection microtube of a 96-well plate (QIAGEN), with two autoclaved glass beads. The processed leaves were kept on ice in individual closed plastic bags to avoid water loss. Leaf fresh weight and surface were then measured using the WinFOLIA®software (Regent Instrument, Canada). Dry weight was measured after drying the leaves in an oven at 65°C for 72h.

On the last date of sampling (July 5), additional environmental samples were collected in each experimental unit. Two old leaves (approximately 70 days old) were collected from the central vine of each experimental unit, close to the place where the inoculation had been performed, and placed in a sterile plastic bag (Whirl-Pak®, USA). The inoculated leaf was collected if it was still attached to the vine. Small pieces of bark (including mosses and lichens) were collected on three vines of each experimental unit and stored in 50mL sterile Falcon tubes. Fragments of ground cover (including upper soil) were also collected beneath three vines of each experimental unit and stored in 50mL sterile Falcon tubes.

Finally, the sub-plot phytosanitary status and canopy vigor were assessed on July 11. The severity (% of lower leaf surface infected) of four diseases, *i.e.* powdery mildew (*E. necator*), downy mildew (*Plasmopara viticola*), black rot (*Guignardia bidwellii*) and grape erineum mite (*Colomerus vitis*), was visually evaluated on 468 leaves randomly chosen within the 8 experimental units (12 leaves per vine). The vigor of the vines was defined as the product of the number of shoots per vine by the number of leaves on the longest shoot.

3.2 DNA extraction and Illumina MiSeq sequencing

Leaf discs were cold ground at 1500 rpm with the Geno/Grinder®for 30 s, then 1 min and 1 min again, with manual shaking between each grinding step. Plates were then centrifuged for 1 min at 6000 g. Total DNA was then extracted using the DNeasy Plant Mini Kit (Qiagen, USA) following the manufacturer's instructions, except that DNA extracts were eluted twice with 50 μl of elution buffer (10 mM Tris-Cl, 0.5 mM EDTA ; pH 9.0). The ITS1 region of the fungal ITS rDNA gene (Schoch et al. 2012) was amplified using the primers ITS1F-ITS2 (Gardes et Bruns, 1993 ; White et al., 1990). To avoid a two-stage PCR protocol, each primer contained the Illumina adaptor sequence and a tag (ITS1F : 5'- CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxxxxxxxxCTTGGTCATTTAGAGGAAGTAA-3' ; ITS2 : 5'- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxxxxxxxxxGCTGCGTTCTTCATCGATGC-3', where "x" is the 12 nucleotides tag). The PCR mixture (20 μL of final volume) consisted of 10 μL of 2X QIAGEN Multiplex PCR Master Mix, 2 μL each of the forward and reverse primers (1 μM), 4 μL of water, 1 μL of 10 mg.ml^{-1} BSA and 1 μL of DNA template. PCR cycling reactions were conducted on a Veriti 96-well Thermal Cycler (Applied Biosystems) using the following conditions : initial denaturation at 95 °C for 15 min followed by 35 cycles at 94°C for 30 s, 57°C for 90 s, 72°C for 90 s with final extension of 72°C for 10 min. ITS1 amplification was confirmed by electrophoresis on a 2% agarose gel. Each PCR plate contained one negative extraction control, one negative PCR control, and one positive control. The negative extraction control corresponded to a microtube left empty on the collection plate. The negative PCR controls were represented by PCR mix without any DNA template. The positive PCR controls were composed of an equimolar mixture of the DNA of two marine fungal strains – *Candida oceani* and *Yamadazyma barbieri* – (Burgaud et al., 2011 ; Burgaud et al., 2016), kindly provided by Dr. Gaétan Burgaud (LUBEM EA 3882 – ESIAB,

Plouzané, France). These strains were chosen as positive controls as they were unlikely to be found in the grapevine samples. PCR products were purified (CleanPCR, MokaScience), quantified (QuBit) and equimolarly pooled (Hamilton robot). Library sequencing on one run of an Illumina MiSeq platform (v2 chemistry, 2×250 bp) and sequence demultiplexing (with exact index search) were performed at the PGTB sequencing facility (Genome Transcriptome Facility of Bordeaux, Pierroton, France).

Environmental samples collected on the last sampling date (old leaves, bark and ground cover samples) were processed separately following the protocols described in Fort et al. (2019). They were manually ground in liquid nitrogen, amplified using a two-step PCR and then sequenced on a separate MiSeq run at the GetPlaGe sequencing facility (Toulouse, France).

3.3 Fungal DNA quantification by ddPCR™

Fungal DNA was quantified with digital droplet PCR assays (ddPCR™ ; Hindson et al., 2011) using the universal fungal primer pair ITS1F (5'-TCCGTAGGTGAACCTGCGG-3') and 5.8S (5'-CGCTGCGTTCTTCATCG-3'), developed and validated by Fierer et al. (2005). Assays were carried out with the QX200 Droplet Digital PCR (ddPCR™) System from Bio-Rad at the PGTB sequencing facility (Pierroton, France). The PCR reactions were carried out in a final volume of 20 μl using the ddPCR™ EvaGreen Supermix (Bio-Rad, USA). The reaction mix consisted of 1X of EvaGreen Supermix, 2.5 μL of each primer at 1.2 μM and 3.3 μL of DNA template or ultrapure water in the negative control. The mix containing the sample was partitioned into droplets with the QX200 Droplet Generator and then transferred to 96-well PCR plates. A thermocycling protocol [95°C × 5 min ; 40 cycles of (95°C × 30 s, 53°C × 1 min 30 s), 4°C × 5 min, 90°C × 5 min] was undertaken in a Bio-Rad C1000 (Bio-Rad, USA). QX200 droplet reader analyzed each droplet individually to detect the fluorescence signal. The number of copies of the target DNA sequence per μl of sample was determined from the number of positive droplets (out of an average of ~20k droplets per sample) estimated from fluorescence signals of both samples and negative controls using the Umbrella procedure implemented in R (Jacobs et al., 2017). Total fungal abundance was then obtained by multiplying the obtained concentration by the mix volume and after adjusting for the 1/100 dilution of the DNA extract.

3.4 Bioinformatic analysis

We used DADA2 v1.8 (Callahan et al., 2016) to describe fungal communities in terms of Amplicon Sequence Variants (ASV, Callahan et al., 2017), because bioinformatic approaches based on DADA2 were shown to recover accurately the composition of an artificial community of 189 fungal strains (Pauvert et al., 2019a). We followed the DADA2 ITS workflow (https://benjjneb.github.io/dada2/ITS_workflow.html) except on the read assembly step. Only forward reads in which the primer sequence was found exactly by cutadapt (Martin, 2011) were processed with DADA2. Quality filtering retained sequences with less than one expected error and longer than 50 bp. ASV were subsequently inferred for each sample. Chimeric sequences were then identified using the consensus method of the *removeBimeras* function. ASV identical in sequence but not in length were finally combined with the *collapseNoMismatch* function. Taxonomic assignments were then performed using the RDP classifier (Wang et al., 2007) implemented in DADA2 and trained with the latest UNITE database (Community, 2019).

ASV table and taxonomic assignments were imported in R through the phyloseq package (McMurdie et Holmes, 2013). Only ASV assigned to a fungal phylum were kept. Positive and negative controls were then used to remove contaminants (Galan et al., 2016). The two ASV corresponding to the marine strains contained in the positive controls (*Candida oceani* and *Yamadazyma barbieri*) were identified after aligning the ASV sequences with the corresponding Genbank reference sequences (*C. oceani* KY102240 and *Y. barbieri* LT547714) using a similarity threshold of 100% with *usearch_global* function of VSEARCH (Rognes et al., 2016). The cross-contamination threshold (T_{CC}) was defined as the maximal number of sequences of each ASV found in negative or positive control samples. The false-assignment threshold (T_{FA}) was defined as $T_{FA} = N \times R_{fa}$ where R_{fa} is the highest sequence count of a positive control strain in a non-control sample, divided by the total number of sequences of the strain in the whole run and N is the total number of sequences of each ASV. ASV were removed from all samples where they harbored fewer sequences than either threshold (T_{FA} or T_{CC}) using a custom script (https://gist.github.com/cpauvert/1ba6a97b01ea6cde4398a8d531fa62f9). The ASV table was finally aggregated at the species level and ASV that could not be assigned at the species level were removed.

3.5 Hypothesis testing

Before testing the three hypotheses, we checked that foliar tissue samples considered as infected based on visual symptom assessment, had a higher fungal load and a higher relative abundance of *Erysiphe necator* than samples considered as healthy. Variations in fungal total abundance (log-transformed) and in *Erysiphe necator* sequence counts (CLR-transformed) were analyzed using linear regressions performed with the *lm* function (R Core Team, 2018). The models had cropping system (CC or NCC), sampling date (40, 62 or 77 dpi), tissue type (HI, HNI or DI) and their interactions as fixed effects. ANOVA *F*-tests were used to assess the significance of the fixed effects and post-hoc pairwise comparisons were performed for the significant interactions using Tukey's adjustment method with the emmeans R package (Lenth, 2018).

Test of H1 : Successful infection events destabilize plant-associated microbial communities and increase their heterogeneity (Anna Karenina principle)

To test H1, we first investigated whether infection altered the composition of foliar fungal communities, using a permutational analysis of variance (PERMANOVA ; Anderson, 2001) performed with the *adonis* function of the vegan R package (Oksanen et al., 2018). The model had tissue type, cropping system, sampling date and their interactions as fixed effects. The analysis was performed using CLR-transformed data, as advised by Gloor et al. (2017). A total of 128 CLR-transformed species × samples matrices were generated using the *aldex.clr* function of ALDEx2 (Fernandes et al., 2014) and averaged. Euclidean distances among samples of the average matrix were then used to test differences in community composition among conditions. These analyses were performed twice, with and without removing the ASV corresponding to the pathogen *Erysiphe necator*.

Then, we investigated whether compositional heterogeneity was larger among infected samples than among healthy ones, using tests of homogeneity of multivariate dispersions (Anderson, 2006) performed with the *betadisper* function of the R vegan package (Oksanen et al., 2018). Compositional heterogeneity within a group of samples (*i.e.* dispersion) was measured

using the distances between each sample and the centroid of its group. Dispersion was calculated for every combination of sampling date and tissue type. Differences in dispersion among groups were tested using ANOVA *F*-tests. Post-hoc pairwise comparisons were performed with Tukey's adjustment method with the emmeans R package (Lenth, 2018).

Test of H2 : Interactions among microorganisms within pathobiomes can be detected by inferring microbial networks from metabarcoding data and environmental covariates

To test H2, we first inferred the foliar microbial network from metabarcoding data and environmental covariates, and then tested some of the associations using co-culture experiments and text-mining. Joint variations in sequence counts between fungal species were modelled with the Poisson Log-Normal model (PLN, Aitchison et Ho, 1989) to account for uneven sequencing depth among samples and for environmental covariates potentially explaining variations in the sequence counts per fungal species across the dataset. Associations between species were then inferred from the residual covariance matrix using graphical-lasso under the sparsity assumption, by using the PLNmodels R package (Chiquet et al., 2019). The extended BIC criteria (Chen et Chen, 2008) was used to perform model selection among a 40-size grid of penalties (Chiquet et al., 2017 ; Chiquet et al., 2019). The stability of each association of the BIC network was then defined as its selection frequency in the bootstrap subsamples of the StARS procedure (Liu et al., 2011). Species included in the network reconstruction were selected based on their prevalence rather than their relative abundance (Röttjers et Faust, 2018). Based on the theoretical work of Cougoul et al. (2019a), we chose to include only species present in more than 20% of the samples of any experimental unit (Methods S6.1). Covariates included several foliar traits and climate variables known to influence fungal development : leaf age, leaf water content, specific leaf area and average vapor pressure deficit experienced by the leaf since unfolding (Table S5). These covariates were introduced in the model to get rid of fungal species associations triggered by similar habitat requirements. The distance between the sampled leaf and environmental sources of microorganisms (old leaves, bark, ground cover) was also included as a covariate to get rid of fungal associations triggered by joint colonization events. Only the distance between the sampled leaf and old leaves was introduced in the model as all three distances were significantly correlated (Figure S6). To account for additional, non-measured environmental variations associated with the experimental design, we also included the three main factors of the experiment (sampling date, experimental unit and tissue type) as covariates. Finally, sequencing depth (log-transformed) was included as an offset in the model to get rid of spurious associations triggered by compositionality. A subset of the fungal network was then tested using co-cultures, to determine whether the statistical associations among fungal species were triggered by ecological interactions. Co-cultures investigated the interactions among eight yeast strains, and between these eight yeast strains and one strain of the pathogen *Erysiphe necator*. The eight yeast strains were chosen among the most abundant fungal species found on grapevine leaves. They were all present in the inferred fungal network. They were also available in the CBS collection (Westerdjik Fungal Biodiversity Institute, Utrecht, The Netherlands) and could grow on PDA (Potato Dextrose Agar, Biokar diagnostics, France) at 25°C : *Buckleyzyma aurantiaca* CBS 8074, *Cystofilobasidium macerans* CBS 9032, *Dioszegia hungarica* CBS 7091, *Filobasidium oeirense* CBS 8681, *Filobasidium wieringae* CBS 1937, *Udeniomyces pyricola* CBS 6754, *Sporobolomyces roseus* CBS 486 and *Vishniacozyma victoriae* CBS 9000. Strains were revived following the CBS instructions and further maintained in collection on MEA and PDA media (Malt Extract Agar : Malt 15 $g.L^{-1}$, Agar 20 $g.L^{-1}$; Potato Dextrose Agar : Potato Extract 4 $g.L^{-1}$, Dextrose 20 $g.L^{-1}$, Agar 15 $g.L^{-1}$). To produce inoculum for the co-culture experiments, yeast strains were grown in liquid ME medium (Malt Extract, 15 $g.L^{-1}$)

Table 1 – Effects of date, cropping system (CS) and tissue type (TT) on pathogen abundance (A), total fungal load (B) and community composition (C). Significant effects ($p < 0.05$) are in bold.

		A. Pathogen abundance			**B. Total fungal abundance**			**C. Fungal community composition**		
Variable	Df	F	Pr(>F)		F	Pr(>F)		F	R2	Pr(>F)
Date	2	**55.29**	**1.00E-20**		**54.27**	**2.10E-20**		**5.1**	**0.0336**	**0.001**
CS	1	0.1	7.50E-01		0.76	3.90E-01		**1.5**	**0.005**	**0.021**
TT	2	**70.82**	**3.10E-25**		**50.87**	**2.40E-19**		1	0.0068	0.377
Date × CS	2	0.29	7.50E-01		1.42	2.40E-01		**9.7**	**0.0642**	**0.001**
Date × TT	4	**6.89**	**2.80E-05**		**17.9**	**5.50E-13**		1.1	0.0148	0.13
CS × TT	2	1.08	3.40E-01		0.34	7.10E-01		1.2	0.0079	0.114
Date × CS × TT	4	1.11	3.50E-01		1.72	1.50E-01		1.1	0.0148	0.13
Residuals	258				NA	NA		NA	0.8529	NA
Total	275							NA	1	NA

at 22°C for 48h. The number of yeast cells was then count using a hemacytometer before being homogenized at the same concentration. *E. necator* strain S19 was isolated in a greenhouse in INRA precincts (Villenave d'Ornon, Bordeaux, France) in 2019. As it is an obligate biotroph parasite of grapevine (*i.e.* requires a living host), its preservation was carried out every 2 weeks on sterilized grapevine leaves by blowing spores from sporulating leaves (approx. 1000 conidia per cm^2 of leaf). Inoculated leaves were then incubated for 10-14 days at 22°C with a 16 h day/ 8 h night photoperiod following the procedure Dufour et al. (2013).

To investigate whether the statistical associations among the yeast species were triggered by ecological interactions, the eight yeast strains were confronted with each other using spot-on-lawn assay (Addis et al., 2001 ; Polonelli et Morace, 1986). Inoculations were performed on MEA medium buffered at pH 4.5 with 0.5 M phosphate-citrate buffer (Heard et Fleet, 1987). Approximately 10^5 $CFU.mL^{-1}$ of each strain was suspended in 15 ml of sterile MEA medium (pH 4.5) maintained at 45°C and then poured into sterile Petri dishes. For each strain, five replicates were made, totalizing 40 plates. Each of the eight strains was then drop-inoculated (50 μl at 10^5 $CFU.mL^{-1}$) (adapted from Polonelli et Morace, 1986) onto the surface of each of the Petri dishes prepared as described above. Once the drops dried, the plates were sealed with Parafilm®and incubated at 22°C. Yeast growth was visually assessed 4 and 7 days after inoculation. We considered that the MEA-included strain inhibited (or promoted) the growth of a drop-inoculated strain, if the latter had a reduced (or increased) growth relative to its own confrontation.

In addition, the subset of the fungal network involving *E. necator* (i.e. its putative pathobiome) was checked for in the literature using a custom R script (Methods S6.2). First, every pair of fungal species belonging to the pathobiome was replicated using obligate and anamorph synonyms of species names, queried from the MycoBank webservice (Robert et al., 2013). For each pair, co-occurrences of species names were then searched for in the Scopus database (that includes article titles, abstracts, keywords and references) using its API through the rscopus R package (Muschelli, 2019). The corresponding articles were read carefully to search for experimental evidence of ecological interaction between the selected species.

Test of H3 : Cropping systems influence the abundance of microorganisms forming pathobiomes

To test H3, we investigated whether the fungal species belonging to the putative pathobiome of *E. necator* differed in abundance between CC and NCC cropping systems. ALDEx2 (Fernandes et al., 2014) was used to detect species differentially abundant between systems as this method is suitable for compositional data (Gloor et al., 2017) and outputs very few false-positives with default values (Thorsen et al., 2016). Species with a false discovery rate below 0.1 after Benjamini-Hochberg adjustment were considered as differentially abundant. To better understand the ecology of the fungal species belonging to the pathobiome of *E. necator*, we also searched for their presence in the environmental samples. Finally, to assess more broadly the effects of cropping systems, we compared grapevine canopy vigor and disease severity between CC and NCC experimental units. Differences in canopy vigor or in disease severity between cropping system were assessed using a single t-test for the former and one test for each of the four recorded diseases for the latter.

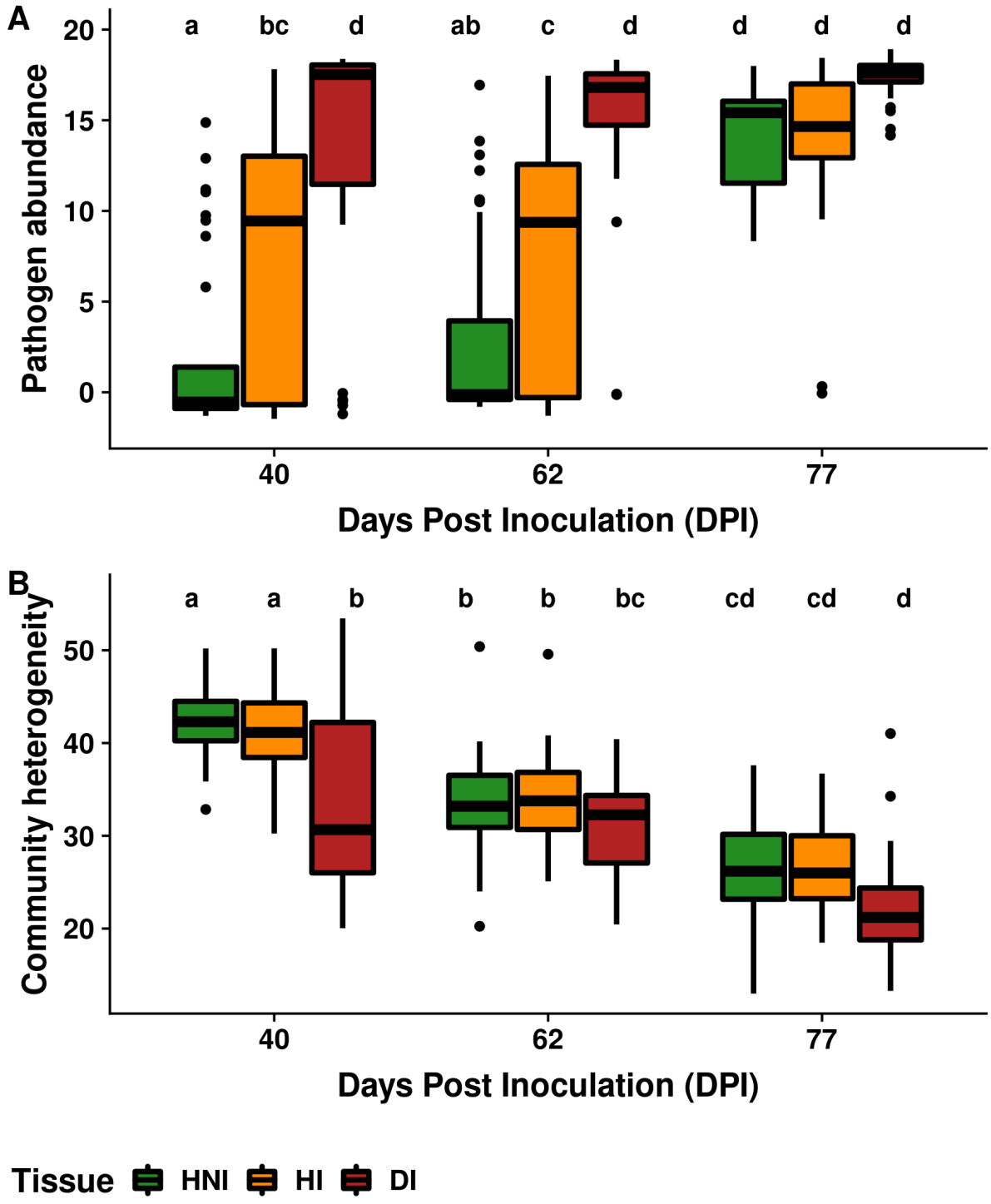


Figure 2 – Variations in pathogen abundance (A) and fungal community heterogeneity (B) among tissue types over time. Tissue types consisted in Healthy tissue in Non Infected leaves (HNI), Healthy tissue in Infected leaves (HI), and Diseased tissue in Infected leaves (DI).

4 Results

Grapevine foliar fungal communities were composed of 4 148 fungal ASV (totalizing 10 195 266 sequences), among which 1454 could be assigned at the species level (totalizing 7 276 628 sequences). These ASV corresponded to 547 fungal species (306 Ascomycetes and 241 Basidiomycetes). The pathogen *Erysiphe necator* was among the ten most abundant species (Table S6). The relative abundance of *E. necator* was significantly higher in tissues considered as infected based on visual assessments (DI), than in tissues considered as healthy (HI and HNI), except on the last sampling date (Table 1A and Figure 2A). On the last sampling date, the pathogen was highly abundant in all tissue types (Figure 2A). The sequences assigned to *E. necator* represented 39%, 39.6% and 65.5% of all the fungal sequences on the three sampling dates, respectively. The total fungal abundance also varied significantly with tissue type and sampling date (Table 1B). As expected, the total fungal abundance was higher in infected tissues, except on the last sampling date (Figure S7).

Infection does not destabilize the microbial community (no Anna Karenina principle)

In contrast to H1, infection did not trigger any significant change in fungal community composition (Table 1C). Fungal community composition did not differ among healthy and infected tissues, but it differed between cropping systems and, to a greater extent, between sampling dates (Table 1C and Figure S8). While there was a significant difference in community heterogeneity among healthy and infected tissues (ANOVA : df = 8 ; F = 37.7 ; $p < 0.01$), the heterogeneity was not higher among infected samples than among healthy ones (Figure 2B) . It was even lower in infected samples on the first sampling date (Figure 2B). The same results were obtained when *E. necator* was removed from the fungal community data (not shown), confirming the absence of Anna Karenina principle.

Inferred microbial networks give relevant insights into the pathobiome

The fungal association network inferred from metabarcoding data and environmental covariates using PLN model was composed of 61 fungal species and 702 statistical associations (Figure 3A). Within this network, *E. necator* was negatively associated with 15 species and positively associated with 8 species (Figure 4). All associations, except two, were robust to subsampling (Table S7). These 23 species are hereafter considered as the putative pathobiome of *E. necator*.

The 8 yeast species selected for microbiology experiments were linked by 7 positive associations in the PLN network (Figure 3B). In contrast, spot-on-lawn experiments (Figure 3C) revealed 3 growth-promoting interactions and 12 growth-inhibiting interactions (Figure 3D) among the 7 yeast strains that grew in the experimental conditions (all strains but *D. hungarica*). The statistical association network (Figure 3B) and the *in vitro* interaction network (Figure 3D) shared two links, one between *C. macerans* and *U. pyricola* and another between *F. oeirense* and *C. macerans*. The positive association between *C. macerans* and *U. pyricola* was upheld by the spot-on-lawn assay as *U. pyricola* enhanced the growth of *C. macerans*. In contrast, the link between *F. oeirense* and *C. macerans* was found as a positive association with PLN, while the growth of *C. macerans* was inhibited by *F. oeirense* in the spot-on-lawn assay.

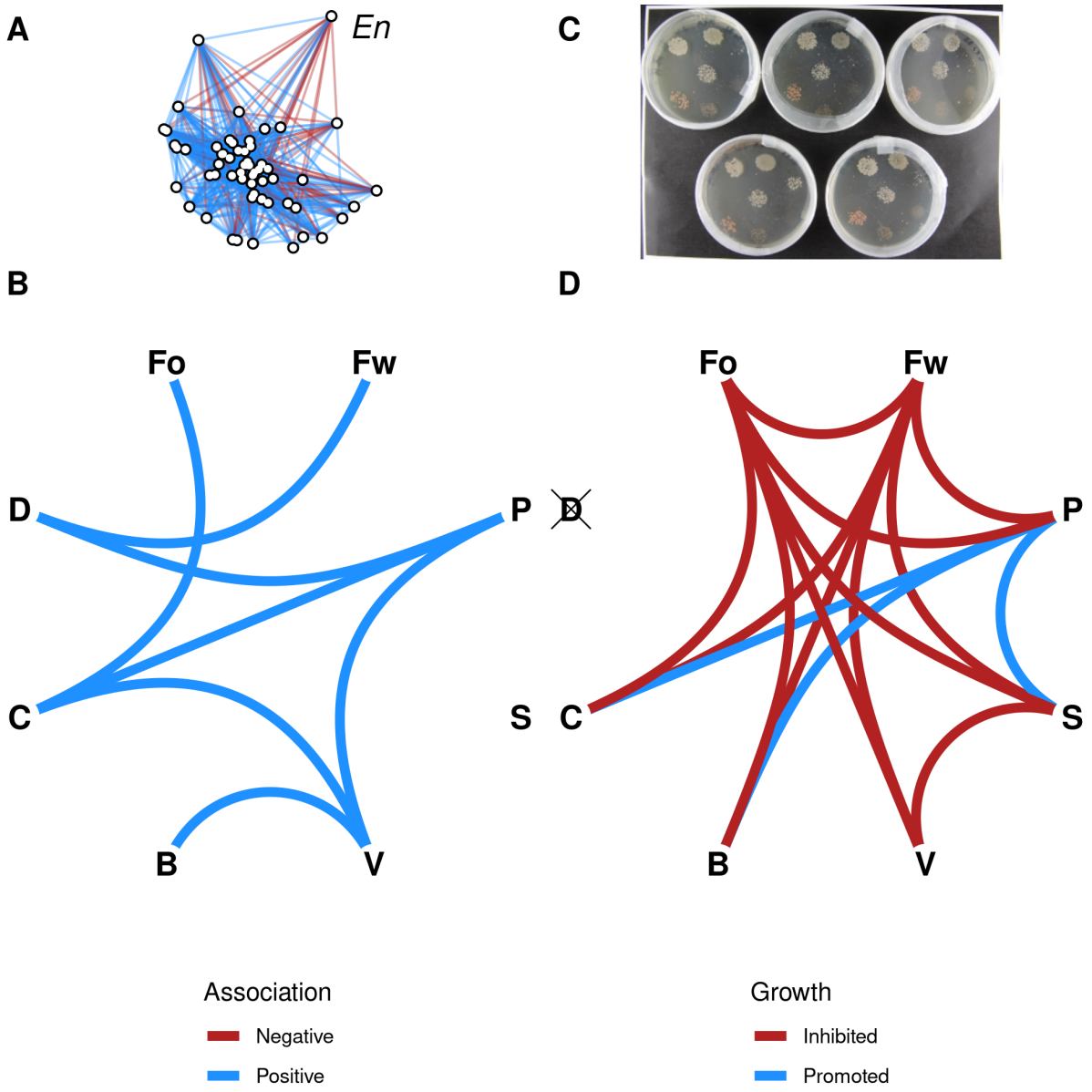


Figure 3 – Association network versus *in vitro* interaction network among eight yeast species. First, an association network (A) was obtained with PLN from metabarcoding data and environmental covariates. The node corresponding to the pathogen *Erysiphe necator* is labelled *En*. A subset of the association network involving 8 yeast species was extracted. It consisted only in positive associations (B). We confronted this network with spot-on-lawn experiments (C) where a single yeast species was seeded in a Petri dish and further inoculated by spots of 8 yeast species to evaluate their interaction. Growth-promoting and -inhibiting interactions were represented as an interaction network (D). The yeast species consisted in *Buckleyzyma aurantiaca* (B), *Cystofilobasidium macerans* (C), *Dioszegia hungarica* (D), *Filobasidium oeirense* (Fo), *Filobasidium wieringae* (Fw), *Udeniomyces pyricola* (P), *Sporobolomyces roseus* (S) and *Vishniacozyma victoriae* (V).

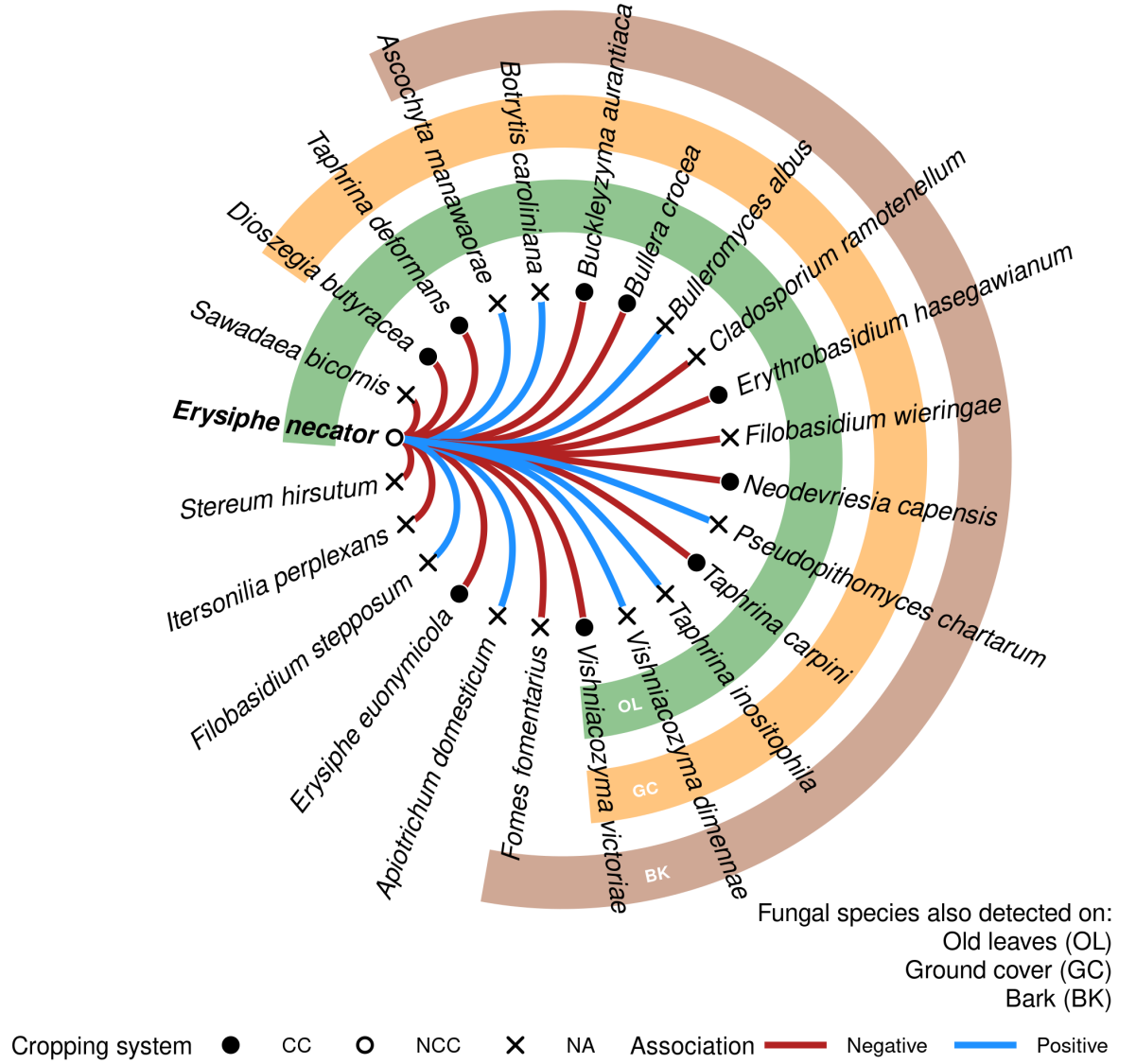


FIGURE 4 – Putative pathobiome of the grapevine powdery mildew agent, *Erysiphe necator*. The subset of the association network (Figure 3A) involving the pathogen was extracted. It consisted of 25 species with positive or negative associations with the pathogen. Node shape indicate whether these species were significantly favored by cover cropping (black circles), by the absence of cover (white circle) or neither (cross). Green, yellow and brown circles indicate whether the species were detected in other microbial environments (old leaves, ground cover and bark, respectively).

Among the 8 yeast species selected for microbiology experiments, 3 were part of the putative pathobiome of *E. necator* : *B. aurantiaca*, *F. wieringae* and *V. victoriae* (Figure 4). The three species were negatively associated with the pathogen in the PLN network (Figure 4). Microbiology experiments supported the antagonistic effect of *B. aurantiaca* on *E. necator*. The supernatant of *B. aurantiaca* centrifuged liquid culture, when applied as a curative treatment, reduced the growth of *E. necator* by about 30% (Table 2) and significantly increased the prevalence of collapsed conidia (Table 3 and Figure 5). The pellet of *U. pyricola* culture also inhibited *E. necator* growth when applied as a curative treatment (Table 2) but there was no association between the two species in the PLN network.

Finally, literature-mining identified 9 abstracts in which *E. necator* and at least one species of its putative pathobiome were mentioned (Table S8). None of these articles provided experimental evidence of an interaction between *E. necator* and the species found in the abstracts.

Overall, network inference provided a list of 15 fungal species that might be antagonists of *E. necator* (Figure 4). Literature mining showed that these interactions have never been studied experimentally so far. We tested experimentally three of these putative interactions and validated one of them, involving the yeast species *B. aurantiaca* (Tables 2, 3 and Figure 5). Hence, in accordance with H2, interactions among microorganisms within the pathobiome can be detected by inferring microbial networks from metabarcoding data and environmental covariates.

Cropping systems impact the pathobiome

Differential abundance analyses revealed that 25 fungal species were favored by cover cropping (Table S9). Among them, 9 fungal species belonged to the putative pathobiome of *E. necator* and were all putative antagonists (Figure 4). The yeast *B. aurantiaca* was part of the species favored by cover cropping. Six of the putative antagonists favored by cover cropping, including *B. aurantiaca*, were generalist fungal species found in grapevine leaves but also on bark and ground cover (Figure 4). These findings indicate that cropping systems do modulate pathobiomes, in accordance with H3. Interestingly, only the pathogen *E. necator* was favored by the absence of cover crop according to differential abundance analyses (Table S9), a pattern which might be accounted for by the drastic increase in vine vigor in the absence of cover crop (Figure S9A). However, visual disease severity assessment performed mid-July did not confirm the higher abundance of *E. necator* in the absence of cover crop (Figure S9B).

Table 2 – Percent increase in *Erysiphe necator* (*En*) growth triggered by preventive or curative treatments with yeast strains or exposure to benzothiadiazole (BTH). Significant effects of the treatments ($p < 0.05$) are in bold.

	Preventive treatment				Curative treatment			
	Pellet		Supernatant		Pellet		Supernatant	
Yeast strain	*En* growth (in %)	p-value	*En* growth (in %)	p-value	*En* growth (in %)	p-value	*En* growth (in %)	p-value
Control (BTH)	**-98.62**	**0**	**-90.62**	**0**	**-85.62**	**1.70E-14**	**-86.25**	**0**
B. aurantiaca	-1.87	0.33	2.25	0.99	-19.63	0.18	**-31.25**	**0**
U. pyricola	-1	0.85	6	0.83	**-32.5**	**0**	-3.75	0.97
V. victoriae	-0.37	0.99	7.63	0.67	-10	0.79	3.88	0.97
D. hungarica	0.25	1	0.63	1	-5	0.98	-10.62	0.44
F. wieringae	0.5	0.98	6.38	0.8	-10.63	0.75	-1.87	1
S. roseus	0.88	0.9	0.13	1	-6.88	0.93	-5.25	0.91
F. oeirense	1.13	0.78	-4.5	0.93	-19.38	0.19	-0.87	1
C. macerans	1.38	0.64	6.25	0.81	-10.63	0.75	5	0.93

5 Discussion

Dealing with pathogens is a matter of host survival, and in the case of crops, it also concerns the species that depend on them, including human beings. Among the range of disease control strategies is the exploitation of ecological interactions between pathogenic microorganisms and microorganisms that inhabit the host plant (Brader et al., 2017 ; Koskella et al., 2017). These interactions, termed the pathobiome (Vayssier-Taussat et al., 2014), were studied here using a metabarcoding approach, with grapevine powdery mildew (*Erysiphe necator* ; Armijo et al., 2016 ; Gadoury et al., 2012) as a model system. Fungal communities sampled from visually healthy and infected foliar zones did not differ in terms of composition, contrasting with the results obtained for oak powdery mildew (*Erysiphe alphitoides* ; Jakuschkin et al., 2016). Furthermore, we had no evidence of the presumed Anna Karenina principle, which predicts more heterogeneous microbial communities following infection (Zaneveld et al., 2017). These findings parallel other studies on plant pathogens (Agler et al., 2016 ; Koskella et al., 2017 ; Rezki et al., 2016 ; Suda et al., 2009 ; Zhang et al., 2019), suggesting that the Anna Karenina principle is not the rule for plants. To discover fungal species with potential antagonistic activity against the pathogen, we sampled hosts in both healthy and diseased conditions (Ellis, 2017 ; Poudel et al., 2016) and used a new statistical approach, the PLN model (Chiquet et al., 2017 ; Chiquet et al., 2019), which allowed us to take into account environmental variability. We were able to generate new hypotheses in the form of a network of associations between microorganisms and some of these hypotheses were tested using co-cultures. Interestingly, the yeast *Buckleyzyma aurantiaca* was found to reduce pathogen colonization and even alter the structure of asexual reproduction organs, the conidia. These findings upheld a negative association between the abundances of *B. aurantiaca* and *E. necator* across foliar samples collected in the field, which makes this yeast a potential biocontrol candidate worthing further investigations. Moreover, our results revealed that *B. aurantiaca* and eight other putative antagonists of *E. necator* were favored by cover cropping, suggesting that cropping systems can be used to favor protective plant microbiota. Finally, another positive association, involving the yeast species *Cystofilobasidium macerans* and *Udeniomyces pyricola*, was also supported by co-cultures. These findings prove that network inference methods can detect relevant signals of interactions between microorganisms and be useful tools for the development of microbiome-based crop protection strategies.

Although informative, all these results are only a glimpse of the complex processes that occur *in planta* and *in natura*. The destructive sampling of foliar tissues prevented us to monitor fungal communities continuously over time and to generate times series that could better inform us on microbial interactions (*e.g.* Carr et al., 2019). We relied on cross-sectional data with a large number of samples instead. In this study, 276 samples were used to reconstruct the fungal network, which is in line with the recent recommendation by Hirano et Takemoto (2019) of at least 200 samples per network, and the inferred associations were robust. To complete this association network, future studies will have to collect data on bacteria, oomycetes, viruses and even insect communities because inter-kingdom interactions influence the structure and dynamics microbial networks (Durán et al., 2018 ; Jakuschkin et al., 2016 ; Tipton et al., 2018). This additional information will also improve the inference of fungal associations because missing species can generate indirect associations (*e.g.* Li et al., 2016). Prior information on

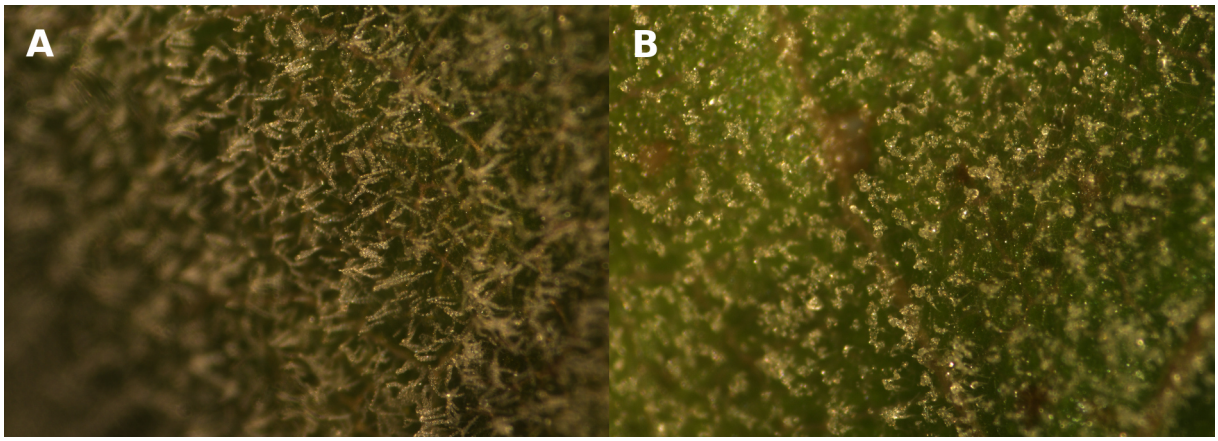


Figure 5 – Pictures of *E. necator* conidia in the *E. necator*-yeast strains confrontation test. Unaltered conidial chains on control foliar discs treated with sterile distilled water (A), collapsed conidia on foliar discs treated with *B. aurantiaca* in the preventive assay (B).

interactions gathered from literature could also be used to improve network inference (Lo et Marculescu, 2017).

In conclusion, microbial networks are no silver bullet but like any other methods provide an insight into the complex processes at play in ecosystems. The field of metabarcoding approaches is now mature enough to be aware of its pitfalls (*e.g.* Zinger et al., 2019). We suggest that the difficult hurdles to overcome should not prevent microbial network inferences but rather be an incentive for cautious interpretations and additional experiments in line with recent recommendations (Armitage et Jones, 2019 ; Carr et al., 2019). Better insights into microbial ecosystems may not necessarily require more metabarcoding samples but more complementary evidence (Widder et al., 2016).

Acknowledgements

We thank Jérôme Jolivet, Lionel Druelle, Hany El Manawy and Benoît Bénéteau (UMR SAVE) for their help during the sampling campaigns, and Isabelle Demeaux (UMR SAVE) for her help with *E. necator* experiments. We thank Lionel Delbac for providing previous pesticide uses in the sub-plot. We also thank Gregory Gambetta, Guilherme Martins, Frédéric Barraquand, Isabelle Lesur, Adrien Rush and all members of the ANR NGB Consortium (ANR-17-CE32-0011) for useful comments on preliminary results. We thank the CLIMATIK portal (INRA Avignon) for providing climatic data concerning the weather station 33550003. We thank the Genotoul sequencing facility (Get-PlaGe) and the PGTB sequencing facility for sequencing, and the Genotoul bioinformatics facility (Bioinfo Genotoul) for providing computing and storage resources. We thank the INRA MEM metaprogram (Meta-Omics of Microbial Ecosystems) for financial and scientific support (Learn-biocontrol project). Additional funding was received from the LABEX COTE (ANR-10-LABX-45), the LABEX CEBA (ANR-10-LABX-25-01) and INRA EcoServ metaprogram on ecosystem services (IBISC project) and the Aquitaine Region (Athene project, n°2016-1R20301-00007218). CP's PhD grant was funded by the INRA and Bordeaux Sciences Agro (BSA). The management of the experimental site was partly funded by the AFB (French Agency for Biodiversity) within the DEPHY network.

Authors contributions

JV and AC performed inoculations. JV, CV and AC performed leaf monitoring and sampling. JV and CV processed the sampled leaves. TF and JFA performed the extractions and amplifications. EC performed the sequencing. CP and MM quantified fungal abundance using ddPCR. JV and CP performed the microbiology experiments. CP performed the bioinformatic and statistical analyses. SR and JC provided methods and advice on network inference. CV and DB had the original idea for the project. CV and JV coordinated all stages of the experiment. CP wrote the first draft of the manuscript, and CV and JV made a major contribution to the writing of the final draft. All authors revised the manuscript.

Table 3 – Effect of yeast strains on the conidia of *E. necator*. The number of foliar discs (out of 8) with altered conidia were counted in different co-culture conditions. Bold entries indicate significant differences in the number of altered conidia relative to the negative control using Fisher exact test. No conidia were observed with the positive control (benzothiadiazole) which was thus not included.

	Preventive treatment				Curative treatment			
	Pellet		Supernatant		Pellet		Supernatant	
Yeast strain	Altered conidia	p-value	Altered conidia	p-value	Altered conidia	p-value	Altered conidia	p-value
V. victoriae	4	0.141	4	0.141	1	0.7667	1	0.7667
B. aurantiaca	3	0.2846	5	0.0594	5	0.0594	**8**	**0.0007**
D. hungarica	1	0.7667	1	0.7667	4	0.141	0	1
U. pyricola	1	0.7667	0	1	5	0.0594	0	1
S. roseus	1	0.7667	1	0.7667	5	0.0594	1	0.7667
C. macerans	0	1	2	0.5	3	0.2846	0	1
F. oeirense	0	1	5	0.0594	3	0.2846	0	1
F. wieringae	0	1	0	1	3	0.2846	1	0.7667
Negative control	1		1		1		1	

6 Supplementary information

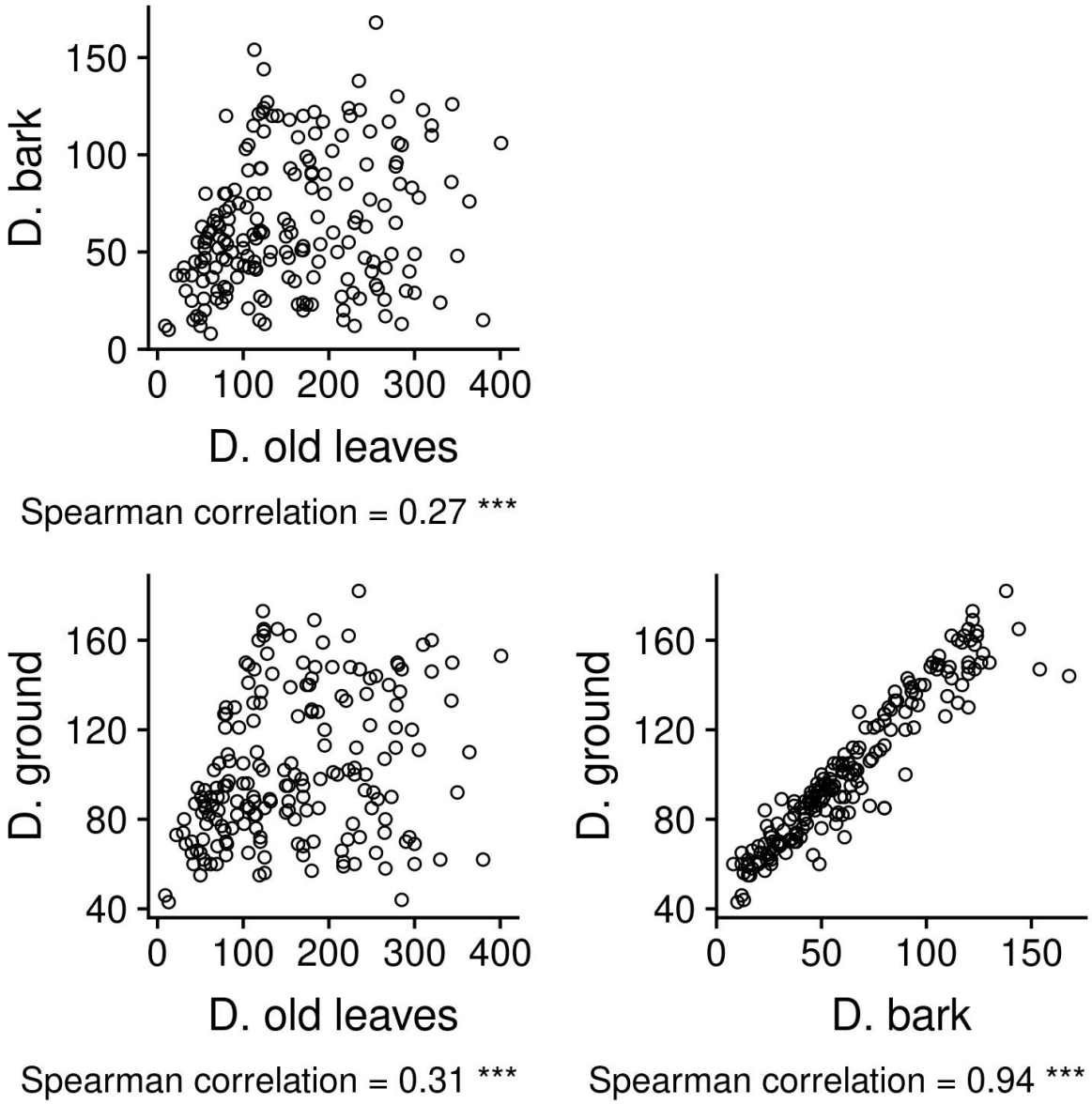


Figure 6 – Relationships among the distances between sampled leaves and the closest environmental sources of microorganisms. The distances (D, in cm) to old inoculated leaves, bark, and ground cover were measured. Spearman's rank correlation and test are presented ($^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

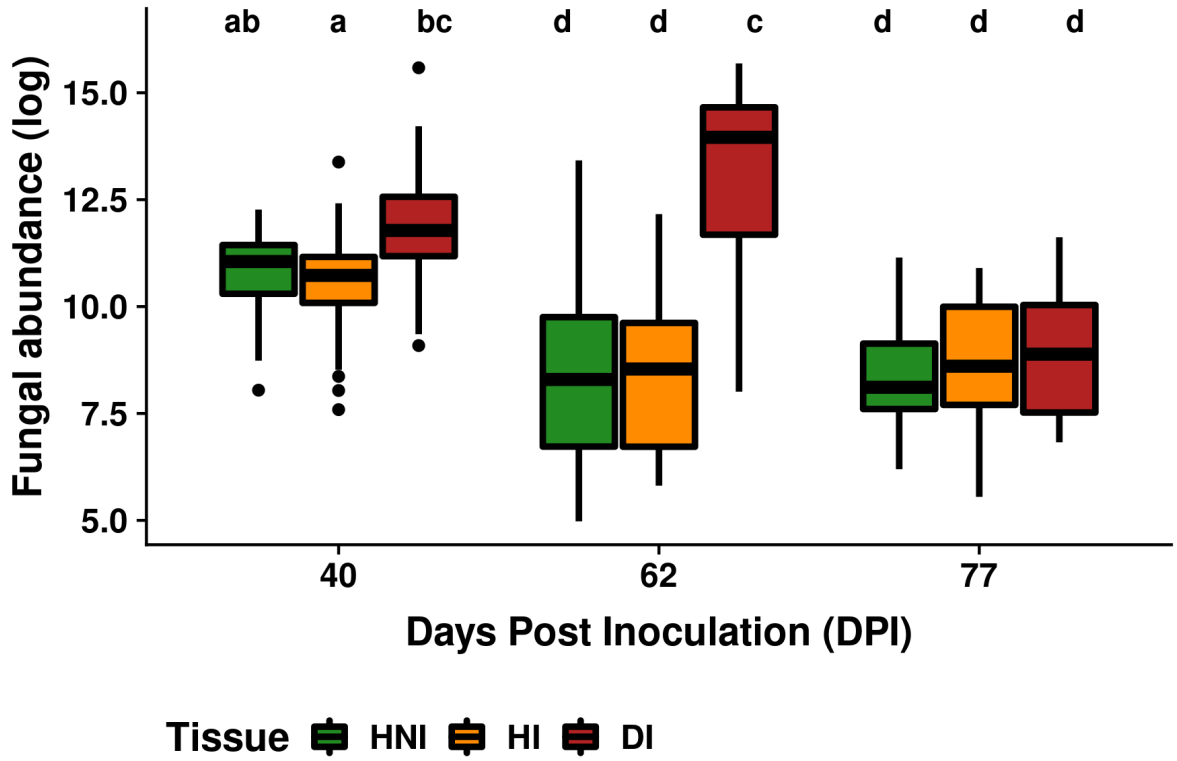


FIGURE 7 – Variations in fungal total abundance among tissue types over time. Different letters indicate significant post-hoc comparisons. Tissue type consisted in Healthy tissue in Non Infected leaves (HNI), Healthy tissue in Infected leaves (HI), and Diseased tissue in Infected leaves (DI).

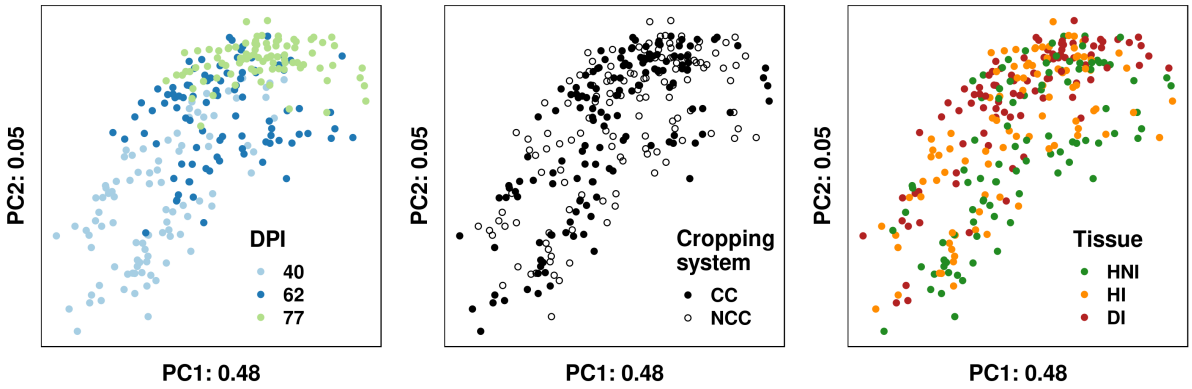


FIGURE 8 – Variations in fungal community compositions among sampling dates (DPI), cropping system (CS) and tissue types (TT). The first two axes of the Principal Component Analysis (PCA) were used to represent the dissimilarities in composition, from left to right, among samples across sampling dates (in days post inoculation, DPI), cropping systems (Cover crop, CC, and no cover crop, NCC) and tissue types (Healthy tissue in Non Infected leaves, HNI, Healthy tissue in Infected leaves, HI, and Diseased tissue in Infected leaves, DI). PCA was performed from the Euclidean distance on clr-transformed sequence counts.

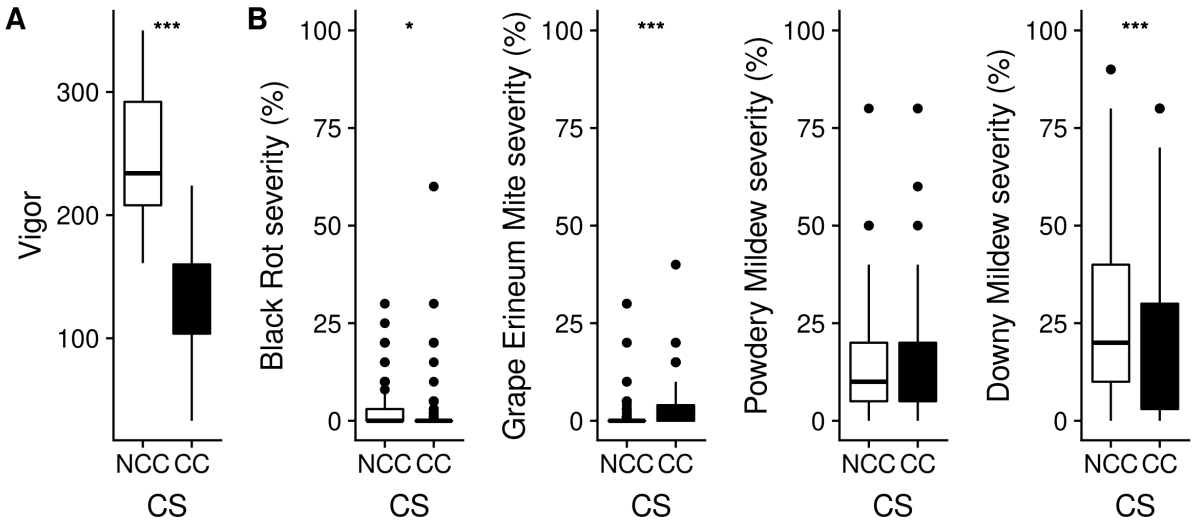


Figure 9 – Canopy vigor and disease pressure indexes per cropping system (CS), *i.e.* with (CC) or without cover crop (NCC). Vigor was estimated on ~20 grapevines per cropping system (A). Grapevine diseases severity were monitored on 468 leaves randomly distributed among the two cropping systems (B) (* p <0.05; ** p <0.01; *** p <0.001).

Table 4 – Pesticides (active compounds) used on the sub-plot since 2006 according to the pest targeted. D : Downy mildew (*Plasmopara viticola*); P : Powdery mildew (*Erysiphe necator*), E : European grapevine moth (*Lobesia botrana*) and S : Flavescence dorée vector (*Scaphoideus titanus*). Applications were done in the 5 rows between May and July 2006-2012.

Pesticide active compounds	**Pest**	**Year (date)**
Chlorpyriphos-ethyl	T, S	2006 (7/3); 2009 (5/20, 7/6)
Copper compounds	D	2008 (7/30)
Fosetyl + Folpet	D	2008 (7/21)
Lambda-cyhalothrin	S	2009 (6/16)
Mancozeb + Cymoxanil	D	2006 (5/23, 6/9); 2007 (6/11); 2008 (5/28, 6/6, 6/26); 2009 (6/26); 2011 (5/19); 2012 (6/22, 7/2)
Quinoxyfen	P	2008 (6/6); 2011 (5/19)
Spiroxamin	P	2008 (6/26, 7/21)

Table 5 – List of covariates included in the network inference. Covariate fluctuations can be tied positively (↗), negatively (↘) to their effect on foliar fungi or without a definite sign (∝). Bold entries indicate references concerning our biological models *Vitis vinifera* or *Erysiphe necator* (*En*). The VPD relates to the gap between a water saturated air to the actual air given the temperature (see Cordier et al., 2012b, for more details on the function f).

Type	Covariates	Definition	Effects on fungi	References
Foliar traits	Leaf water content (LWC)	1 – (D/F) with fresh (F) and dry (D) leaf weight	↗ with endophytic fungal richness	Sanchez-Azofeifa et al. (2012)
			∝ the occurrence of endophytes	Eschen et al. (2010)
	Leaf age (LA)	Days since leaf unfolding	↗ with endophytic fungal richness	Osono (2008)
			↗ with endophytic fungal richness and number of isolates	López-González et al. (2017)
			↘ **the susceptibility to *En***	Calonnec et al. (2018) et Gadoury et al. (2012)
	Specific Leaf Area (SLA)	S/D with leaf surface (S)	↘ the development of mildew and rust pathogens	Itagaki et Shibuya (2017) et Toome et al. (2010)
			↘ the establishment of endophytes and hence ∝ fungal abundance	Nezhad et Geitmann (2013)
			S ↗ collected water and hence ∝ the infection susceptibility	Bradley et al. (2003)
			S ↗ with the log number of individuals fungi	Andrews et al. (1987)
Climatic variables	Vapor Pressure Deficit (VPD)	f(RH × T) with relative humidity (RH) and temperature (T)	RH × T ∝ with spore density	Crandall et Gilbert (2017)
			↘ with fungal richness and abundance	Talley et al. (2002)
			↘ **with the incidence and severity of *En***	Carroll et Wilcox (2003)
Dispersal	Distance	Distances to microbial sources	↘ **with microbial invasion**	Miura et al. (2017)
Experimental design	Sampling date		∝ **the fungal community composition**	This study ; Fort et al. (2016) et Pinto et al. (2014)
	Experimental unit		∝ **the fungal community composition**	This study ; Setati et al. (2012)
	Tissue type		∝ **the fungal community composition**	This study ; Bruez et al. (2014)

Table 6 – Most abundant species in grapevine foliar fungal communities in the sampled foliar tissues. The relative abundances (RA, in %) and ranks of species were calculated for all leaf samples (TOTAL ; n = 276) and for samples collected from healthy zones of non-infected (HNI ; n = 93) and infected (HI ; n = 92) leaves, and the center of disease spots of infected leaves (DI ; n = 91).

	TOTAL		**HNI**		**HI**		**DI**	
	Rank	**RA**	**Rank**	**RA**	**Rank**	**RA**	**Rank**	**RA**
Erysiphe necator	1	35.1	3	14.9	1	22	1	68.9
Mycosphaerella tassiana	2	17.8	1	26.2	2	20.2	3	7
Filobasidium wieringae	3	13.9	2	15.5	3	19.8	4	6.2
Sporobolomyces roseus	4	9.2	4	9.9	4	10.3	2	7.4
Udeniomyces pyricola	5	4.4	5	6.3	5	5.3	5	1.6
Vishniacozyma victoriae	6	1.9	7	1.9	6	2.6	6	1.2
Erysiphe euonymicola	7	1.6	6	2.3	7	1.7	8	0.6
Dioszegia hungarica	8	1.2	8	1.6	8	1.3	9	0.6
Itersonilia pannonica	9	1	10	1.2	9	1.2	7	0.7
Symmetrospora coprosmae	10	0.8	11	1.2	11	1	16	0.2

Table 7 – List of the species associated with *Erysiphe necator* (Figure 4). Associations are described with the partial correlation provided by PLN and with the stability, which is the fraction of bootstrap subsamples that contained this association. The detection of theses species in other microbial environments (ME) such as old leaves (OL), ground cover (GC) or bark (BK) is indicated or left blank if undetected. Their significant differential abundance in either one of the cropping systems (CS) (*i.e.* cover crop CC or no cover crop NCC) is indicated or left blank if insignificant.

Species	**Correlation**	**Stability**	**Other ME**	**CS**
Dioszegia butyracea	-0.087	0.96	OL / GC	CC
Cladosporium ramotenellum	-0.074	0.94	OL / GC / BK	
Sawadaea bicornis	-0.071	0.96	OL	
Taphrina carpini	-0.059	0.96	OL / GC / BK	CC
Stereum hirsutum	-0.043	0.84		
Buckleyzyma aurantiaca	-0.035	0.82	OL / GC / BK	CC
Erythrobasidium hasegawianum	-0.03	0.76	OL / GC / BK	CC
Taphrina deformans	-0.02	0.56	OL / GC	CC
Erysiphe euonymicola	-0.015	0.58		CC
Fomes fomentarius	-0.011	0.6	BK	
Bullera crocea	-0.011	0.68	OL / GC / BK	CC
Filobasidium wieringae	-0.01	0.52	OL / GC / BK	
Neodevriesia capensis	-0.01	0.38	OL / GC / BK	CC
Vishniacozyma victoriae	-0.005	0.62	OL / GC / BK	CC
Itersonilia perplexans	-0.003	0.38		
Ascochyta manawaorae	0.011	0.52	OL / GC / BK	
Taphrina inositophila	0.017	0.68	OL / GC / BK	
Vishniacozyma dimennae	0.02	0.64	OL / GC / BK	
Botrytis caroliniana	0.035	0.86	OL / GC / BK	
Apiotrichum domesticum	0.038	0.8		
Filobasidium stepposum	0.055	1		
Bulleromyces albus	0.149	1	OL / GC / BK	
Pseudopithomyces chartarum	0.158	1	OL / GC / BK	

Table 8 – Literature mining of the pathobiome. The following papers were reported by the text-mining approach because both the pathogen *Erysiphe necator* and the name of the potential antagonists presented in the Figure 4 were found in abstracts, titles, keywords or references of the Scopus database. This search also included fungal synonyms through a custom script (Methods 6.2).

Pathogen	Potential antagonist searched for	References
Erysiphe necator	*Erythrobasidium hasegawianum*	Mueller, G. M., Bills, G. F., & Foster, M. S. (Eds.). (2004). *Biodiversity of fungi : inventory and monitoring methods*. Amsterdam ; Boston : Elsevier.
Erysiphe necator	*Fomes fomentarius*	Tabata, J., De Moraes, C. M., & Mescher, M. C. (2011). Olfactory Cues from Plants Infected by Powdery Mildew Guide Foraging by a Mycophagous Ladybird Beetle. *PLoS ONE, 6*(8), e23799. doi : 10.1371/journal.pone.0023799
Erysiphe necator	*Sawadaea bicornis*	Kiss, L. (1998). Natural occurrence of ampelomyces intracellular mycoparasites in mycelia of powdery mildew fungi. *New Phytologist, 140*(4), 709–714. doi : 10.1046/j.1469-8137.1998.00316.x
Erysiphe necator	*Taphrina deformans*	Chang, H.-X., Noel, Z. A., Sang, H., & Chilvers, M. I. (2018). Annotation resource of tandem repeat-containing secretory proteins in sixty fungi. *Fungal Genetics and Biology, 119*, 7–19. doi : 10.1016/j.fgb.2018.07.004
Erysiphe necator	*Taphrina deformans*	Kües, U., Khonsuntia, W., & Subba, S. (2018). Complex fungi. *Fungal Biology Reviews, 32*(4), 205–218. doi : 10.1016/j.fbr.2018.08.001
Erysiphe necator	*Taphrina deformans*	Caubel, J., Launay, M., Lannou, C., & Brisson, N. (2012). Generic response functions to simulate climate-based processes in models for the development of airborne fungal crop pathogens. *Ecological Modelling, 242*, 92–104. doi : 10.1016/j.ecolmodel.2012.05.012
Erysiphe necator	*Taphrina deformans*	Weete, J. D., Abril, M., & Blackwell, M. (2010). Phylogenetic Distribution of Fungal Sterols. *PLoS ONE, 5*(5), e10899. doi : 10.1371/journal.pone.0010899
Erysiphe necator	*Taphrina deformans*	Mysyakina, I. S., & Funtikova, N. S. (2007). The role of sterols in morphogenetic processes and dimorphism in fungi. *Microbiology, 76*(1), 1–13. doi :10.1134/S0026261707010018
Erysiphe necator	*Taphrina deformans*	Hernandez, A., Cooke, D. T., Lewis, M., & Clarkson, D. T. (1997). Fungicides and sterol-deficient mutants of Ustilago maydis : plasma membrane physico-chemical characteristics do not explain growth inhibition. *Microbiology, 143*(10), 3165–3174. doi :10.1099/00221287-143-10-3165

Table 9 – List of the species differentially abundant between cropping systems. Species were assigned to a preferred cropping system : either no cover crop (NCC) or cover crop (CC) depending on the sign of the effect size generated by ALDEx2. Selection of the species was based on the false discovery rate (FDR) after Benjamini-Hochberg correction using the default threshold of ALDEX2 (FDR < 0.1).

Species	CS	Effect size	FDR (BH)
Erysiphe necator	NCC	0.45	2.80E-07
Udeniomyces pyricola	CC	-0.17	6.90E-02
Taphrina caerulescens	CC	-0.17	7.10E-02
Blumeria graminis	CC	-0.18	6.10E-02
Symmetrospora gracilis	CC	-0.21	4.00E-02
Filobasidium oeirense	CC	-0.22	9.30E-02
Articulospora proliferata	CC	-0.25	2.60E-02
Gelidatrema spencermartinsiae	CC	-0.26	3.10E-02
Dioszegia butyracea	CC	-0.26	3.50E-02
Erythrobasidium hasegawianum	CC	-0.26	1.30E-02
Limonomyces culmigenus	CC	-0.27	1.60E-02
Vishniacozyma victoriae	CC	-0.3	4.90E-04
Dioszegia hungarica	CC	-0.3	4.80E-04
Neodevriesia capensis	CC	-0.3	3.00E-02
Taphrina deformans	CC	-0.31	1.60E-03
Bullera crocea	CC	-0.32	6.10E-03
Erysiphe euonymicola	CC	-0.39	1.70E-07
Buckleyzyma aurantiaca	CC	-0.46	2.30E-05
Curvibasidium cygneicollum	CC	-0.46	5.60E-05
Mycosphaerella tassiana	CC	-0.51	6.80E-08
Taphrina carpini	CC	-0.51	7.20E-06
Cystofilobasidium macerans	CC	-0.55	5.10E-07
Mycosphaerella punctiformis	CC	-0.57	3.50E-07
Angustimassarina acerina	CC	-0.62	1.70E-08
Itersonilia pannonica	CC	-0.74	5.40E-12

6.1 Methods S1

Species selection needs to be done prior to network inference (Röttjers et Faust, 2018). However, there is reluctance to filter out infrequent species, or rare, that could be specialists or possess an important role (Harrison et al., 2019 ; Lynch et Neufeld, 2015). Röttjers et Faust (2018) have highlighted the importance of removing infrequent ASV prior to network inference, as they can cause spurious associations due to matching zeros. Others have previously indicated the removal of infrequent species but without consensus on the prevalence threshold at which to keep species : present in at least 20% of the samples (Berry et Widder, 2014), at least in 2 (Durán et al., 2018), 5 (Poudel et al., 2016) or 10 samples (Agler et al., 2016). Indeed, the more sparsity in the ASV table, the less reliable are the inferred networks and no more than 50% of zeroes was suggested (Weiss et al., 2016). A recent study did not trim species but suggest instead that associations between pairs of rare species should not be inferred but those between more abundant species and rare species could be (Cougoul et al., 2019a). They provide a test based on the prevalence of the two species to determine whether the association could be inferred and their approach improved the inference of networks according to simulations (Cougoul et al., 2019a).

However, the method we used to infer the network at the time did not allow for discarding associations *a priori* and left us with the option of trimming species. But we decided to evaluate the testability of the associations to guide us for species trimming. Moreover, we performed the species selection for each experimental unit (Figure 1D) separately to ensure that both ubiquitous and locally-distributed species were kept.

Therefore, we assessed the testability of every ASV pairs in the table for different minimum values of ASV prevalence. We started from 1 sample minimum and increased to the maximum number of samples containing two ASV. We used the fraction of testable pairs over the total number of pairs as a trade-off criterium to identify the threshold. Keeping species in the network inference if they were detected in 20% of the samples considered, seemed to well balance in each experimental unit (Figure 10) the testability of associations and the number of species. The final network inference was conducted with the ASV table of all samples but only with the species present in at least 20% of the samples of any experimental unit.

6.2 Methods S2

```
### TEXT-MINING OF PUTATIVE ASSOCIATIONS WITH SYNONYMS SEARCH
### Charlie Pauvert
### 2019-05-31

### SYNONYMS SEARCH

# Extracting fungal synonyms from MycoBank (unpublished)
# based on the API documented here
# http://www.mycobank.org/Services/
#  Generic/Help.aspx?s=searchservice
get_synonyms<-function(species){
```

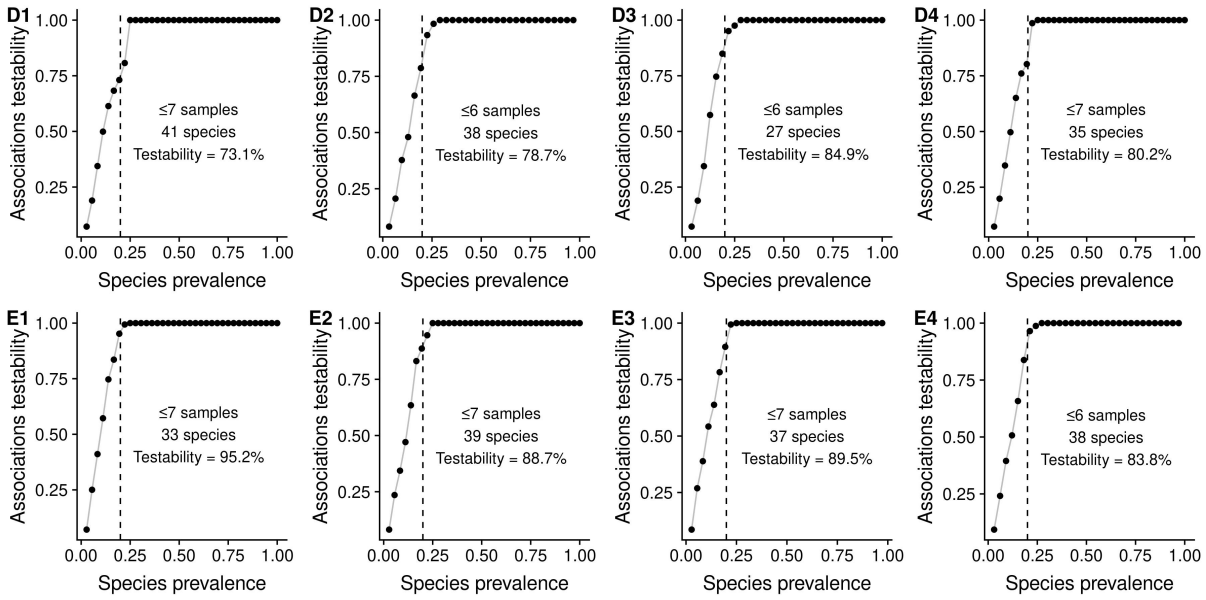


Figure 10 – Fraction of testable associations depending on species prevalence in each experimental unit. The dashed vertical line indicates the species prevalence threshold of 20%, which appeared as a good trade-off between testability and species number. Several metrics are provided (minimum number of samples, final number of species and testability value) for this threshold

```
  require(xml2)
  base_url<-paste("http://www.mycobank.org/",
                  "Services/Generic/SearchService.svc/",
                  "rest/xml?layout=14682616000000161&filter=name CONTAINS")
  # Sanitize species name and url
  url<-paste0(base_url,' "',gsub("_"," ",species),'"') # Add the quotes!
  url<-gsub(" ","%20", url)
  message("Checking for ", species)
  # Fetch XML from MycoBank
  x<-read_xml(url)
  # Get the obligate synonyms
  oblig<-xml_text(xml_find_first(x,"//obligatesynonyms_pt_"))
  anamo<-xml_text(xml_find_first(x,"//anamorph_pt_"))
  # Extract the name
  f<-function(x) gsub(".*<Name>([A-Z][a-z]+ [a-z]+)</Name>.*","\\1",x)
  # Output
  c("CurrentName"=gsub("_"," ",species),
    "ObligateSyn"=f(oblig),
    "AnamorphSyn"=f(anamo)
  )
}
# List of lists of synonyms! Empty strings when unavailable
# The fungal name can have space or underscore

(examples.syn<-lapply(
c("Erysiphe_necator","Mycosphaerella tassiana"), get_synonyms))
```

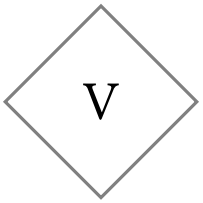
```
# [[1]]
# CurrentName        ObligateSyn        AnamorphSyn
# "Erysiphe necator" "Uncinula necator"   "Oidium tuckeri"
#
# [[2]]
# CurrentName                ObligateSyn                AnamorphSyn
# "Mycosphaerella tassiana"      "Davidiella tassiana" "Helmisporium vesiculosum"

# Function to restore the current name if looking at a synonym
get_current<-function(sp,synonyms){
  # Find in which sublists the species belongs
  i<-sapply(synonyms, function(bar) sp %in% bar)
  # and get its name
  synonyms[i][[1]][["CurrentName"]]
}
get_current("Oidium tuckeri",examples.syn)
# [1] "Erysiphe necator"

### TEXT-MINING OF PUTATIVE ASSOCIATIONS
# Construct a List of lists of synonyms!
# Empty strings when unavailable
patho.syn<-lapply(V(patho.net)$name, get_synonyms)
# Reduce to a single vector
patho.syn.vec<-Reduce(c,patho.syn) %>% unique() %>% .[.!=""]
# Generate pairs between (Erysiphe necator and its synonyms) AND the others
patho.pairs<-expand.grid(patho.syn.vec[1:3],patho.syn.vec[-c(1:3)],
                         KEEP.OUT.ATTRS = F,stringsAsFactors = F)

# Basd on rscopus library
# to be installed using devtools::install_github("muschellij2/rscopus")
library(rscopus)
rscopus::set_api_key("INSERT SCOPUS API KEY HERE")
# Needs a data.frame Var1 Var2 with all the combinations
#  of species wanted to be searched for
text_cooccurrence<-function(pair){
  # Abbreviate the species name like E. necator
  i<-function(x) gsub("([A-Z])[a-z]+ ([a-z]+)","\\1. \\2",x)
  # Build query
  query<-paste0('ALL("',pair$Var1,'" OR "',i(pair$Var1),'") ',
                'AND ALL("',pair$Var2,'" OR "',i(pair$Var2),'")')
  message("Querying ",pair$Var1," vs. ",pair$Var2)
  r_search <- scopus_search(query = query,view = "STANDARD",
                  field = "dc:title,prism:doi",verbose = F,
                  max_count = 25,count = 25)
  gen_entries_to_df(r_search$entries)$df
}
# Apply the search function to all pairs constructed earlier
patho.tm.scopus<-adply(patho.pairs,1, text_cooccurrence)
# Rename back the species to current name to build a summary of the search
```

```
patho.tm.scopus %>%
  mutate(
    Var1 = gsub(" ","_",sapply(Var1,get_current,synonyms = patho.syn)),
    Var2 = gsub(" ","_",sapply(Var2,get_current,synonyms = patho.syn))
    ) %>% group_by(Var1,Var2) %>%
    dplyr::summarise(  n = length(na.omit(unique(‘prism:doi‘)))) %>%
  ungroup() %>% filter(n > 0) %>% data.frame()
```



Conclusion et perspectives

1 Points forts et limites des approches utilisées dans la thèse

Les résultats de cette thèse soulignent plusieurs points importants à considérer pour reconstruire les réseaux d'interactions entre micro-organismes associés aux plantes. Le chapitre II (p. 35) a mis en évidence que différents pipelines bioinformatiques produisent différents relevés des micro-organismes. Ces résultats sont importants car la détermination des micro-organismes – futurs nœuds – est un prérequis à la construction de réseaux microbiens basés sur ces relevés. Mais choisir une approche bioinformatique dépend du critère à optimiser : la sensibilité de détection des espèces, un faible taux de faux positifs ou encore une composition similaire à la composition réelle. Le chapitre II propose un pipeline bioinformatique performant pour les trois critères évalués. À l'aide de ce pipeline, j'ai pu établir avec confiance un relevé des champignons sur les feuilles de vigne dans le chapitre III (p. 60). J'ai pu mettre en évidence que les communautés fongiques sont influencées par la pratique culturale. Mais que cet effet n'est pas détecté sur les propriétés des réseaux microbiens. Les réseaux d'associations étaient distincts entre pratiques culturales mais ils étaient surtout distincts entre parcelles d'une même pratique indiquant une faible réplicabilité. Un nouveau réseau microbien a été reconstruit dans le chapitre IV (p. 85) en augmentant le nombre d'échantillons et en mesurant des covariables environnementales pouvant expliquer les variations de comptages de séquences. Ce réseau inféré a généré des hypothèses d'interactions entre champignons grâce à une nouvelle méthode – PLN – apte à inclure ces covariables. Deux associations prédites ont pu être confirmées par co-cultures. Par exemple, la levure *Buckleyzyma aurantiaca* a réduit la colonisation de l'agent causal de l'oïdium de la vigne et même modifié sa morphologie. Ces résultats confirment une corrélation négative partielle, toutes choses égales par ailleurs, entre *B. aurantiaca* et *E. necator* obtenue avec PLN. Ceci suggère que les signaux d'interactions entre micro-organismes peuvent être détectés par les méthodes d'inférence. L'ensemble de ces résultats est prometteur pour l'étude des pathobiomes des plantes cultivées. Les points forts et les limites des approches utilisées dans la thèse sont discutés dans les paragraphes suivants.

Les communautés artificielles – *mock* – sont un atout majeur pour les approches de métabarcoding où les multiples décisions prises ont une influence sur la communauté microbienne observée (e.g., Alberdi et al., 2018 ; Vorholt et al., 2017). Elles permettent de contourner les biais documentés – liés à l'échantillonnage ou l'extraction et l'amplification de l'ADN – afin de se concentrer sur une étape clé : les pipelines bioinformatiques. Nguyen et al. (2015) recommandent que les communautés artificielles assemblent des espèces éloignées pour couvrir un large éventail taxonomique. La communauté artificielle du Chapitre II (p. 35) est relativement complexe et réaliste avec 181 espèces fongiques couramment détectés dans les sols agricoles et forestiers ainsi que dans les tissus végétaux. Elle est d'une complexité intermédiaire entres les communautés artificielles plus petites (10 – 30 espèces, e.g. Amend et al., 2010 ; Bakker, 2018 ; Nguyen et al., 2015) – et les échantillons environnementaux (10^2 – 10^3 espèces, Peay et al., 2016). Les petites communautés artificielles permettent plus facilement des analyses complémentaires telles que la qPCR pour identifier le nombre de copies de l'ITS et pouvoir vérifier l'homogénéité des concentrations d'ADN avant le séquençage (Bakker, 2018). Mais la forte variabilité génétique des marqueurs moléculaires bactériens et fongiques (Edgar, 2018 ; Vu et al., 2019) nécessite des communautés artificielles qui imitent – même partiellement – une

telle complexité (Vorholt et al., 2017). Sur les 181 espèces de la communauté artificielle utilisée dans le chapitre II, 8 espèces ont plus d'une souche représentative ce qui permet d'introduire de la complexité. Les ADN de la communauté artificielle sont malheureusement épuisés ce qui limite son partage pour être réutilisée par d'autres (mais voir Bakker, 2018). Toutefois, il est possible de repartir des séquences qui sont disponibles publiquement. Il serait même possible d'intégrer les séquences de cette mock complexe dans le *mockrobiota* (Bokulich et al., 2016). Cette ressource publique met à jour une liste d'une dizaine de communautés artificielles de domaines variés – bactéries, archées et eucaryotes – analysées par séquençage à haut-débit de marqueurs moléculaires (Bokulich et al., 2016). La comparaison des pipelines aurait pu se faire avec une de ces communautés artificielles. Néanmoins, les deux seules communautés artificielles fongiques disponibles dans mockrobiota au début de la thèse correspondaient à des séquences d'ITS obtenues avec la technologie Illumina HiSeq générant des séquences plus courtes que le MiSeq utilisé dans la thèse (Goodwin et al., 2016). La communauté artificielle du chapitre II était plus adaptée à l'étude des pathobiomes des plantes et comprenait notamment le champignon *Erysiphe necator*.

Le marqueur moléculaire fongique de référence – ITS – a été utilisé dans les trois chapitres (Schoch et al., 2012). Sa qualité de référence implique la disponibilité de banques de séquences fournies et toujours croissantes, facilitant la comparaison entre les études (Hibbett et al., 2016 ; Kõljalg et al., 2013 ; Yahr et al., 2016). Le couple d'amorces utilisé – ITS1F-ITS2 – amplifient certes préférentiellement les séquences d'ITS1 taxons ciblés – les Ascomycota plus que les Basidiomycota – et d'autres non ciblés comme les plantes (Bellemain et al., 2010). Néanmoins, ces phylums fongiques sont ceux qui prédominent les communautés associées aux parties aériennes des plantes (Hardoim et al., 2015 ; Peay et al., 2016). L'analyse de milliers de séquences d'ITS expertisées indiquent que le marqueur ITS1 est résolutif à l'espèce grâce à des variations de quelques nucléotides (Vu et al., 2019). Notre approche dans cette thèse a exploité correctement cette variabilité génétique en analysant les séquences obtenues avec une méthode – DADA2 – adaptée à cette finesse (Callahan et al., 2016).

Les approches basées sur DADA2 pour l'analyse de variations des séquences se classaient parmi celles avec le moins d'ASV faux-positifs et avec les richesses et compositions les plus similaires à celle de la communauté artificielle d'après le comparatif du chapitre II. Ce comparatif ambitieux a évalué 360 combinaisons unique d'outils et de paramètres mais ne couvre pas hélas la myriade d'outils bioinformatiques développés pour l'analyse des séquences issues du métabarcoding. En effet, le comparatif n'incluait pas d'autres outils comme Swarm (Mahé et al., 2014 ; Mahé et al., 2015), ou des pipelines complets comme mothur (Schloss et al., 2009). Anslan et al. (2018) ont évalué les performances de pipelines complets – tels que QIIME 2 ou PIPITS (Bolyen et al., 2019 ; Gweon et al., 2015) – sur des séquences fongiques environnementales et concluent également à des différences importantes de richesses entre pipeline à partir du même jeu de données. Ils suggèrent pour limiter les différences de ne conserver que les ASV (ou OTU) avec des assignations fongiques ou non-artéfactuelles, avec plus de 10 séquences par échantillon et les ASV (ou OTU) conservés après la procédure LULU (Anslan et al., 2018). Néanmoins la richesse réelle des communautés séquencées est inconnue dans leur cas et nos résulats indiquent que des filtres trop drastiques omettent des ASV (ou OTU) présentes dans la communauté (Chapitre II). Parmi les différentes approches qui génèrent des ASV, nous n'avons inclus que DADA2 dans notre comparatif du chapitre II, mais les résultats mis en évidence dans ce chapitre sont confirmées par une autre étude basée sur une communauté artificielle bactérienne. Nearing et al. (2018) ont en effet noté que la richesse, c'est-à-dire

le nombre d'ASV ou d'OTU, était très dépendante de l'approche bioinformatique utilisée. Ils confirment également que les méthodes basées sur l'agrégation des séquences en OTU sont une approche plus sensible que les méthodes basées sur les ASV (Nearing et al., 2018). Ces résultats soulignent qu'il n'existe pas de pipeline "miracle" mais que le critère à maximiser dans l'étude – une détection sensible, précise ou une communauté similaire – doit guider le choix du pipeline (Chapitre II).

Dans cette thèse, les communautés fongiques ont été caractérisées en priorité. Ces communautés comprennent les pathogènes majoritaires de la vigne (Armijo et al., 2016) d'où l'importance d'évaluer la réponse à un changement de pratique culturale dans ces communautés (Chapitre III). L'empreinte de la pratique culturale detectée sur les communautés fongiques est en accord avec celles de la littérature (Castañeda et al., 2018 ; Kernaghan et al., 2017 ; Pancher et al., 2012 ; Schmid et al., 2011 ; Varanda et al., 2016). Mais les communautés bactériennes associées peuvent aussi être impactées par les pratiques viticoles, ce qui peut influencer les réseaux microbiens (Burns et al., 2016 ; Campisano et al., 2014 ; Kecskeméti et al., 2016 ; Likar et al., 2017 ; Lupatini et al., 2017). De plus, les communautés fongiques étaient robustes à l'infection par *E. necator* (Chapitre IV) mais d'autres communautés microbiennes peuvent y répondre différemment. D'autant plus que les champignons responsables de l'oïdium chez plusieurs plantes sont supposés interagir avec d'autres micro-organismes (Panstruga et Kuhn, 2019). Par exemple, les bactéries associées aux feuilles sont sujettes à des perturbations lors de l'infection par l'oïdium du concombre (Suda et al., 2009). De même, la baisse de virulence déclenchée par certains virus – notamment le genre *Mitovirus* – infectant les champignons pathogènes des plantes (Nuss, 2005) pourrait être une alternative à explorer dans le biocontrôle d'*Erysiphe necator* (Pandey et al., 2018) et plus généralement des champignons phytopathogènes des plantes (Roossinck, 2019). Cependant, peu d'études incluent les virus dans l'inférence de réseau (mais voir Chow et al., 2014). Il n'existe pas de marqueur conservé entre les nombreuses espèces de virus qui pourrait être ciblé lors du séquençage (Roossinck, 2019). Certains motifs protéiques détectés lors du séquençage non ciblé de l'ADN (*shotgun*) pourraient servir d'OTU virales (Lefebvre et al., 2019). D'autres études ont initialement inclus des données de métabarcoding issues de plusieurs règnes du vivant (bactéries, champignons et oomycètes) pour construire un réseau de corrélation inter-règnes (Agler et al., 2016 ; Jakuschkin et al., 2016). Durán et al. (2018) ont montré expérimentalement que ces interactions inter-règnes étaient fondamentales pour la survie de la plante. Malgré ces résultats encourageants, il est impératif de tenir compte de la différence de séquençage entre les règnes dans l'inférence de réseaux. Les marqueurs moléculaires ont des longueurs différentes par règne, ce qui implique souvent des runs de séquençage distincts. Or les différences entre runs de séquençage existent (Song et al., 2018) et les variations d'effort de séquençage doivent être prises en compte avant d'inférer des associations inter-règnes. Les auteurs de SpiecEasi (Kurtz et al., 2015) ont récemment implémenté cette fonctionnalité, et mis en évidence que les nœuds fongiques stabilisaient les réseaux bactériens chez l'homme (Tipton et al., 2018). La méthode d'inférence PLN (Chiquet et al., 2017 ; Chiquet et al., 2019) peut nativement inclure les différences d'effort de séquençage entre espèces et donc prendre en compte les différences possibles entre les communautés issues de règnes distincts. La combinaison de règnes microbiens implique néanmoins un décalage de résolution taxonomique entre les différents marqueurs utilisés nécessitant potentiellement d'agglomérer les nœuds au genre par exemple pour unifier l'analyse (Röttjers et Faust, 2018).

Les analyses des diversités α et β classiquement appliquées aux communautés ont également été appliquées dans le chapitre III aux réseaux en suivant le cadre théorique de comparaisons

des réseaux (Pellissier et al., 2018 ; Poisot et al., 2012). La comparaison des liens – diversité β – permet d'aller au-delà de la comparaison des propriétés des réseaux – diversité α – plus fréquemment utilisés (e.g., Durand et al., 2017 ; Faust et al., 2015 ; Karimi et al., 2019). De plus, cette comparaison a mis en évidence une réplicabilité faible entre les réseaux inférés au sein de la même pratique culturale (Chapitre III). Les réseaux sont fréquemment reconstruits à partir de tous les échantillons de l'étude (e.g., Agler et al., 2016 ; Durán et al., 2018) ou tous ceux d'un compartiment (e.g., Durand et al., 2017 ; Kerdraon et al., 2019) et il existe peu d'études avec des réplicats de réseaux (mais voir Karimi et al., 2019). Dans le chapitre III, des réseaux répliqués ont été reconstruits à partir d'échantillons – une feuille par cep – issus de parcelles de vigne proches. Il est possible que la faible réplicabilité des réseaux mise en évidence provienne d'une variabilité des comptages de séquences entre feuilles d'un même bloc. En effet, une variabilité intra-parcelle dans les communautés de levures des grappes de vigne a déjà été mise en évidence (Setati et al., 2012). Une telle variabilité pourrait induire de la variabilité dans les comptages de séquences obtenues dans le chapitre III et donc une variabilité des réseaux. Deux analyses possibles permettraient de préciser si un tel effet existe. L'utilisation du modèle HMSC permettrait de partitionner la variance des comptages à l'échelle de la parcelle et de la pratique indiquant potentiellement une forte contribution de la parcelle à la variance (Annexe 2 et Ovaskainen et al., 2017). Il serait possible de reconstruire les réseaux différemment par exemple avec PLN pour évaluer cet effet. Si l'on postule que la structure de dépendance est similaire au sein d'une pratique culturale, on pourrait regrouper les échantillons issus des trois parcelles pour réaliser l'inférence de réseau. Avec PLN on peut introduire une covariable "parcelle" pour contrôler pour les différences de comptages moyens entre parcelles et générer un réseau par pratique culturale. Dans le cas où le coefficient "parcelle" est très variable pour toutes les espèces incluses dans le réseau cela pourrait indiquer que le manque de réplicabilité observée avec SparCC est dû à la trop grande fluctuation des comptages des espèces entre les parcelles.

L'absence de réplicabilité pourrait être due à un nombre limité d'échantillons par réseau. Cordero et Datta (2016) soulignent qu'il est nécessaire d'avoir un maximum d'échantillons répliqués et à la plus petite échelle possible pour mettre en évidence des associations se rapprochant des interactions. Le nombre minimum d'échantillons nécessaires pour reconstruire des réseaux a été évalué par simulations et les auteurs indiquent qu'avec moins de 25 échantillons le nombre de liens faux positifs s'accroît (Berry et Widder, 2014). Dans le chapitre III, le point fort a été de réaliser des réseaux répliqués mais le nombre d'échantillons par réseau est devenu légèrement en deçà de cette limite. L'inférence réalisée dans le chapitre IV est plus robuste et provient de 276 échantillons ce qui place cette étude dans la gamme des recommandations récentes : $n > 200$ pour un réseau de 50 nœuds (Hirano et Takemoto, 2019).

Dans cette thèse, les deux méthodes d'inférence de réseaux utilisées – SparCC et PLN (Chiquet et al., 2019 ; Friedman et Alm, 2012) – ont permis de surmonter certaines des barrières à la reconstruction de réseaux (voir 2.2.2, p. 20). La première a été choisi sur les recommandations d'une comparaison de réseaux et pour sa robustesse à la compositionnalité des données (Weiss et al., 2016). La seconde méthode plus récente est également robuste à la compositionnalité, produit des corrélations partielles entre les comptages et permet surtout d'inclure des covariables environnementales (Chiquet et al., 2019).

L'inférence robuste réalisée avec PLN a généré des hypothèses d'interactions entre champignons qui ont pu être confrontées à des expériences de co-cultures dans le chapitre IV. Tester les

hypothèses avec des expériences supplémentaires est une force pour aller au-delà des réseaux corrélatifs (Carr et al., 2019). Or les ASV obtenus sont des unités bioinformatiques et ne sont pas des micro-organismes cultivables (McLaren et Callahan, 2018). Mais leur assignations jusqu'au niveau de l'espèce grâce au RDP Classifier et à la banque de référence UNITE nous a permis de relier tout de même ASV et souches fongiques cultivables (Kõljalg et al., 2013 ; Wang et al., 2007). Les souches utilisées dans les co-cultures ont été commandées depuis la souchothèque du Westerdijk Institute et pourrait être différente de celles associées au feuilles de vigne du chapitre IV. Ce décalage entre les souches peut entraîner des différences dans leur adaptation aux feuilles, leur métabolisme et leurs potentiels traits liés au biocontrôle. Afin de limiter ce décalage, l'inférence de réseau dans le chapitre IV a été réalisé après avoir agglomeré les ASV au niveau de l'espèce en se basant sur la taxonomie. Cette étape va néanmoins à l'encontre de recommandations sur la précision taxonomique nécessaire dans l'inférence de réseaux (Berry et Widder, 2014) et met de côté de plus les comptages potentiellement informatifs d'ASV avec des assignations seulement au genre.

L'échantillonnage dans le chapitre IV a été conduit pour mettre en évidence de potentiels agents de biocontrôle. Ainsi des feuilles ont été prélevées sur des hôtes sains et malades conformément aux recommandations (Ellis, 2017 ; Poudel et al., 2016). De plus une inoculation en parcelle a été réalisée, combinant les avantages des approches de manipulations et des approches *in planta*. Une date supplémentaire d'échantillonnage permettrait de caractériser la communauté juste avant l'inoculation du pathogène et évaluer la robustesse des communautés fongiques à l'infection. La collecte de deux échantillons dans les tissus sains des feuilles sans symptômes visuels permettrait d'évaluer les différences de communautés au sein d'une même feuille. Néanmoins, l'inférence de réseau réalisée avec cet échantillonnage a mis en évidence une association négative entre *B. aurantiaca* et *E. necator* confirmée par l'effet négatif sur la croissance et la morphologie de *E. necator* (Chapitre IV).

Cette levure *B. aurantiaca* est ubiquiste et colonise des milieux aquatiques, des sédiments ainsi que racines et feuilles (Buzzini et al., 2017). Les résultats encourageants présentés dans le chapitre IV permettent de considérer de nouveau rôle pour cette levure. En effet, les antagonistes potentiels ne se limitent pas à cette levure. Le réseau inféré a également identifié la levure *Vishniacozyma victoriae* comme un antagoniste potentiel bien que les expériences de co-cultures ne le confirme pas (Chapitre IV). Les fonctions de cette levure sont plus documentés que *B. aurantiaca*. La levure *V. victoriae* est considéré comme une levure foliaire et ubiquiste (Buzzini et al., 2017 ; Glushakova et Chernov, 2010), elle est également détectée à la surface des baies de raisins (Martins et al., 2014) et associée aux glands du chêne (Annexe 2). La levure *V. victoriae* est utilisée comme agent de lutte biologique contre les maladies causées par des champignons après la récolte des fruits (Lutz et al., 2013) et inhibe moyennement la croissance de champignons filamenteux (Hilber-Bodmer et al., 2017). D'ailleurs, une association négative a été détectée de façon robuste dans le chapitre III entre *V. victoriae* et *Neofusicoccum parvum* un champignon associé aux maladies du bois de la vigne (Bruez et al., 2014). Ce qui suggère que les réseaux microbiens pourraient détecter des signaux d'interactions biotiques utile au biocontôle (Poudel et al., 2016).

2 Pistes de recherche futures

2.1 Quel échantillonnage pour inférer les réseaux microbiens ?

Plusieurs auteurs ont insisté sur l'importance d'échantillonner à l'échelle appropriée pour les interactions entre micro-organismes (Armitage et Jones, 2019 ; Cordero et Datta, 2016). Mais cette échelle est difficile à déterminer. Une étude pionnière a déterminé l'échelle à laquelle deux espèces de bactéries se partagent la surface des feuilles (Esser et al., 2015) et indiquent que la sphère d'influence des bactéries est estimée à environ 10 μm (Remus-Emsermann et Schlechter, 2018). Mais ce type d'étude reste à être appliqué aux champignons associés aux plantes. Lodge et al. (1996) ont montré que des espèces très différentes de champignons endophytes dominaient des rectangles de 1×2 mm en utilisant des méthodes de cultures, suggérant une hétérogénéité des communautés fongiques à la surface de la feuille. Le champignon *Aureobasidium pullulans* colonise en effet principalement les veines de la feuille mais recouvre ensuite la totalité de la feuille (McGrath et Andrews, 2006). Cette colonisation serait plutôt due à des substances promouvant la croissance dans ces régions qu'à une incapacité à coloniser ou croître ailleurs (McGrath et Andrews, 2007). Déterminer l'échelle des interactions sur les feuilles pourrait se faire en combinant ces études avec la méthodologie de Esser et al. (2015) en ajoutant d'autres champignons en compétition avec *A. pullulans*. Ce champignon est en effet capable de fortes compétitions pour l'espace et les nutriments (Bozoudi et Tsaltas, 2018 ; Pertot et al., 2017). Néanmoins, les échantillons prélevés sur des tissus végétaux pour être séquencés sont de l'ordre de quelques millimètres (e.g., chapitres III et IV ; Kernaghan et al., 2017). Échantillonner en deçà du millimètre pour se rapprocher des échelles supposées d'interactions pose des problèmes techniques. À ces échelles microscopiques, la quantité de micro-organismes est mécaniquement moindre ce qui complique l'obtention de quantité suffisante d'ADN microbien pour le séquençage, au risque de devoir multiplier les cycles d'amplification et *in fine* le nombre d'erreurs induites par la polymérase. De plus, la faible concentration d'ADN complique la caractérisation de la communauté en augmentant le risque de contamination (Salter et al., 2014). En diluant des mocks bactériennes pour imiter de faibles concentrations d'ADN, Caruso et al. (2019) montrent que les approches qui produisent des ASV ne sont pas suffisantes pour retirer les contaminants. Mais ces approches distinguent mieux les séquences de contaminants des séquences erronées par rapport aux approches produisant des OTU ce qui facilite l'utilisation d'outils (voir p. 17) pour les identifier et les retirer (Caruso et al., 2019). L'échantillonnage actuel des communautés (quelques millimètres) constitue un bon compromis quantité d'ADN et représentativité dans l'attente de progrès technologique.

L'utilisation de données temporelles est souvent proposée comme une approche permettant d'inférer les interactions entre micro-organismes (e.g., Fisher et Mehta, 2014 ; Gao et al., 2018 ; Mounier et al., 2008 ; Venturelli et al., 2018). Mais ces études ne se basent pas sur des communautés microbiennes associées aux plantes. Bien que ces dernières soient également caractérisées dans le temps, la fréquence d'échantillonnage adaptée est inconnue (e.g., Copeland et al., 2015 ; Hertz et al., 2016 ; Shade et al., 2013). Or cette fréquence doit suivre les dynamiques microbiennes et les interactions qui les maintiennent et tout décalage pourrait révéler des

associations trompeuses (Carr et al., 2019). La colonisation de champignons endophytes a récemment été suivie sur des disques foliaires à l'aide d'outils de microscopie (Huang et al., 2018). Différentes dynamiques ont été mises en évidence : la souche de *Cladosporium* s'est développée sur les disques dès 6 heures après l'inoculation, alors que les souches de *Penicillium* et de *Trichoderma* se sont respectivement développées après 48 et 72 heures (Huang et al., 2018). Même si une fréquence d'échantillonnage adéquate était déterminée, les protocoles actuels détruisent les tissus végétaux pour l'extraction d'ADN. Ceci limite fortement le suivi temporel des communautés associées aux plantes pour mettre en évidence des interactions. Cette destruction limite également l'utilisation d'analyses complémentaires, telles que la microscopie sur les mêmes échantillons. Choisir la stratégie d'échantillonnage idéal des communautés associées aux plantes, même destructive, nécessiterait une comparaison détaillée entre multi-site ou temporelle (Figure 6, p. 21).

Quelque soit la stratégie d'échantillonnage utilisée un nombre suffisant d'échantillons (minimum 25 et idéalement 200) est nécessaire pour reconstruire les réseaux (Berry et Widder, 2014 ; Hirano et Takemoto, 2019). Les résultats du chapitre III vont également dans le sens de cette recommandation puisque un nombre limite d'échantillons pourrait expliquer la faible replicabilité observée. Bien que les échantillons doivent être suffisamment homogènes (Berry et Widder, 2014), les comptages de séquences doivent être également fluctuants et contrastés pour pouvoir détecter des signaux d'interactions (Armitage et Jones, 2019). Ces recommandations sont faisables dans un run MiSeq de séquençage d'échantillons environnementaux qui inclurait aussi des témoins positifs et négatifs (voir chapitre IV).

2.2 Comment réduire les biais moléculaires ?

L'identification précise des nœuds du réseau est primordiale. Le marqueur moléculaire utilisée nécessite donc la plus fine résolution taxonomique possible. L'ITS et le 16S constituent actuellement les marqueurs fongiques et bactériens – principalement via leurs régions courtes et variables respectives – de référence et les banques sont alimentées de milliers de séquences (Kõljalg et al., 2013 ; Quast et al., 2012). Augmenter la longueur des séquences de ces marqueurs promet une meilleure identification taxonomique pour les ITS fongique (Motooka et al., 2017 ; Wurzbacher et al., 2018) et les bactéries (Callahan et al., 2019) d'après des comparaisons sur des communautés artificielles. Ces longueurs étendues sont obtenues avec des technologies de séquençage de troisième génération tels que le MinION ou le PacBio (Goodwin et al., 2016). Mais elles souffrent de taux de chimères ou d'erreurs élevés, et leur faible débit a même été critiqué comme une limite à l'inférence d'hypothèses robustes (Kennedy et al., 2018). Le faible nombre d'échantillons ou de séquence obtenues entrave l'inférence de réseau. Ce problème difficile nécessite de se baser sur des technologies de séquençage déjà éprouvées. Une comparaison de réseaux inférés à partir de séquences PacBio ou MinION *versus* Illumina serait un plus pour aider à la décision. Néanmoins, la précision taxonomique des approches de troisième génération les rendent toutes indiquées plutôt pour identifier quelques espèces, par exemple pour caractériser rapidement des pathogènes de plantes (Loit et al., 2019) que pour l'inférence de réseau.

La région ITS complète n'est pas accessible avec le séquençage MiSeq (Tedersoo et Lindahl, 2016). Néanmoins la région ITS2 est recommandée comme une alternative valide pour obtenir une taxonomie proche de celle obtenue avec l'ITS complet (Yang et al., 2018) et les amorces

suivantes – gITS7ngs ou ITS3ngsmix en forward et ITS4ngsUni – ont été recommandées pour l'amplifier correctement (Tedersoo et Lindahl, 2016). L'idéal pour minimiser les biais moléculaires serait de se passer de l'étape d'amplification comme dans une approche récente qui capture les ARN ribosomiques via sélection de longueur permettant le séquençage de séquences ribosomiques de bactéries, archées et eucaryotes (Karst et al., 2018). Dans l'attente de la généralisation de ces méthodes, le séquençage Illumina MiSeq avec une région résolutive à l'espèce – comme l'ITS1 (Vu et al., 2019) – génère des données en quantité sur lesquelles les erreurs sont connues et les pipelines sont adaptés permettant d'obtenir un relevé des micro-organismes présents utilisable pour l'inférence.

L'approche de métabarcoding n'est pas tout à fait quantitative et fournit des données compositionnelles (Gloor et al., 2017 ; Lamb et al., 2019). L'obtention d'abondances absolues des micro-organismes a été suggérée pour améliorer l'inférence de réseaux (Berry et Widder, 2014 ; Röttjers et Faust, 2018 ; Widder et al., 2016). Des approches de communautés synthétiques avec une dizaine de souches identifient effectivement des interactions biotiques à partir de ces données absolues et de modèles de Lotka-Volterra (Friedman et al., 2017 ; Mounier et al., 2008 ; Venturelli et al., 2018). Les abondances totales des ADN ou des cellules sont obtenues avec qPCR, ddPCR ou cytométrie en flux mais le couplage de ces données avec les données de communautés n'est pas trivial (voir chapitre 2, p. 26). Nous avons besoin d'une méthode qui puisse tenir compte adéquatement des covariables et traiter les abondances absolues obtenues pour chaque taxon. Xiao et al. (2017) se basent sur les abondances absolues des micro-organismes obtenues à l'état stationnaire du système étudié pour inférer des interactions. Il est cependant difficile de pouvoir dire si l'échantillonnage d'une communauté dans l'environnement est à l'état stationnaire et leur méthode ne propose pas de moyen d'inclure des covariables. La méthode proposée par Biswas et al. (2016) ne se base pas sur les données compositionnelles mais sur des comptages issus d'un processus de Poisson. Cette méthode permet d'inclure les effets de différentes covariables et est implémenté dans une librairie R MInt. Yoon et al. (2019) ont proposé une nouvelle méthode appelée SPRING pour inférer les réseaux à partir de données quantitatives et de compositions, mais elle n'inclut pas les covariables pour le moment. L'intégration de données compositionnelles et de données d'abondance absolues n'est donc pas directe et semble nécessiter un axe de recherche statistique spécifique afin d'éviter de propager des erreurs liées aux différentes méthodes.

2.3 Comment réduire les biais statistiques ?

En absence de réplicats de réseaux, une approche intéressante pour évaluer la réplicabilité de la méthode d'inférence de réseaux utilisée sur ses propres données est suggérée par les développeurs de MPLasso (Lo et Marculescu, 2017). Elle consiste à reconstruire un réseau avec la totalité du jeu de données, puis de générer une vingtaine de réseaux à partir d'un sous-échantillonnage aléatoire du jeu de données et quantifier les liens communs entre le réseau total et ceux issus des sous-échantillons (Lo et Marculescu, 2017). Cette approche non automatisée se rapproche de l'approche de sélection de la pénalité λ StARS (Liu et al., 2010) utilisée par SpiecEasi et PLN (voir p. 24). En plus d'appliquer cette procédure robuste de sélection, PLN permet d'extraire pour chaque lien du réseau inféré la proportion de sous-échantillons dans lesquels ils ont été identifiés (Chiquet et al., 2019). Cela donne une estimation de la fiabilité des liens et permet de trier les hypothèses à tester (Röttjers et Faust, 2018).

Limiter les associations dues aux exigences environnementales nécessite de pouvoir inclure des covariables durant l'inférence de réseaux. Les approches issues de l'écologie statistique telles que HMSC impliquent de modéliser conjointement les abondances de chaque ASV de la communauté (Ovaskainen et al., 2017 ; Warton et al., 2015b). La covariance estimée dans ces modèles une fois les facteurs environnementaux et même phylogénétiques pris en compte, indique les associations entre ASV (Warton et al., 2015a). L'utilisation d'un tel modèle – HMSC – a permis de mettre en évidence des associations entres micro-organismes, soutenus par ailleurs dans la littérature (voir Annexe 2, p. 168). Un arbre de décision guide l'interprétation écologique des résultats de ces modèles (D'Amen et al., 2018) ce qui permettrait de faciliter l'utilisation de ces approches peu appliquées aux données de séquençage dans le cadre de l'étude des pathobiomes. Ces approches sont capables de prédire qualitativement des interactions de compétition ou de mutualisme au sein d'environnements homogènes simulés (Zurell et al., 2018). La capacité de ces modèles à capturer les signaux des interactions biotiques au sein d'échantillons hétérogènes reste encore à être évaluée, notamment dans le cas de données issues du séquençage.

La méthode de reconstruction PLN (Chiquet et al., 2019) paraît être une bonne option pour générer des réseaux à partir de données de séquençage. Elle est capable de prendre en compte la compositionnalité des données, d'inférer des corrélations partielles entre ASV, d'inclure des covariables, d'estimer la répétabilité des associations, d'intégrer différents runs de séquençage microbiens et d'être utilisable rapidement sur R quelque soit le système d'exploitation. De plus elle offre la possibilité de forcer l'inclusion ou l'exclusion de certaines associations lors de l'inférence en modifiant la matrice de pénalités [1]. Des associations peuvent être exclues si elles sont entre des ASV détectés dans peu d'échantillons (voir la méthode de filtre proposé par Cougoul et al., 2019a) ou bien incluses s'il existe un *a priori* sur elles (Lo et Marculescu, 2017). Néanmoins, l'absence d'un ou plusieurs nœuds au moment de l'inférence peut générer des liens fallacieux, que ce soit faute d'un ASV (ou d'OTU) non détecté ou mis de côté pour l'inférence. Une approche récente de modèles graphiques gaussiens basées sur les structures en arbres propose de quantifier le nombre de nœuds manquants au réseau inféré (Robin et al., 2018). Cette méthode, hélas, ne prend pas en compte les comptages, ni les comptages spécifiques issus du séquençage.

2.4 Comment mieux valider les interactions prédites ?

Contrairement aux pipelines bioinformatiques (voir section 1, §2, p. 119), les méthodes d'inférence de réseaux ne disposent pas d'équivalent de communauté artificielle où des interactions entre micro-organismes seraient assemblées artificiellement puis la communauté produite serait séquencée (Röttjers et Faust, 2018). Les réseaux inférés sont souvent comparés à des simulations. Ces simulations peuvent provenir de modèles de Lotka-Volterra généralisés qui décrivent l'évolution des abondances des espèces au cours du temps en fonction de leur taux de croissance et des interactions entre elles (Gonze et al., 2018). Les principales études de comparaison de réseaux citées dans cette thèse se basent sur ces approches (e.g., Berry et Widder, 2014 ; Hirano et Takemoto, 2019 ; Weiss et al., 2016). Les simulations peuvent démarrer d'une matrice de covariance – servant de réseau de référence – pour échantillonner des comptages selon différents modèles statistiques et en introduisant des biais pour imiter les données de séquençage (e.g., Biswas et al., 2016 ; Chiquet et al., 2019 ; Fang et al., 2015 ; Kurtz et al.,

[1] https://github.com/jchiquet/PLNmodels/issues/44

2015). Les méthodes d'inférence sont comparées avec ces deux types de simulation mais il n'est pas clair à quel point ces simulations reflètent les interactions microbiennes (Röttjers et Faust, 2018). L'assemblage de communautés microbiennes synthétiques pour mieux comprendre leurs interactions n'est pas neuf (e.g., Friedman et al., 2017 ; Hoek et al., 2016 ; Li et al., 2019 ; Mounier et al., 2008 ; Venturelli et al., 2018). Mais la plupart de ces études mesurent les abondances des micro-organismes par densité optique ou cytométrie de flux et non avec des techniques de séquençage, ce qui ne permet pas de les utiliser comme un jeu de données expertisé pour évaluer l'inférence de réseaux. Plusieurs auteurs soulignent qu'un tel jeu de données permettrait de comparer de manière reproductible les méthodes actuelles et de guider les nouveaux développements (Hirano et Takemoto, 2019 ; Röttjers et Faust, 2018). L'étude de Venturelli et al. (2018) avec le suivi temporel de 12 bactéries dans près de 600 échantillons pourrait servir de base mais le réseau proposé par les auteurs nécessite d'être validé par des études indépendantes.

Les réseaux prédits sont souvent validés par des expérimentations de cultures des micro-organismes sur des milieux synthétiques (e.g., Das et al., 2018 ; Tipton et al., 2018 ; Wang et al., 2017). Des approches automatisées se développent pour surmonter la longueur de ces tâches et permettre de tester de multiples combinaisons de nutriments et d'espèces (Kehe et al., 2019 ; Lagier et al., 2018). Des stratégies complémentaires visent à étendre les approches réductionnistes de culture en réintégrant la plante, soit dans les approches *in vitro* avec des plantes sur des puces miniatures (Stanley et van der Heijden, 2017), ou encore avec des boîtes de Petri moulées sur la microstructure de la surface des feuilles (Soffe et al., 2019). Ces approches sont réductionnistes et ne peuvent mimer la totalité des conditions abiotiques (UV, température, etc.) et biotiques (les micro-organismes non inclus), mais elles permettent au moins de démêler la complexité des mécanismes qui opèrent dans les écosystèmes (Crowther et al., 2018).

L'approche de cette thèse s'inscrit dans les approches – dites *top-down* – qui caractérisent les communautés naturelles (*top*) et cherchent à inférer et à valider les interactions qui la constitue (*down* ; Vega et Gore, 2018). Une approche complémentaire – dite *bottom-up* (Vega et Gore, 2018) – évalue d'abord les interactions paires à paires (*bottom*) dans des communautés synthétiques d'une dizaine de micro-organismes, puis des interactions avec plus de membres afin de reconstruire un réseau d'interactions complet (*up* ; e.g., Friedman et al., 2017 ; Li et al., 2019 ; Venturelli et al., 2018). Ces deux approches ne sont pas incompatibles. Par exemple, la reconstruction de réseaux *top-down* peut s'allier d'expériences d'inhibition paires à paires, d'assemblage de communautés synthétiques et d'inoculations contrôlés pour déduire des interactions nécessaires à la survie de la plante (Durán et al., 2018). En dehors de plantes modèles et de micro-organismes associées, l'approche *bottom-up* est plus difficile à mettre en œuvre et l'approche utilisée ici (*top-bottom*) pourra être utilisée sous réserve de tests supplémentaires.

En définitive, identifier les interactions dans les pathobiomes des plantes est important particulièrement lorsque ces interactions influencent l'état de santé de l'hôte. Mais comprendre des systèmes aussi complexes que ces communautés microbiennes nécessite plusieurs échelles et méthodes complémentaires (Cordero et Datta, 2016 ; Widder et al., 2016). Parmi elles, les approches de métabarcoding fournissent un type de relevé des micro-organismes présents avec ses limites certes, mais ces approches sont suffisamment matures pour que des recommandations

soient formulées (e.g., Zinger et al., 2019). Reconstruire les réseaux d'associations constitue une approche pour identifier de la structure dans les comptages de ces relevés microbiens. Mais les réseaux d'associations doivent être considérés pour ce qu'ils sont : des générateurs d'hypothèses (Carr et al., 2019 ; Röttjers et Faust, 2018). Dans l'attente d'une communauté artificielle de réseau microbien standardisée, il est nécessaire d'être vigilant lors de la reconstruction – e.g. en ayant suffisamment d'échantillons homogènes. Ceci permettra de générer les hypothèses les plus plausibles possibles dès lors qu'une méthode apte à surmonter les barrières est utilisée (e.g. PLN). Ces approches ne sont qu'une pierre à l'édifice scientifique interdisciplinaire qui vise à mieux comprendre les mécanismes en cours au sein des holobiontes et notamment des pathobiomes.

VI

Bibliographie

Bibliographie

Aas AB, Davey ML et Kauserud H (2017). "ITS All Right Mama : Investigating the Formation of Chimeric Sequences in the ITS2 Region by DNA Metabarcoding Analyses of Fungal Mock Communities of Different Complexities". In : *Molecular Ecology Resources* 17.4, p. 730–741. doi : 10.1111/1755-0998.12622.

Abarenkov K, Henrik Nilsson R, Larsson KH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Ursing BM, Vrålstad T, Liimatainen K, Peintner U et Kõljalg U (2010). "The UNITE Database for Molecular Identification of Fungi - Recent Updates and Future Perspectives : Letters". In : *New Phytologist* 186.2, p. 281–285. doi : 10.1111/j.1469-8137.2009.03160.x.

Abarenkov K, Somervuo P, Nilsson RH, Kirk PM, Huotari T, Abrego N et Ovaskainen O (2018). "Protax-Fungi : A Web-Based Tool for Probabilistic Taxonomic Placement of Fungal Internal Transcribed Spacer Sequences". In : *New Phytologist* 220.2, p. 517–525. doi : 10.1111/nph.15301.

Abdelfattah A, Malacrinò A, Wisniewski M, Cacciola SO et Schena L (2018). "Metabarcoding : A Powerful Tool to Investigate Microbial Communities and Shape Future Plant Protection Strategies". In : *Biological Control.* Metagenomics and the Science of Biological Control 120, p. 1–10. doi : 10.1016/j.biocontrol.2017.07.009.

Addis E, Fleet G, Cox J, Kolak D et Leung T (2001). "The Growth, Properties and Interactions of Yeasts and Bacteria Associated with the Maturation of Camembert and Blue-Veined Cheeses". In : *International Journal of Food Microbiology* 69.1-2, p. 25–36. doi : 10.1016/S0168-1605(01)00569-4.

Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D et Kemen EM (2016). "Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation". In : *PLOS Biology* 14.1. Sous la dir. de MK Waldor, e1002352. doi : 10.1371/journal.pbio.1002352.

Agtmaal M van, Straathof A, Termorshuizen A, Teurlincx S, Hundscheid M, Ruyters S, Busschaert P, Lievens B et Boer W de (2017). "Exploring the Reservoir of Potential Fungal Plant Pathogens in Agricultural Soil". In : *Applied Soil Ecology* 121, p. 152–160. doi : 10.1016/j.apsoil.2017.09.032.

Aitchison J (1982). "The Statistical Analysis of Compositional Data". In : *Journal of the Royal Statistical Society : Series B (Methodological)* 44.2, p. 139–160. doi : 10.1111/j.2517-6161.1982.tb01195.x.

Aitchison J et Ho CH (1989). "The Multivariate Poisson-Log Normal Distribution". In : *Biometrika* 76.4, p. 643–653. doi : 10.1093/biomet/76.4.643.

Alberdi A, Aizpurua O, Gilbert MTP et Bohmann K (2018). "Scrutinizing Key Steps for Reliable Metabarcoding of Environmental Samples". In : *Methods in Ecology and Evolution* 9.1, p. 134–147. doi : 10.1111/2041-210X.12849.

Almeida RPP, De La Fuente L, Koebnik R, Lopes JRS, Parnell S et Scherm H (2019). "Addressing the New Global Threat of Xylella Fastidiosa". In : *Phytopathology* 109.2, p. 172–174. doi : 10.1094/PHYTO-12-18-0488-FI.

Altschul SF, Gish W, Miller W, Myers EW et Lipman DJ (1990). "Basic Local Alignment Search Tool". In : *Journal of Molecular Biology* 215.3, p. 403–410. DOI : 10.1016/S0022-2836(05)80360-2. pmid : 2231712.

Amend AS, Cobian GM, Laruson AJ, Remple K, Tucker SJ, Poff KE, Antaky C, Boraks A, Jones CA, Kuehu D, Lensing BR, Pejhanmehr M, Richardson DT et Riley PP (2019). "Phytobiomes Are Compositionally Nested from the Ground Up". In : *PeerJ* 7, e6609. DOI : 10.7717/peerj.6609.

Amend AS, Seifert KA et Bruns TD (2010). "Quantifying Microbial Communities with 454 Pyrosequencing : Does Read Abundance Count ?" In : *Molecular Ecology* 19.24, p. 5555–5565. DOI : 10.1111/j.1365-294X.2010.04898.x.

Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, Kightley EP, Thompson LR, Hyde ER, Gonzalez A et Knight R (2017). "Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns". In : *mSystems* 2.2, e00191–16. DOI : 10.1128/mSystems.00191-16. pmid : 28289731.

Anderson MJ (2001). "A New Method for Non-parametric Multivariate Analysis of Variance". In : *Austral Ecology* 26.1, p. 32–46. DOI : 10.1111/j.1442-9993.2001.01070.pp.x.

— (2006). "Distance-Based Tests for Homogeneity of Multivariate Dispersions". In : *Biometrics* 62.1, p. 245–253. DOI : 10.1111/j.1541-0420.2005.00440.x.

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR et Daszak P (2004). "Emerging Infectious Diseases of Plants : Pathogen Pollution, Climate Change and Agrotechnology Drivers". In : *Trends in Ecology & Evolution* 19.10, p. 535–544. DOI : 10.1016/j.tree.2004.07.021.

Andrews JH, Kinkel LL, Berbee FM et Nordheim EV (1987). "Fungi, Leaves, and the Theory of Island Biogeography". In : *Microbial Ecology* 14.3, p. 277–290. DOI : 10.1007/BF02012947.

Angeli D, Pellegrini E et Pertot I (2009). "Occurrence of Erysiphe Necator Chasmothecia and Their Natural Parasitism by Ampelomyces Quisqualis". In : *Phytopathology* 99.6, p. 704–710. DOI : 10.1094/PHYTO-99-6-0704. pmid : 19453229.

Anonyme (2019a). *Liste Des Produits Phytopharmaceutiques de Biocontrôle, Au Titre Des Articles L.253-5 et L.253-7 Du Code Rural et de La Pêche Maritime*. DGAL/SDQSPV/2019-525. Ministère de l'Agriculture et de l'Alimentation. DOI : https://agriculture.gouv.fr/telecharger/100505?token=e6cf91ddb4eb4bff9e3ec73a2a8e0ee0.

— (2019b). *Statistique Agricole Annuelle 2017-2018*. Ministère de l'Agriculture et de l'Alimentation. DOI : http://agreste.agriculture.gouv.fr/spip.php?article15405.

Anslan S, Nilsson RH, Wurzbacher C, Baldrian P, Tedersoo L et Bahram M (2018). "Great Differences in Performance and Outcome of High-Throughput Sequencing Data Analysis Platforms for Fungal Metabarcoding". In : *MycoKeys* 39, p. 29–40. DOI : 10.3897/mycokeys.39.28109.

Armijo G, Schlechter R, Agurto M, Muñoz D, Nuñez C et Arce-Johnson P (2016). "Grapevine Pathogenic Microorganisms : Understanding Infection Strategies and Host Response Scenarios". In : *Frontiers in Plant Science* 7, p. 382. DOI : 10.3389/fpls.2016.00382.

Armitage DW et Jones SE (2019). "How Sample Heterogeneity Can Obscure the Signal of Microbial Interactions". In : *The ISME Journal*, p. 1. DOI : 10.1038/s41396-019-0463-3.

Arnold AE, Mejía LC, Kyllo D, Rojas EI, Maynard Z, Robbins N et Herre EA (2003). "Fungal Endophytes Limit Pathogen Damage in a Tropical Tree". In : *Proceedings of the National Academy of Sciences* 100.26, p. 15649–15654. DOI : 10.1073/pnas.2533483100. pmid : 14671327.

Ayala FJ (2011). "Elixir of Life : *In Vino Veritas*". In : *Proceedings of the National Academy of Sciences* 108.9, p. 3457–3458. DOI : 10.1073/pnas.1019729108.

Baird DJ et Hajibabaei M (2012). “Biomonitoring 2.0 : A New Paradigm in Ecosystem Assessment Made Possible by next-Generation DNA Sequencing”. In : *Molecular Ecology* 21.8, p. 2039–2044. doi : 10.1111/j.1365-294X.2012.05519.x.

Bakker MG (2018). “A Fungal Mock Community Control for Amplicon Sequencing Experiments”. In : *Molecular Ecology Resources* 18.3, p. 541–556. doi : 10.1111/1755-0998.12760.

Baldrian P (2017). “Forest Microbiome : Diversity, Complexity and Dynamics”. In : *FEMS Microbiology Reviews* 41.2, p. 109–130. doi : 10.1093/femsre/fuw040.

Bálint M, Bahram M, Eren AM, Faust K, Fuhrman JA, Lindahl B, O’Hara RB, Öpik M, Sogin ML, Unterseher M et Tedersoo L (2016). “Millions of Reads, Thousands of Taxa : Microbial Community Structure and Associations Analyzed via Marker Genes”. In : *FEMS Microbiology Reviews* 40.5, p. 686–700. doi : 10.1093/femsre/fuw017.

Bálint M, Bartha L, O’Hara RB, Olson MS, Otte J, Pfenninger M, Robertson AL, Tiffin P et Schmitt I (2015). “Relocation, High-latitude Warming and Host Genetic Identity Shape the Foliar Fungal Microbiome of Poplars”. In : *Molecular Ecology* 24.1, p. 235–248. doi : 10.1111/mec.13018.

Bálint M, Schmidt PA, Sharma R, Thines M et Schmitt I (2014). “An Illumina Metabarcoding Pipeline for Fungi”. In : *Ecology and Evolution* 4.13, p. 2642–2653. doi : 10.1002/ece3.1107.

Ban Y, An L et Jiang H (2015). “Investigating Microbial Co-Occurrence Patterns Based on Metagenomic Compositional Data”. In : *Bioinformatics* 31.20, p. 3322–3329. doi : 10.1093/bioinformatics/btv364.

Banerjee S, Schlaeppi K et Heijden MGA (2018). “Keystone Taxa as Drivers of Microbiome Structure and Functioning”. In : *Nature Reviews Microbiology*, p. 1. doi : 10.1038/s41579-018-0024-1.

Barabási AL, Albert R et Jeong H (2000). “Scale-Free Characteristics of Random Networks : The Topology of the World-Wide Web”. In : *Physica A : Statistical Mechanics and its Applications* 281.1, p. 69–77. doi : 10.1016/S0378-4371(00)00018-2.

Barberán A, Bates ST, Casamayor EO et Fierer N (2012). “Using Network Analysis to Explore Co-Occurrence Patterns in Soil Microbial Communities”. In : *The ISME Journal* 6.2, p. 343–351. doi : 10.1038/ismej.2011.119.

Barner AK, Coblentz KE, Hacker SD et Menge BA (2018). “Fundamental Contradictions among Observational and Experimental Estimates of Non-trophic Species Interactions”. In : *Ecology* 99.3, p. 557–566. doi : 10.1002/ecy.2133.

Barret M, Briand M, Bonneau S, Préveaux A, Valière S, Bouchez O, Hunault G, Simoneau P et Jacques MA (2015). “Emergence Shapes the Structure of the Seed Microbiota”. In : *Applied and Environmental Microbiology* 81.4. Sous la dir. de HL Drake, p. 1257–1266. doi : 10.1128/AEM.03722-14.

Bass D, Stentiford GD, Wang HC, Koskella B et Tyler CR (2019). “The Pathobiome in Animal and Plant Diseases”. In : *Trends in Ecology & Evolution*, S0169534719302241. doi : 10.1016/j.tree.2019.07.012. Online ahead of print.

Bates D, Mächler M, Bolker B et Walker S (2015). “Fitting Linear Mixed-Effects Models Using Lme4”. In : *Journal of Statistical Software* 67.1, p. 1–48. doi : 10.18637/jss.v067.i01.

Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P et Kauserud H (2010). “ITS as an Environmental DNA Barcode for Fungi : An in Silico Approach Reveals Potential PCR Biases”. In : *BMC Microbiology* 10, p. 189. doi : 10.1186/1471-2180-10-189.

Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DGJ et Nilsson RH (2015). “Metaxa2 : Improved Identification and Taxonomic Classification of Small and Large Subunit rRNA in Metagenomic Data”. In : *Molecular Ecology Resources* 15.6, p. 1403–1414. doi : 10.1111/1755-0998.12399.

Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, de Sousa F, Amend AS, Jumpponen A, Unterseher M, Kristiansson E, Abarenkov K, Bertrand YJK, Sanli K, Eriksson KM, Vik U, Veldre V et Nilsson RH (2013). "Improved Software Detection and Extraction of ITS1 and ITS2 from Ribosomal ITS Sequences of Fungi and Other Eukaryotes for Analysis of Environmental Sequencing Data". In : *Methods in Ecology and Evolution* 4.10, p. 914–919. doi : 10.1111/2041-210X.12073.

Berendsen RL, Pieterse CMJ et Bakker PAHM (2012). "The Rhizosphere Microbiome and Plant Health". In : *Trends in Plant Science* 17.8, p. 478–486. doi : 10.1016/j.tplants.2012.04.001. pmid : 22564542.

Berry D et Widder S (2014). "Deciphering Microbial Interactions and Detecting Keystone Species with Co-Occurrence Networks". In : *Frontiers in Microbiology* 5, p. 219. doi : 10.3389/fmicb.2014.00219.

Berthold T, Centler F, Hübschmann T, Remer R, Thullner M, Harms H et Wick LY (2016). "Mycelia as a Focal Point for Horizontal Gene Transfer among Soil Bacteria". In : *Scientific Reports* 6.1, p. 36390. doi : 10.1038/srep36390.

Biswas S, Mcdonald M, Lundberg DS, Dangl JL et Jojic V (2016). "Learning microbial interaction networks from metagenomic count data". In : *Journal of Computational Biology* 23.6, p. 526–535. doi : 10.1089/cmb.2016.0061.

Blancard D (2019). *Vigne - Index Des Maladies, Ravageurs, Vecteurs et Auxiliaires*. url : http://ephytia.inra.fr/fr/C/6045/Vigne-Index-des-maladies-ravageurs-vecteurs-et-auxiliaires (visité le 23/08/2019).

Bohan DA, Caron-Lormier G, Muggleton S, Raybould A et Tamaddoni-Nezhad A (2011). "Automated Discovery of Food Webs from Ecological Data Using Logic-Based Machine Learning". In : *PLoS ONE* 6.12. Sous la dir. de RKF Unsworth, e29028. doi : 10.1371/journal.pone.0029028.

Bohan DA, Vacher C, Tamaddoni-Nezhad A, Raybould A, Dumbrell AJ et Woodward G (2017). "Next-Generation Global Biomonitoring : Large-Scale, Automated Reconstruction of Ecological Networks". In : *Trends in Ecology & Evolution* 32.7, p. 477–487. doi : 10.1016/j.tree.2017.03.001.

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA et Gregory Caporaso J (2018). "Optimizing Taxonomic Classification of Marker-Gene Amplicon Sequences with QIIME 2's Q2-Feature-Classifier Plugin". In : *Microbiome* 6.1, p. 90. doi : 10.1186/s40168-018-0470-z.

Bokulich NA, Rideout JR, Mercurio WG, Shiffer A, Wolfe B, Maurice CF, Dutton RJ, Turnbaugh PJ, Knight R et Caporaso JG (2016). "Mockrobiota : A Public Resource for Microbiome Bioinformatics Benchmarking". In : *mSystems* 1.5. Sous la dir. de JD Neufeld, e00062–16, /msys/1/5/e00062–16.atom. doi : 10.1128/mSystems.00062-16.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA et Caporaso JG (2013). "Quality-Filtering Vastly Improves Diversity Estimates from Illumina Amplicon Sequencing". In : *Nature methods* 10.1, p. 57–59. doi : 10.1038/nmeth.2276. pmid : 23202435.

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH et White JSS (2009). "Generalized Linear Mixed Models : A Practical Guide for Ecology and Evolution". In : *Trends in Ecology & Evolution* 24.3, p. 127–135. doi : 10.1016/j.tree.2008.10.008. pmid : 19185386.

Bolyen E et al. (2019). "Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2". In : *Nature Biotechnology* 37.8, p. 852–857. doi : 10.1038/s41587-019-0209-9.

Bonfante P et Genre A (2010). “Mechanisms Underlying Beneficial Plant–Fungus Interactions in Mycorrhizal Symbiosis”. In : *Nature Communications* 1, p. 48. doi : 10.1038/ncomms1046.

Bordenstein SR et Theis KR (2015). “Host Biology in Light of the Microbiome : Ten Principles of Holobionts and Hologenomes”. In : *PLOS Biology* 13.8. Sous la dir. de MK Waldor, e1002226. doi : 10.1371/journal.pbio.1002226.

Botnen SS, Davey ML, Halvorsen R et Kauserud H (2018). “Sequence Clustering Threshold Has Little Effect on the Recovery of Microbial Community Structure”. In : *Molecular Ecology Resources* 18.5, p. 1064–1076. doi : 10.1111/1755-0998.12894.

Bozoudi D et Tsaltas D (2018). “The Multiple and Versatile Roles of Aureobasidium Pullulans in the Vitivinicultural Sector”. In : *Fermentation* 4.4, p. 85. doi : 10.3390/fermentation4040085.

Brader G, Compant S, Vescio K, Mitter B, Trognitz F, Ma LJ et Sessitsch A (2017). “Ecology and Genomic Insights on Plant-Pathogenic and -Nonpathogenic Endophytes”. In : *Annual Review of Phytopathology* 55.1, p. 61–83. doi : 10.1146/annurev-phyto-080516-035641.

Bradley DJ, Gilbert GS et Parker IM (2003). “Susceptibility of Clover Species to Fungal Infection : The Interaction of Leaf Surface Traits and Environment”. In : *American Journal of Botany* 90.6, p. 857–864. doi : 10.3732/ajb.90.6.857.

Brown SP, Veach AM, Rigdon-Huss AR, Grond K, Lickteig SK, Lothamer K, Oliver AK et Jumpponen A (2015). “Scraping the Bottom of the Barrel : Are Rare High Throughput Sequences Artifacts ?” In : *Fungal Ecology* 13, p. 221–225. doi : 10.1016/j.funeco.2014.08.006.

Bruez E, Vallance J, Gerbore J, Lecomte P, Costa JPD, Guerin-Dubrana L et Rey P (2014). “Analyses of the Temporal Dynamics of Fungal Communities Colonizing the Healthy Wood Tissues of Esca Leaf-Symptomatic and Asymptomatic Vines”. In : *PLOS ONE* 9.5, e95928. doi : 10.1371/journal.pone.0095928.

Buée M, Boer WD, Martin F, Overbeek L van et Jurkevitch E (2009a). “The Rhizosphere Zoo : An Overview of Plant-Associated Communities of Microorganisms, Including Phages, Bacteria, Archaea, and Fungi, and of Some of Their Structuring Factors”. In : *Plant and Soil* 321.1-2, p. 189–212. doi : 10.1007/s11104-009-9991-3.

Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S et Martin F (2009b). “454 Pyrosequencing Analyses of Forest Soils Reveal an Unexpectedly High Fungal Diversity”. In : *New Phytologist* 184.2, p. 449–456. doi : 10.1111/j.1469-8137.2009.03003.x.

Bulgarelli D, Schlaeppi K, Spaepen S, van Themaat EVL et Schulze-Lefert P (2013). “Structure and Functions of the Bacterial Microbiota of Plants”. In : *Annual Review of Plant Biology* 64.1, p. 807–838. doi : 10.1146/annurev-arplant-050312-120106.

Burgaud G, Arzur D, Sampaio JP et Barbier G (2011). “Candida Oceani Sp. Nov., a Novel Yeast Isolated from a Mid-Atlantic Ridge Hydrothermal Vent (-2300 Meters)”. In : *Antonie van Leeuwenhoek* 100.1, p. 75–82. doi : 10.1007/s10482-011-9566-1.

Burgaud G, Coton M, Jacques N, Debaets S, Maciel NOP, Rosa CA, Gadanho M, Sampaio JP et Casaregola S (2016). “Yamadazyma Barbieri f.a. Sp. Nov., an Ascomycetous Anamorphic Yeast Isolated from a Mid-Atlantic Ridge Hydrothermal Site (-2300 m) and Marine Coastal Waters”. In : *International Journal of Systematic and Evolutionary Microbiology* 66.9, p. 3600–3606. doi : 10.1099/ijsem.0.001239.

Burns KN, Bokulich NA, Cantu D, Greenhut RF, Kluepfel DA, O’Geen AT, Strauss SL et Steenwerth KL (2016). “Vineyard Soil Bacterial Diversity and Composition Revealed by 16S rRNA Genes : Differentiation by Vineyard Management”. In : *Soil Biology and Biochemistry* 103, p. 337–348. doi : 10.1016/j.soilbio.2016.09.007.

Buzzini P, Lachance MA et Yurkov A, éds. (2017). *Yeasts in Natural Ecosystems : Diversity*. Cham : Springer International Publishing. isbn : 978-3-319-62682-6 978-3-319-62683-3. doi : 10.1007/978-3-319-62683-3.

Callahan BJ, McMurdie PJ et Holmes SP (2017). “Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis”. In : *The ISME Journal* 11.12, p. 2639–2643. doi : 10.1038/ismej.2017.119.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA et Holmes SP (2016). “DADA2 : High-Resolution Sample Inference from Illumina Amplicon Data”. In : *Nature Methods* 13.7, p. 581–583. doi : 10.1038/nmeth.3869.

Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK et Dougherty MK (2019). “High-Throughput Amplicon Sequencing of the Full-Length 16S rRNA Gene with Single-Nucleotide Resolution”. In : *Nucleic Acids Research*. doi : 10.1093/nar/gkz569.

Calonnec A, Cartolaro P et Chadœuf J (2009). “Highlighting Features of Spatiotemporal Spread of Powdery Mildew Epidemics in the Vineyard Using Statistical Modeling on Field Experimental Data”. In : *Phytopathology* 99.4, p. 411–422. doi : 10.1094/PHYTO-99-4-0411.

Calonnec A, Cartolaro P, Delière L et Chadœuf J (2006). “Powdery Mildew on Grapevine : The Date of Primary Contamination Affects Disease Development on Leaves and Damage on Grape”. In : *Bulletin OILB/SROP* 29.11, p. 67–73.

Calonnec A, Cartolaro P, Poupot C, Dubourdieu D et Darriet P (2004). “Effects of Uncinula Necator on the Yield and Quality of Grapes (Vitis Vinifera) and Wine”. In : *Plant Pathology* 53.4, p. 434–445. doi : 10.1111/j.0032-0862.2004.01016.x.

Calonnec A, Jolivet J, Vivin P et Schnee S (2018). “Pathogenicity Traits Correlate With the Susceptible Vitis Vinifera Leaf Physiology Transition in the Biotroph Fungus Erysiphe Necator : An Adaptation to Plant Ontogenic Resistance”. In : *Frontiers in Plant Science* 9, p. 1808. doi : 10.3389/fpls.2018.01808.

Campisano A, Antonielli L, Pancher M, Yousaf S, Pindo M et Pertot I (2014). “Bacterial Endophytic Communities in the Grapevine Depend on Pest Management”. In : *PLOS ONE* 9.11, e112763. doi : 10.1371/journal.pone.0112763.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J et Knight R (2010). “QIIME Allows Analysis of High-Throughput Community Sequencing Data”. In : *Nature Methods* 7.5, p. 335–336. doi : 10.1038/nmeth.f.303.

Carmichael PC, Siyoum N, Chidamba L et Korsten L (2019). “Exploring the Microbial Communities Associated with Botrytis Cinerea during Berry Development in Table Grape with Emphasis on Potential Biocontrol Yeasts”. In : *European Journal of Plant Pathology*. doi : 10.1007/s10658-019-01710-5.

Carr A, Diener C, Baliga NS et Gibbons SM (2019). “Use and Abuse of Correlation Analyses in Microbial Ecology”. In : *The ISME Journal*, p. 1. doi : 10.1038/s41396-019-0459-z.

Carroll JE et Wilcox WF (2003). “Effects of Humidity on the Development of Grapevine Powdery Mildew”. In : *Phytopathology* 93.9, p. 1137–1144. doi : 10.1094/PHYTO.2003.93.9.1137.

Caruso V, Song X, Asquith M et Karstens L (2019). “Performance of Microbiome Sequence Inference Methods in Environments with Varying Biomass”. In : *mSystems* 4.1, e00163–18. doi : 10.1128/mSystems.00163-18.

Castañeda LE, Miura T, Sánchez R et Barbosa O (2018). “Effects of Agricultural Management on Phyllosphere Fungal Diversity in Vineyards and the Association with Adjacent Native Forests”. In : *PeerJ* 6, e5715. doi : 10.7717/peerj.5715.

Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, Crowther TW, Danovaro R, Foreman CM, Huisman J,

Hutchins DA, Jansson JK, Karl DM, Koskella B, Mark Welch DB, Martiny JBH, Moran MA, Orphan VJ, Reay DS, Remais JV, Rich VI, Singh BK, Stein LY, Stewart FJ, Sullivan MB, van Oppen MJH, Weaver SC, Webb EA et Webster NS (2019). "Scientists' Warning to Humanity : Microorganisms and Climate Change". In : *Nature Reviews Microbiology*. DOI : 10.1038/s41579-019-0222-5.

Chaffron S, Rehrauer H, Pernthaler J et Mering C von (2010). "A Global Network of Coexisting Microbes from Environmental and Whole-Genome Sequence Data". In : *Genome Research* 20.7, p. 947–959. DOI : 10.1101/gr.104521.109. pmid : 20458099.

Chao A, Chazdon RL, Colwell RK et Shen TJ (2006). "Abundance-Based Similarity Indices and Their Estimation When There Are Unseen Species in Samples". In : *Biometrics* 62.2, p. 361–371. DOI : 10.1111/j.1541-0420.2005.00489.x.

Chen C, Xin K, Liu H, Cheng J, Shen X, Wang Y et Zhang L (2017). "*Pantoea Alhagi*, a Novel Endophytic Bacterium with Ability to Improve Growth and Drought Tolerance in Wheat". In : *Scientific Reports* 7, srep41564. DOI : 10.1038/srep41564.

Chen H (2018). *VennDiagram : Generate High-Resolution Venn and Euler Plots*.

Chen J et Chen Z (2008). "Extended Bayesian Information Criteria for Model Selection with Large Model Spaces". In : *Biometrika* 95.3, p. 759–771. DOI : 10.1093/biomet/asn034.

Chen W, Turkington TK, Lévesque CA, Bamforth JM, Patrick SK, Lewis CT, Chapados JT, Gaba D, Tittlemier SA, MacLeod A et Gräfenhan T (2016). "Geography and Agronomical Practices Drive Diversification of the Epiphytic Mycoflora Associated with Barley and Its Malt End Product in Western Canada". In : *Agriculture, Ecosystems & Environment* 226, p. 43–55. DOI : 10.1016/j.agee.2016.03.030.

Chiquet J, Mariadassou M et Robin S (2017). "Variational Inference for Probabilistic Poisson PCA". In : arXiv : 1703.06633 [stat].

— (2018). "Variational Inference for Sparse Network Reconstruction from Count Data". In : arXiv : 1806.03120 [stat].

Chiquet J, Robin S et Mariadassou M (2019). "Variational Inference for Sparse Network Reconstruction from Count Data". In : *Proceedings of the 36th International Conference on Machine Learning*. Sous la dir. de K Chaudhuri et R Salakhutdinov. T. 97. Proceedings of Machine Learning Research. Long Beach, California, USA : PMLR, p. 1162–1171.

Chow CET, Kim DY, Sachdeva R, Caron DA et Fuhrman JA (2014). "Top-down Controls on Bacterial Community Structure : Microbial Network Analysis of Bacteria, T4-like Viruses and Protists". In : *The ISME Journal* 8.4, p. 816–829. DOI : 10.1038/ismej.2013.199. pmid : 24196323.

Cline LC, Song Z, Al-Ghalith GA, Knights D et Kennedy PG (2017). "Moving beyond de Novo Clustering in Fungal Community Ecology". In : *New Phytologist*. DOI : 10.1111/nph.14752.

Coleman TF et Moré JJ (1983). "Estimation of Sparse Jacobian Matrices and Graph Coloring Blems". In : *SIAM Journal on Numerical Analysis* 20.1, p. 187–209. DOI : 10.1137/0720013.

Community U (2019). *UNITE General FASTA Release for Fungi 2*. DOI : 10.15156/bio/786353.

Compant S et Mathieu F, éds. (2016). *Biocontrol of Major Grapevine Diseases : Leading Research*. Boston, MA : CABI. ISBN : 978-1-78064-712-8.

Compant S, Mitter B, Colli-Mull JG, Gangl H et Sessitsch A (2011). "Endophytes of Grapevine Flowers, Berries, and Seeds : Identification of Cultivable Bacteria, Comparison with Other Plant Parts, and Visualization of Niches of Colonization". In : *Microbial Ecology* 62.1, p. 188–197. DOI : 10.1007/s00248-011-9883-y.

Comtet T, Sandionigi A, Viard F et Casiraghi M (2015). "DNA (Meta)Barcoding of Biological Invasions : A Powerful Tool to Elucidate Invasion Processes and Help Managing Aliens". In : *Biological Invasions* 17.3, p. 905–922. DOI : 10.1007/s10530-015-0854-y.

Connor N, Barberán A et Clauset A (2017). "Using Null Models to Infer Microbial Co-Occurrence Networks". In : *PLOS ONE* 12.5, e0176751. doi : 10.1371/journal.pone.0176751.

Copeland JK, Yuan L, Layeghifard M, Wang PW et Guttman DS (2015). "Seasonal Community Succession of the Phyllosphere Microbiome". In : *Molecular Plant-Microbe Interactions* 28.3, p. 274–285. doi : 10.1094/MPMI-10-14-0331-FI.

Cordero OX et Datta MS (2016). "Microbial Interactions and Community Assembly at Microscales". In : *Current Opinion in Microbiology* 31, p. 227–234. doi : 10.1016/j.mib.2016.03.015.

Cordier T, Forster D, Dufresne Y, Martins CIM, Stoeck T et Pawlowski J (2018a). "Supervised Machine Learning Outperforms Taxonomy-Based Environmental DNA Metabarcoding Applied to Biomonitoring". In : *Molecular Ecology Resources* 18.6, p. 1381–1391. doi : 10.1111/1755-0998.12926.

Cordier T, Lanzén A, Apothéloz-Perret-Gentil L, Stoeck T et Pawlowski J (2018b). "Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring". In : *Trends in Microbiology*. doi : 10.1016/j.tim.2018.10.012.

Cordier T, Robin C, Capdevielle X, Desprez-Loustau ML et Vacher C (2012a). "Spatial Variability of Phyllosphere Fungal Assemblages : Genetic Distance Predominates over Geographic Distance in a European Beech Stand (Fagus Sylvatica)". In : *Fungal Ecology* 5.5, p. 509–520. doi : 10.1016/j.funeco.2011.12.004.

Cordier T, Robin C, Capdevielle X, Fabreguettes O, Desprez-Loustau ML et Vacher C (2012b). "The Composition of Phyllosphere Fungal Assemblages of European Beech (*Fagus Sylvatica*) Varies Significantly along an Elevation Gradient". In : *New Phytologist* 196.2, p. 510–519. doi : 10.1111/j.1469-8137.2012.04284.x.

Corio-Costet MF (2007). *Erysiphe necator*. Monographies de microbiologie. Paris : Tec et Doc. 132 p. isbn : 978-2-7430-1034-8.

Corpet F (1988). "Multiple Sequence Alignment with Hierarchical Clustering". In : *Nucleic Acids Research* 16.22, p. 10881–10890. doi : 10.1093/nar/16.22.10881.

Cortesi P, Pizzatti C, Bertocchi D et Milgroom MG (2008). "Persistence and Spatial Autocorrelation of Clones of *Erysiphe Necator* Overwintering as Mycelium in Dormant Buds in an Isolated Vineyard in Northern Italy". In : *Phytopathology* 98.2, p. 148–152. doi : 10.1094/PHYTO-98-2-0148.

Cougoul A, Bailly X, Vourc'h G et Gasqui P (2019a). "Rarity of Microbial Species : In Search of Reliable Associations". In : *PLOS ONE* 14.3. Sous la dir. de H Smidt, e0200458. doi : 10.1371/journal.pone.0200458.

Cougoul A, Bailly X et Wit EC (2019b). "MAGMA : Inference of Sparse Microbial Association Networks". In : *bioRxiv*, p. 538579. doi : 10.1101/538579.

Coyte KZ, Schluter J et Foster KR (2015). "The Ecology of the Microbiome : Networks, Competition, and Stability". In : *Science* 350.6261, p. 663–666. doi : 10.1126/science.aad2602. pmid : 26542567.

Crandall SG et Gilbert GS (2017). "Meteorological Factors Associated with Abundance of Airborne Fungal Spores over Natural Vegetation". In : *Atmospheric Environment* 162, p. 87–99. doi : 10.1016/j.atmosenv.2017.05.018.

Creamer R, Hannula S, Leeuwen J, Stone D, Rutgers M, Schmelz R, Ruiter P, Hendriksen N, Bolger T, Bouffaud M, Buee M, Carvalho F, Costa D, Dirilgen T, Francisco R, Griffiths B, Griffiths R, Martin F, Silva P da, Mendes S, Morais P, Pereira C, Philippot L, Plassart P, Redecker D, Römbke J, Sousa J, Wouterse M et Lemanceau P (2016). "Ecological Network Analysis Reveals the Inter-Connection between Soil Biodiversity

and Ecosystem Function as Affected by Land Use across Europe". In : *Applied Soil Ecology* 97, p. 112–124. DOI : 10.1016/j.apsoil.2015.08.006.

Crowther TW, Boddy L et Maynard DS (2018). "The Use of Artificial Media in Fungal Ecology". In : *Fungal Ecology* 32, p. 87–91. DOI : 10.1016/j.funeco.2017.10.007.

Csardi G et Nepusz T (2006). "The Igraph Software Package for Complex Network Research". In : *InterJournal* Complex Systems, p. 1695.

Daly AJ, Baetens JM et De Baets B (2018). "Ecological Diversity : Measuring the Unmeasurable". In : *Mathematics* 6.7, p. 119. DOI : 10.3390/math6070119.

D'Amen M, Mod HK, Gotelli NJ et Guisan A (2018). "Disentangling Biotic Interactions, Environmental Filters, and Dispersal Limitation as Drivers of Species Co-Occurrence". In : *Ecography* 41.8, p. 1233–1244. DOI : 10.1111/ecog.03148.

Dannemiller KC, Lang-Yona N, Yamamoto N, Rudich Y et Peccia J (2014). "Combining Real-Time PCR and next-Generation DNA Sequencing to Provide Quantitative Comparisons of Fungal Aerosol Populations". In : *Atmospheric Environment* 84, p. 113–121. DOI : 10.1016/j.atmosenv.2013.11.036.

Das P, Ji B, Kovatcheva-Datchary P, Bäckhed F et Nielsen J (2018). "In Vitro Co-Cultures of Human Gut Bacterial Species as Predicted from Co-Occurrence Network Analysis". In : *PLOS ONE* 13.3, e0195161. DOI : 10.1371/journal.pone.0195161.

Davey JS, Vaughan IP, Andrew King R, Bell JR, Bohan DA, Bruford MW, Holland JM et Symondson WOC (2013). "Intraguild Predation in Winter Wheat : Prey Choice by a Common Epigeal Carabid Consuming Spiders". In : *Journal of Applied Ecology* 50.1, p. 271–279. DOI : 10.1111/1365-2664.12008.

Davey ML, Nybakken L, Kauserud H et Ohlson M (2009). "Fungal Biomass Associated with the Phyllosphere of Bryophytes and Vascular Plants". In : *Mycological Research* 113.11, p. 1254–1260. DOI : 10.1016/j.mycres.2009.08.001.

Davis NM, Proctor DM, Holmes SP, Relman DA et Callahan BJ (2018). "Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data". In : *Microbiome* 6.1, p. 226. DOI : 10.1186/s40168-018-0605-2.

Delière L, Petitgenet M, Goutouly JP, Forget D, Coulon T, Davidou L, Guilbault P, Christen M, Rochas A et Alonso Ugaglia A (2014). "Le réseau Ecoviti Bordeaux expérimente des systèmes de culture viticoles « bas-intrants »". In : *Phytoma* 673, p. 51–55.

Delmas E, Besson M, Brice MH, Burkle LA, Riva GVD, Fortin MJ, Gravel D, Guimarães PR, Hembry DH, Newman EA, Olesen JM, Pires MM, Yeakel JD et Poisot T (2019). "Analysing Ecological Networks of Species Interactions". In : *Biological Reviews* 94.1, p. 16–36. DOI : 10.1111/brv.12433.

Derocles SAP, Bohan DA, Dumbrell AJ, Kitson JJN, Massol F, Pauvert C, Plantegenest M, Vacher C et Evans DM (2018). "Chapter One - Biomonitoring for the 21st Century : Integrating Next-Generation Sequencing Into Ecological Network Analysis". In : *Advances in Ecological Research*. Sous la dir. de DA Bohan, AJ Dumbrell, G Woodward et M Jackson. T. 58. Next Generation Biomonitoring : Part 1. Academic Press, p. 1–62. DOI : 10.1016/bs.aecr.2017.12.001.

Desjardins-Proulx P, Laigle I, Poisot T et Gravel D (2017). "Ecological Interactions and the Netflix Problem". In : *PeerJ* 5, e3644. DOI : 10.7717/peerj.3644.

Deveau A, Bonito G, Uehling J, Paoletti M, Becker M, Bindschedler S, Hacquard S, Hervé V, Labbé J, Lastovetsky OA, Mieszkin S, Millet LJ, Vajna B, Junier P, Bonfante P, Krom BP, Olsson S, van Elsas JD et Wick LY (2018). "Bacterial–Fungal Interactions : Ecology, Mechanisms and Challenges". In : *FEMS Microbiology Reviews* 42.3, p. 335–352. DOI : 10.1093/femsre/fuy008.

Di Bella JM, Bao Y, Gloor GB, Burton JP et Reid G (2013). "High Throughput Sequencing Methods and Analysis for Microbiome Research". In : *Journal of Microbiological Methods* 95.3, p. 401–414. doi : 10.1016/j.mimet.2013.08.011.

Dighton J, White JF et Oudemans P, éds. (2005). *The Fungal Community : Its Organization and Role in the Ecosystem*. 3rd ed. Mycology Series 23. Boca Raton, FL : Taylor & Francis. 936 p. isbn : 978-0-8247-2355-2.

Dissanayake AJ, Purahong W, Wubet T, Hyde KD, Zhang W, Xu H, Zhang G, Fu C, Liu M, Xing Q, Li X et Yan J (2018). "Direct Comparison of Culture-Dependent and Culture-Independent Molecular Approaches Reveal the Diversity of Fungal Endophytic Communities in Stems of Grapevine (Vitis Vinifera)". In : *Fungal Diversity* 90.1, p. 85–107. doi : 10.1007/s13225-018-0399-3.

Dopheide A, Xie D, Buckley TR, Drummond AJ et Newcomb RD (2019). "Impacts of DNA Extraction and PCR on DNA Metabarcoding Estimates of Soil Biodiversity". In : *Methods in Ecology and Evolution* 10.1, p. 120–133. doi : 10.1111/2041-210X.13086.

Döring J, Collins C, Frisch M et Kauer R (2019). "Organic and Biodynamic Viticulture Affect Biodiversity and Properties of Vine and Wine : A Systematic Quantitative Review". In : *American Journal of Enology and Viticulture* 70.3, p. 221–242. doi : 10.5344/ajev.2019.18047.

Doster MA et Schnathorst WC (1985). "Comparative Susceptibility of Various Grapevine Cultivars to the Powdery Mildew Fungus Uncinula Necator". In : *American Journal of Enology and Viticulture* 36.2, p. 101–104.

Dufour MC, Lambert C, Bouscaut J, Mérillon JM et Corio-Costet MF (2013). "Benzothiadiazole-Primed Defence Responses and Enhanced Differential Expression of Defence Genes in *Vitis Vinifera* Infected with Biotrophic Pathogens *Erysiphe Necator* and *Plasmopara Viticola* : Elicitation and Grapevine Responses to Mildews". In : *Plant Pathology* 62.2, p. 370–382. doi : 10.1111/j.1365-3059.2012.02628.x.

Dufour MC, Fontaine S, Montarry J et Corio-Costet MF (2011). "Assessment of Fungicide Resistance and Pathogen Diversity in Erysiphe Necator Using Quantitative Real-Time PCR Assays". In : *Pest Management Science* 67.1, p. 60–69. doi : 10.1002/ps.2032.

Dupont J, Dequin S, Giraud T, Tacon FL, Marsit S, Ropars J, Richard F et Selosse MA (2017). "Fungi as a Source of Food". In : *Microbiology Spectrum* 5.3, p. 1063–1085. doi : 10.1128/microbiolspec.FUNK-0030-2016.

Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P et Hacquard S (2018). "Microbial Interkingdom Interactions in Roots Promote Arabidopsis Survival". In : *Cell* 175.4, 973–983.e14. doi : 10.1016/j.cell.2018.10.020.

Durand A, Maillard F, Foulon J, Gweon HS, Valot B et Chalot M (2017). "Environmental Metabarcoding Reveals Contrasting Belowground and Aboveground Fungal Communities from Poplar at a Hg Phytomanagement Site". In : *Microbial Ecology* 74.4, p. 795–809. doi : 10.1007/s00248-017-0984-0.

Durling MB, Clemmensen KE, Stenlid J et Lindahl B (2011). *SCATA -An Efficient Bioinformatic Pipeline for Species Identification and Quantification after High-Throughput Sequencing of Tagged Amplicons*.

Edgar RC (2010). "Search and Clustering Orders of Magnitude Faster than BLAST". In : *Bioinformatics* 26.19, p. 2460–2461. doi : 10.1093/bioinformatics/btq461.

— (2016). *UNOISE2 : Improved Error-Correction for Illumina 16S and ITS Amplicon Sequencing*. preprint. Bioinformatics. doi : 10.1101/081257.

— (2018). "Updating the 97% Identity Threshold for 16S Ribosomal RNA OTUs". In : *Bioinformatics* 34.14, p. 2371–2375. doi : 10.1093/bioinformatics/bty113.

Edgar RC et Flyvbjerg H (2015). "Error Filtering, Pair Assembly and Error Correction for next-Generation Sequencing Reads". In : *Bioinformatics* 31.21, p. 3476–3482. doi : 10.1093/bioinformatics/btv401.

Edgar RC, Haas BJ, Clemente JC, Quince C et Knight R (2011). "UCHIME Improves Sensitivity and Speed of Chimera Detection". In : *Bioinformatics* 27.16, p. 2194–2200. doi : 10.1093/bioinformatics/btr381.

Ellis JG (2017). "Can Plant Microbiome Studies Lead to Effective Biocontrol of Plant Diseases ?" In : *Molecular Plant-Microbe Interactions* 30.3, p. 190–193. doi : 10.1094/MPMI-12-16-0252-CR.

English-Loeb G, Norton AP, Gadoury D, Seem R et Wilcox W (2007). "Biological Control of Grape Powdery Mildew Using Mycophagous Mites". In : *Plant Disease* 91.4, p. 421–429. doi : 10.1094/PDIS-91-4-0421.

Eren AM, Vineis JH, Morrison HG et Sogin ML (2013). "A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End Technology". In : *PLOS ONE* 8.6, e66643. doi : 10.1371/journal.pone.0066643.

Eschen R, Hunt S, Mykura C, Gange AC et Sutton BC (2010). "The Foliar Endophytic Fungal Community Composition in Cirsium Arvense Is Affected by Mycorrhizal Colonization and Soil Nutrient Content". In : *Fungal Biology* 114.11, p. 991–998. doi : 10.1016/j.funbio.2010.09.009.

Escudie F, Auer L, Bernard M, Cauquil L, Vidal K, Maman S, Mariadassou M, Hernandez-Raquet G et Pascal G (2015). "FROGS : Find Rapidly OTU with Galaxy Solution". In : Environmental Genomic Conference. Montpellier, France.

Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S et Pascal G (2018). "FROGS : Find, Rapidly, OTUs with Galaxy Solution". In : *Bioinformatics* 34.8. Sous la dir. de B Berger, p. 1287–1294. doi : 10.1093/bioinformatics/btx791.

Esser DS, Leveau JHJ, Meyer KM et Wiegand K (2015). "Editor's Choice : Spatial Scales of Interactions among Bacteria and between Bacteria and the Leaf Surface". In : *FEMS Microbiology Ecology* 91.3. doi : 10.1093/femsec/fiu034. pmid : 25764562.

European Commision (2007). "Council Regulation (EC) No 834/2007 of 28 June 2007 on Organic Production and Labelling of Organic Products and Repealing Regulation (EEC) No 2092/91". In : *Official Journal of the European Union* 50, p. 1–23.

— (2009). "Directive 2009/128/EC of the European Parliament and of the Council of 21 October 2009 Establishing a Framework for Community Action to Achieve the Sustainable Use of Pesticides". In : *Official Journal of the European Union* 52, p. 71–86.

Evans DM, Kitson JJN, Lunt DH, Straw NA et Pocock MJO (2016). "Merging DNA Metabarcoding and Ecological Network Analysis to Understand and Build Resilient Terrestrial Ecosystems". In : *Functional Ecology* 30.12, p. 1904–1916. doi : 10.1111/1365-2435.12659.

Ewing B et Green P (1998). "Base-Calling of Automated Sequencer Traces Using *Phred*. II. Error Probabilities". In : *Genome Research* 8.3, p. 186–194. doi : 10.1101/gr.8.3.186.

Falkowski PG, Fenchel T et Delong EF (2008). "The Microbial Engines That Drive Earth's Biogeochemical Cycles". In : *Science* 320.5879, p. 1034–1039. doi : 10.1126/science.1153213. pmid : 18497287.

Fang H, Huang C, Zhao H et Deng M (2015). "CCLasso : Correlation Inference for Compositional Data through Lasso". In : *Bioinformatics* 31.19, p. 3172–3180. doi : 10.1093/bioinformatics/btv349.

— (2017). "gCoda : Conditional Dependence Network Inference for Compositional Data". In : *Journal of Computational Biology* 24.7, p. 699–708. doi : 10.1089/cmb.2017.0054.

Farré-Armengol G et Junker RR (2019). "Bacteria Associated with Flowers Affect the Reproductive Success of Brassica Napus L. via Scent-Mediated Interactions with Pollinators". In : *bioRxiv*, p. 685453. doi : 10.1101/685453.

Faust K, Lima-Mendez G, Lerat JS, Sathirapongsasuti JF, Knight R, Huttenhower C, Lenaerts T et Raes J (2015). "Cross-Biome Comparison of Microbial Association Networks". In : *Frontiers in Microbiology* 6, p. 1200. doi : 10.3389/fmicb.2015.01200.

Faust K et Raes J (2012). "Microbial Interactions : From Networks to Models". In : *Nature Reviews Microbiology* 10.8, p. 538–550. doi : 10.1038/nrmicro2832.

— (2016). "CoNet App : Inference of Biological Association Networks Using Cytoscape". In : *F1000Research* 5, p. 1519. doi : 10.12688/f1000research.9050.2.

Fávaro LCdL, Sebastianes FLdS et Araújo WL (2012). "Epicoccum Nigrum P16, a Sugarcane Endophyte, Produces Antifungal Compounds and Induces Root Growth". In : *PLOS ONE* 7.6, e36826. doi : 10.1371/journal.pone.0036826.

Feibelman T, Bayman P et Cibula WG (1994). "Length Variation in the Internal Transcribed Spacer of Ribosomal DNA in Chanterelles". In : *Mycological Research* 98.6, p. 614–618. doi : 10.1016/S0953-7562(09)80407-3.

Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR et Gloor GB (2014). "Unifying the Analysis of High-Throughput Sequencing Datasets : Characterizing RNA-Seq, 16S rRNA Gene Sequencing and Selective Growth Experiments by Compositional Data Analysis". In : *Microbiome* 2, p. 15. doi : 10.1186/2049-2618-2-15.

Fernandes AD, Vu MTHQ, Edward LM, Macklaim JM et Gloor GB (2018). "A Reproducible Effect Size Is More Useful than an Irreproducible Hypothesis Test to Analyze High Throughput Sequencing Datasets". In : arXiv : 1809.02623 [q-bio].

Fierer N, Jackson JA, Vilgalys R et Jackson RB (2005). "Assessment of Soil Microbial Community Structure by Use of Taxon-Specific Quantitative PCR Assays". In : *Applied and Environmental Microbiology* 71.7, p. 4117–4120. doi : 10.1128/AEM.71.7.4117-4120.2005.

Fiers M, Edel-Hermann V, Héraud C, Gautheron N, Chatot C, Hingrat YL, Bouchek-Mechiche K et Steinberg C (2011). "Genetic Diversity of *Rhizoctonia Solani* Associated with Potato Tubers in France". In : *Mycologia* 103.6, p. 1230–1244. doi : 10.3852/10-231.

Fisher CK et Mehta P (2014). "Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries Using Sparse Linear Regression". In : *PLOS ONE* 9.7, e102451. doi : 10.1371/journal.pone.0102451.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL et Gurr SJ (2012). "Emerging Fungal Threats to Animal, Plant and Ecosystem Health". In : *Nature* 484.7393, p. 186–194. doi : 10.1038/nature10947.

Fisher PJ et Petrini O (1992). "Fungal Saprobes and Pathogens as Endophytes of Rice (Oryza Sativa L.)" In : *New Phytologist* 120.1, p. 137–143. doi : 10.1111/j.1469-8137.1992.tb01066.x.

Fitzpatrick CR, Lu-Irving P, Copeland J, Guttman DS, Wang PW, Baltrus DA, Dlugosch KM et Johnson MTJ (2018). "Chloroplast Sequence Variation and the Efficacy of Peptide Nucleic Acids for Blocking Host Amplification in Plant Microbiome Studies". In : *Microbiome* 6.1, p. 144. doi : 10.1186/s40168-018-0534-0.

Fort T, Pauvert C, Zanne AE, Ovaskainen O, Caignard T, Barret M, Compant S, Hampe A, Delzon S et Vacher C (2019). "Maternal Effects and Environmental Filtering Shape Seed Fungal Communities in Oak Trees". In : *bioRxiv*, p. 691121. doi : 10.1101/691121.

Fort T, Robin C, Capdevielle X, Delière L et Vacher C (2016). "Foliar Fungal Communities Strongly Differ between Habitat Patches in a Landscape Mosaic". In : *PeerJ* 4, e2656. doi : 10.7717/peerj.2656.

Foster KR et Bell T (2012). "Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species". In : *Current Biology* 22.19, p. 1845–1850. doi : 10.1016/j.cub.2012.08.005.

Freilich MA, Wieters E, Broitman BR, Marquet PA et Navarrete SA (2018). "Species Co-Occurrence Networks : Can They Reveal Trophic and Non-Trophic Interactions in Ecological Communities ?" In : *Ecology* 99.3, p. 690–699. doi : 10.1002/ecy.2142.

Friedman J, Hastie T et Tibshirani R (2008). "Sparse Inverse Covariance Estimation with the Graphical Lasso". In : *Biostatistics* 9.3, p. 432–441. doi : 10.1093/biostatistics/kxm045.

Friedman J et Alm EJ (2012). "Inferring Correlation Networks from Genomic Survey Data". In : *PLoS Computational Biology* 8.9, e1002687.

Friedman J, Higgins LM et Gore J (2017). "Community Structure Follows Simple Assembly Rules in Microbial Microcosms". In : *Nature Ecology & Evolution* 1.5, p. 0109. doi : 10.1038/s41559-017-0109.

Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C et Hansen AJ (2017). "Algorithm for Post-Clustering Curation of DNA Amplicon Data Yields Reliable Biodiversity Estimates". In : *Nature Communications* 8.1, p. 1188. doi : 10.1038/s41467-017-01312-x.

Gadoury DM, Cadle-Davidson L, Wilcox WF, Dry IB, Seem RC et Milgroom MG (2012). "Grapevine Powdery Mildew (Erysiphe Necator) : A Fascinating System for the Study of the Biology, Ecology and Epidemiology of an Obligate Biotroph". In : *Molecular Plant Pathology* 13.1, p. 1–16. doi : 10.1111/j.1364-3703.2011.00728.x.

Galan M, Razzauti M, Bard E, Bernard M, Brouat C, Charbonnel N, Dehne-Garcia A, Loiseau A, Tatard C, Tamisier L, Vayssier-Taussat M, Vignes H et Cosson JF (2016). "16S rRNA Amplicon Sequencing for Epidemiological Surveys of Bacteria in Wildlife". In : *mSystems* 1.4, e00032–16. doi : 10.1128/mSystems.00032-16.

Ganley ARD et Kobayashi T (2007). "Highly Efficient Concerted Evolution in the Ribosomal DNA Repeats : Total rDNA Repeat Variation Revealed by Whole-Genome Shotgun Sequence Data". In : *Genome Research* 17.2, p. 184–191. doi : 10.1101/gr.5457707. pmid : 17200233.

Gao X, Huynh BT, Guillemot D, Glaser P et Opatowski L (2018). "Inference of Significant Microbial Interactions from Longitudinal Metagenomics Sequencing Data". In : *bioRxiv*, p. 305326. doi : 10.1101/305326.

Gardes M et Bruns TD (1993). "ITS Primers with Enhanced Specificity for Basidiomycetes - Application to the Identification of Mycorrhizae and Rusts". In : *Molecular Ecology* 2.2, p. 113–118. doi : 10.1111/j.1365-294X.1993.tb00005.x.

Gdanetz K, Benucci GMN, Vande Pol N et Bonito G (2017). "CONSTAX : A Tool for Improved Taxonomic Resolution of Environmental Fungal ITS Sequences". In : *BMC Bioinformatics* 18.1, p. 538. doi : 10.1186/s12859-017-1952-x.

Ghoul M et Mitri S (2016). "The Ecology and Evolution of Microbial Competition". In : *Trends in Microbiology* 24.10, p. 833–845. doi : 10.1016/j.tim.2016.06.011.

Gilbert JA et Neufeld JD (2014). "Life in a World without Microbes". In : *PLoS Biology* 12.12, e1002020. doi : 10.1371/journal.pbio.1002020.

Gloor GB, Macklaim JM, Pawlowsky-Glahn V et Egozcue JJ (2017). "Microbiome Datasets Are Compositional : And This Is Not Optional". In : *Frontiers in Microbiology* 8, p. 2224. doi : 10.3389/fmicb.2017.02224.

Glushakova AM et Chernov IY (2010). "Seasonal Dynamics of the Structure of Epiphytic Yeast Communities". In : *Microbiology* 79.6, p. 830–839. doi : 10.1134/S0026261710060160.

Gonze D, Coyte KZ, Lahti L et Faust K (2018). "Microbial Communities as Dynamical Systems". In : *Current Opinion in Microbiology*. Microbiota 44, p. 41–49. doi : 10.1016/j.mib.2018.07.004.

Goodwin S, McPherson JD et McCombie WR (2016). "Coming of Age : Ten Years of next-Generation Sequencing Technologies". In : *Nature Reviews Genetics* 17.6, p. 333–351. doi : 10.1038/nrg.2016.49.

Grangeteau C, David V, Hervé A, Guilloux-Benatier M et Rousseaux S (2017). "The Sensitivity of Yeasts and Yeasts-like Fungi to Copper and Sulfur Could Explain Lower Yeast Biodiversity in Organic Vineyards". In : *FEMS Yeast Research* 17.8, p. 1–6. doi : 10.1093/femsyr/fox092.

Gweon HS, Oliver A, Taylor J, Booth T, Gibbs M, Read DS, Griffiths RI et Schonrogge K (2015). "PIPITS : An Automated Pipeline for Analyses of Fungal Internal Transcribed Spacer Sequences from the Illumina Sequencing Platform". In : *Methods in Ecology and Evolution* 6.8, p. 973–980. doi : 10.1111/2041-210X.12399. pmid : 27570615.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methe B, DeSantis TZ, The Human Microbiome Consortium, Petrosino JF, Knight R et Birren BW (2011). "Chimeric 16S rRNA Sequence Formation and Detection in Sanger and 454-Pyrosequenced PCR Amplicons". In : *Genome Research* 21.3, p. 494–504. doi : 10.1101/gr.112730.110.

Hacquard S, Spaepen S, Garrido-Oter R et Schulze-Lefert P (2017). "Interplay Between Innate Immunity and the Plant Microbiota". In : *Annual Review of Phytopathology* 55.1, p. 565–589. doi : 10.1146/annurev-phyto-080516-035623.

Halwachs B, Madhusudhan N, Krause R, Nilsson RH, Moissl-Eichinger C, Högenauer C, Thallinger GG et Gorkiewicz G (2017). "Critical Issues in Mycobiota Analysis". In : *Frontiers in Microbiology* 8, p. 180. doi : 10.3389/fmicb.2017.00180.

Hamonts K, Trivedi P, Garg A, Janitz C, Grinyer J, Holford P, Botha FC, Anderson IC et Singh BK (2018). "Field Study Reveals Core Plant Microbiota and Relative Importance of Their Drivers". In : *Environmental Microbiology* 20.1, p. 124–140. doi : 10.1111/1462-2920.14031.

Hardoim PR, van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A, Döring M et Sessitsch A (2015). "The Hidden World within Plants : Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes". In : *Microbiology and Molecular Biology Reviews* 79.3, p. 293–320. doi : 10.1128/MMBR.00050-14.

Harris DJ (2016). "Inferring Species Interactions from Co-Occurrence Data with Markov Networks". In : *Ecology* 97.12, p. 3308–3314. doi : 10.1002/ecy.1605.

Harrison JG, Beltran LP, Buerkle CA, Cook D, Gardner DR, Parchman TL et Forister ML (2019). "A Suite of Rare Microbes Interacts with a Dominant, Heritable, Fungal Endophyte to Influence Plant Trait Expression". In : *bioRxiv*, p. 608729. doi : 10.1101/608729.

Harrison JG, Forister ML, Parchman TL et Koch GW (2016). "Vertical Stratification of the Foliar Fungal Community in the World's Tallest Trees". In : *American Journal of Botany* 103.12, p. 2087–2095. doi : 10.3732/ajb.1600277.

Hartman K, van der Heijden MGA, Wittwer RA, Banerjee S, Walser JC et Schlaeppi K (2018). "Cropping Practices Manipulate Abundance Patterns of Root and Soil Microbiome Members Paving the Way to Smart Farming". In : *Microbiome* 6, p. 14. doi : 10.1186/s40168-017-0389-9.

Hassani MA, Durán P et Hacquard S (2018). "Microbial Interactions within the Plant Holobiont". In : *Microbiome* 6.1, p. 58. doi : 10.1186/s40168-018-0445-0.

Hathaway NJ, Parobek CM, Juliano JJ et Bailey JA (2018). "SeekDeep : Single-Base Resolution de Novo Clustering for Amplicon Deep Sequencing". In : *Nucleic Acids Research* 46.4, e21–e21. doi : 10.1093/nar/gkx1201.

Hawkes CV et Connor EW (2017). “Translating Phytobiomes from Theory to Practice : Ecological and Evolutionary Considerations”. In : *Phytobiomes* 1.2, p. 57–69. doi : 10.1094/PBIOMES-05-17-0019-RVW.

Heard GM et Fleet GH (1987). “Occurrence and Growth of Killer Yeasts during Wine Fermentation”. In : *Applied and Environmental Microbiology* 53.9, p. 2171–2174. pmid : 16347438.

Hebert PDN, Cywinska A, Ball SL et deWaard JR (2003). “Biological Identifications through DNA Barcodes”. In : *Proceedings of the Royal Society of London. Series B : Biological Sciences* 270.1512, p. 313–321. doi : 10.1098/rspb.2002.2218.

Hertz M, Jensen IR, Jensen LØ, Thomsen SN, Winde J, Dueholm MS, Sørensen LH, Wollenberg RD, Sørensen HO, Sondergaard TE et Sørensen JL (2016). “The Fungal Community Changes over Time in Developing Wheat Heads”. In : *International Journal of Food Microbiology* 222, p. 30–39. doi : 10.1016/j.ijfoodmicro.2016.01.018.

Hibbett DS, Ohman A et Kirk PM (2009). “Fungal Ecology Catches Fire”. In : *New Phytologist* 184.2, p. 279–282. doi : 10.1111/j.1469-8137.2009.03042.x.

Hibbett DS et Taylor JW (2013). “Fungal Systematics : Is a New Age of Enlightenment at Hand ?” In : *Nature Reviews Microbiology* 11.2, p. 129–133. doi : 10.1038/nrmicro2963.

Hibbett D, Abarenkov K, Kõljalg U, Öpik M, Chai B, Cole J, Wang Q, Crous P, Robert V, Helgason T, Herr JR, Kirk P, Lueschow S, O’Donnell K, Nilsson RH, Oono R, Schoch C, Smyth C, Walker DM, Porras-Alfaro A, Taylor JW et Geiser DM (2016). “Sequence-Based Classification and Identification of Fungi”. In : *Mycologia* 108.6, p. 1049–1068. doi : 10.3852/16-130.

Hilber-Bodmer M, Schmid M, Ahrens CH et Freimoser FM (2017). “Competition Assays and Physiological Experiments of Soil and Phyllosphere Yeasts Identify Candida Subhashii as a Novel Antagonist of Filamentous Fungi”. In : *BMC Microbiology* 17.1, p. 4. doi : 10.1186/s12866-016-0908-z.

Hill MO (1973). “Diversity and Evenness : A Unifying Notation and Its Consequences”. In : *Ecology* 54.2, p. 427–432. doi : 10.2307/1934352.

Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, Kitano TK, Hodel MR, Petersen JF, Wyatt PW, Steenblock ER, Shah PH, Bousse LJ, Troup CB, Mellen JC, Wittmann DK, Erndt NG, Cauley TH, Koehler RT, So AP, Dube S, Rose KA, Montesclaros L, Wang S, Stumbo DP, Hodges SP, Romine S, Milanovich FP, White HE, Regan JF, Karlin-Neumann GA, Hindson CM, Saxonov S et Colston BW (2011). “High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number”. In : *Analytical Chemistry* 83.22, p. 8604–8610. doi : 10.1021/ac202028g.

Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL et Tewari M (2013). “Absolute Quantification by Droplet Digital PCR versus Analog Real-Time PCR”. In : *Nature Methods* 10.10, p. 1003–1005. doi : 10.1038/nmeth.2633.

Hirano H et Takemoto K (2019). “Difficulty in Inferring Microbial Community Structure Based on Co-Occurrence Network Approaches”. In : *BMC Bioinformatics* 20.1, p. 329. doi : 10.1186/s12859-019-2915-1.

Hoek TA, Axelrod K, Biancalani T, Yurtsev EA, Liu J et Gore J (2016). “Resource Availability Modulates the Cooperative and Competitive Nature of a Microbial Cross-Feeding Mutualism”. In : *PLOS Biol* 14.8, e1002540. doi : 10.1371/journal.pbio.1002540.

Hoffman MT et Arnold AE (2010). “Diverse Bacteria Inhabit Living Hyphae of Phylogenetically Diverse Fungal Endophytes”. In : *Applied and Environmental Microbiology* 76.12, p. 4063–4075. doi : 10.1128/AEM.02928-09.

Holt RD et Bonsall MB (2017). "Apparent Competition". In : *Annual Review of Ecology, Evolution, and Systematics* 48.1, p. 447–471. doi : 10.1146/annurev-ecolsys-110316-022628.

Hornung BVH, Zwittink RD et Kuijper EJ (2019). "Issues and Current Standards of Controls in Microbiome Research". In : *FEMS Microbiology Ecology* 95.5. doi : 10.1093/femsec/fiz045.

Huang YL, Zimmerman NB et Arnold AE (2018). "Observations on the Early Establishment of Foliar Endophytic Fungi in Leaf Discs and Living Leaves of a Model Woody Angiosperm, Populus Trichocarpa (Salicaceae)". In : *Journal of Fungi* 4.2, p. 58. doi : 10.3390/jof4020058.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC et Banfield JF (2016). "A New View of the Tree of Life". In : *Nature Microbiology* 1.5, p. 16048. doi : 10.1038/nmicrobiol.2016.48.

Ihrmark K, Bödeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandström-Durling M, Clemmensen KE et Lindahl BD (2012). "New Primers to Amplify the Fungal ITS2 Region – Evaluation by 454-Sequencing of Artificial and Natural Communities". In : *FEMS Microbiology Ecology* 82.3, p. 666–677. doi : 10.1111/j.1574-6941.2012.01437.x.

Ings TC, Montoya JM, Bascompte J, Blüthgen N, Brown L, Dormann CF, Edwards F, Figueroa D, Jacob U, Jones JI, Lauridsen RB, Ledger ME, Lewis HM, Olesen JM, Veen FJFV, Warren PH et Woodward G (2009). "Review : Ecological Networks – beyond Food Webs". In : *Journal of Animal Ecology* 78.1, p. 253–269. doi : 10.1111/j.1365-2656.2008.01460.x.

Itagaki K et Shibuya T (2017). "Differences in Early Hyphal Development of Podosphaera Xanthii on Cucumis Sativus Leaves Acclimatized to High or Low Relative Humidity". In : *Botany* 96.1, p. 67–71. doi : 10.1139/cjb-2017-0159.

Jaccard P (1900). "Contribution Au Probleme de l'immigration Post-Glaciare de La Flore Alpine". In : *Bull Soc Vaudoise Sci Nat* 36, p. 87–130.

Jacobs BKM, Goetghebeur E, Vandesompele J, De Ganck A, Nijs N, Beckers A, Papazova N, Roosens NH et Clement L (2017). "Model-Based Classification for Digital PCR : Your Umbrella for Rain". In : *Analytical Chemistry* 89.8, p. 4461–4467. doi : 10.1021/acs.analchem.6b04208.

Jakuschkin B, Fievet V, Schwaller L, Fort T, Robin C et Vacher C (2016). "Deciphering the Pathobiome : Intra- and Interkingdom Interactions Involving the Pathogen *Erysiphe Alphitoides*". In : *Microbial Ecology* 72.4, p. 870–880. doi : 10.1007/s00248-016-0777-x.

Jost L (2006). "Entropy and Diversity". In : *Oikos* 113.2, p. 363–375. doi : 10.1111/j.2006.0030-1299.14714.x.

Jumpponen A et Jones KL (2009). "Massively Parallel 454 Sequencing Indicates Hyperdiverse Fungal Communities in Temperate Quercus Macrocarpa Phyllosphere". In : *New Phytologist* 184.2, p. 438–448. doi : 10.1111/j.1469-8137.2009.02990.x.

Kamada N, Chen GY, Inohara N et Núñez G (2013). "Control of Pathogens and Pathobionts by the Gut Microbiota". In : *Nature Immunology* 14.7, p. 685–690. doi : 10.1038/ni.2608.

Kamoun S, Furzer O, Jones JDG, Judelson HS, Ali GS, Dalio RJD, Roy SG, Schena L, Zambounis A, Panabières F, Cahill D, Ruocco M, Figueiredo A, Chen XR, Hulvey J, Stam R, Lamour K, Gijzen M, Tyler BM, Grünwald NJ, Mukhtar MS, Tomé DFA, Tör M, Van Den Ackerveken G, McDowell J, Daayf F, Fry WE, Lindqvist-Kreuze H, Meijer HJG, Petre B, Ristaino J, Yoshida K, Birch PRJ et Govers F (2015). "The Top 10 Oomycete Pathogens in Molecular Plant Pathology : Top 10 Oomycete Plant Pathogens". In : *Molecular Plant Pathology* 16.4, p. 413–434. doi : 10.1111/mpp.12190.

Karimi B, Dequiedt S, Terrat S, Jolivet C, Arrouays D, Wincker P, Cruaud C, Bispo A, Prévost-Bouré NC et Ranjard L (2019). "Biogeography of Soil Bacterial Networks along a Gradient of Cropping Intensity". In : *Scientific Reports* 9.1, p. 3812. doi : 10.1038/s41598-019-40422-y.

Karimi B, Maron PA, Boure NCP, Bernard N, Gilbert D et Ranjard L (2017). "Microbial Diversity and Ecological Networks as Indicators of Environmental Quality". In : *Environmental Chemistry Letters* 15.2, p. 265–281. doi : 10.1007/s10311-017-0614-6.

Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH et Albertsen M (2018). "Retrieval of a Million High-Quality, Full-Length Microbial 16S and 18S rRNA Gene Sequences without Primer Bias". In : *Nature Biotechnology* 36.2, p. 190–195. doi : 10.1038/nbt.4045.

Kassemeyer HH (2017). "Fungi of Grapes". In : *Biology of Microorganisms on Grapes, in Must and in Wine*. Sous la dir. de H König, G Unden et J Fröhlich. Cham : Springer International Publishing, p. 103–132. isbn : 978-3-319-60020-8 978-3-319-60021-5. doi : 10.1007/978-3-319-60021-5_4.

Kecskeméti E, Berkelmann-Löhnertz B et Reineke A (2016). "Are Epiphytic Microbial Communities in the Carposphere of Ripening Grape Clusters (Vitis Vinifera L.) Different between Conventional, Organic, and Biodynamic Grapes ?" In : *PLOS ONE* 11.8, e0160852. doi : 10.1371/journal.pone.0160852.

Kehe J, Kulesa A, Ortiz A, Ackerman CM, Thakku SG, Sellers D, Kuehn S, Gore J, Friedman J et Blainey PC (2019). "Massively Parallel Screening of Synthetic Microbial Communities". In : *Proceedings of the National Academy of Sciences* 116.26, p. 12804–12809. doi : 10.1073/pnas.1900102116. pmid : 31186361.

Kemen E (2014). "Microbe–Microbe Interactions Determine Oomycete and Fungal Host Colonization". In : *Current Opinion in Plant Biology* 20, p. 75–81. doi : 10.1016/j.pbi.2014.04.005.

Kemen E et Jones JDG (2012). "Obligate Biotroph Parasitism : Can We Link Genomes to Lifestyles ?" In : *Trends in Plant Science* 17.8, p. 448–457. doi : 10.1016/j.tplants.2012.04.005.

Kemler M, Witfeld F, Begerow D et Yurkov A (2017). "Phylloplane Yeasts in Temperate Climates". In : *Yeasts in Natural Ecosystems : Diversity*. Sous la dir. de P Buzzini, MA Lachance et A Yurkov. Springer International Publishing, p. 171–197. isbn : 978-3-319-62683-3. doi : 10.1007/978-3-319-62683-3_6.

Kennedy PG, Cline LC et Song Z (2018). "Probing Promise versus Performance in Longer Read Fungal Metabarcoding". In : *New Phytologist* 217.3, p. 973–976. doi : 10.1111/nph.14883.

Kerdraon L, Barret M, Laval V et Suffert F (2019). "Differential Dynamics of Microbial Community Networks Help Identify Microorganisms Interacting with Residue-Borne Pathogens : The Case of Zymoseptoria Tritici in Wheat". In : *Microbiome* 7.1, p. 125. doi : 10.1186/s40168-019-0736-0.

Kernaghan G, Mayerhofer M et Griffin A (2017). "Fungal Endophytes of Wild and Hybrid Vitis Leaves and Their Potential for Vineyard Biocontrol". In : *Canadian Journal of Microbiology*. Canadian Journal of Microbiology 63.7, p. 583–595. doi : 10.1139/cjm-2016-0740. pmid : 28407475.

Kiss L (1998). "Natural Occurrence of Ampelomyces Intracellular Mycoparasites in Mycelia of Powdery Mildew Fungi". In : *New Phytologist* 140.4, p. 709–714. doi : 10.1046/j.1469-8137.1998.00316.x.

Klassen R, Schaffrath R, Buzzini P et Ganter PF (2017). "Antagonistic Interactions and Killer Yeasts". In : *Yeasts in Natural Ecosystems : Ecology*. Sous la dir. de P Buzzini, MA Lachance et A Yurkov. Cham : Springer International Publishing, p. 229–275. isbn : 978-3-319-61574-5 978-3-319-61575-2. doi : 10.1007/978-3-319-61575-2_9.

Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R, von Mering C et Vorholt JA (2012). "Metaproteogenomic Analysis of Microbial Communities in the Phyllosphere and Rhizosphere of Rice". In : *The ISME Journal* 6.7, p. 1378–1390. doi : 10.1038/ismej.2011.192.

Kobayashi DY et Crouch JA (2009). "Bacterial/Fungal Interactions : From Pathogens to Mutualistic Endosymbionts". In : *Annual Review of Phytopathology* 47.1, p. 63–82. doi : 10.1146/annurev-phyto-080508-081729.

Koch H et Schmid-Hempel P (2011). "Socially Transmitted Gut Microbiota Protect Bumble Bees against an Intestinal Parasite". In : *Proceedings of the National Academy of Sciences* 108.48, p. 19288–19292. doi : 10.1073/pnas.1110474108. pmid : 22084077.

Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Põldmaa K, Saag L, Saar I, Schüssler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M et Larsson KH (2013). "Towards a Unified Paradigm for Sequence-Based Identification of Fungi". In : *Molecular Ecology* 22.21, p. 5271–5277. doi : 10.1111/mec.12481.

Kosel J, Raspor P et Čadež N (2019). "Maximum Residue Limit of Fungicides Inhibits the Viability and Growth of Desirable Non-Saccharomyces Wine Yeasts". In : *Australian Journal of Grape and Wine Research* 25.1, p. 43–52. doi : 10.1111/ajgw.12364.

Koskella B, Meaden S, Crowther WJ, Leimu R et Metcalf CJE (2017). "A Signature of Tree Health ? Shifts in the Microbiome and the Ecological Drivers of Horse Chestnut Bleeding Canker Disease". In : *New Phytologist* 215.2, p. 737–746. doi : 10.1111/nph.14560.

Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N et Knight R (2010). "Microbial Community Resemblance Methods Differ in Their Ability to Detect Biologically Relevant Patterns". In : *Nature Methods* 7.10, p. 813–819. doi : 10.1038/nmeth.1499.

Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ et Bonneau RA (2015). "Sparse and Compositionally Robust Inference of Microbial Ecological Networks". In : *PLOS Computational Biology* 11.5. Sous la dir. de C von Mering, e1004226. doi : 10.1371/journal.pcbi.1004226.

Kurtzman CP, Fell JW et Boekhout T (2011). "Definition, Classification and Nomenclature of the Yeasts". In : *The Yeasts : A Taxonomic Study*, p. 3–5. isbn : 978-0-444-52149-1.

Lafferty KD, Dobson AP et Kuris AM (2006). "Parasites Dominate Food Web Links". In : *Proceedings of the National Academy of Sciences* 103.30, p. 11211–11216. doi : 10.1073/pnas.0604755103. pmid : 16844774.

Lagier JC, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, Levasseur A, Rolain JM, Fournier PE et Raoult D (2018). "Culturing the Human Microbiota and Culturomics". In : *Nature Reviews Microbiology* 16.9, p. 540–550. doi : 10.1038/s41579-018-0041-0.

Lamb PD, Hunter E, Pinnegar JK, Creer S, Davies RG et Taylor MI (2019). "How Quantitative Is Metabarcoding : A Meta-Analytical Approach". In : *Molecular Ecology* 28.2, p. 420–430. doi : 10.1111/mec.14920.

Laur J, Ramakrishnan GB, Labbé C, Lefebvre F, Spanu PD et Bélanger RR (2018). "Effectors Involved in Fungal–Fungal Interaction Lead to a Rare Phenomenon of Hyperbiotrophy in the Tritrophic System Biocontrol Agent–Powdery Mildew–Plant". In : *New Phytologist* 217.2, p. 713–725. doi : 10.1111/nph.14851.

Lauritzen SL (1996). *Graphical Models*. Oxford Statistical Science Series 17. Oxford : New York : Clarendon Press ; Oxford University Press. 298 p. isbn : 978-0-19-852219-5.

Layeghifard M, Hwang DM et Guttman DS (2017). "Disentangling Interactions in the Microbiome : A Network Perspective". In : *Trends in Microbiology* 25.3, p. 217–228. doi : 10.1016/j.tim.2016.11.008.

Lebreton L, Guillerm-Erckelboudt AY, Gazengel K, Linglin J, Ourry M, Glory P, Sarniguet A, Daval S, Manzanares-Dauleux MJ et Mougel C (2019). "Temporal Dynamics of Bacterial and Fungal Communities during the Infection of Brassica Rapa Roots by the Protist Plasmodiophora Brassicae". In : *PLOS ONE* 14.2. Sous la dir. de RA Wilson, e0204195. doi : 10.1371/journal.pone.0204195.

Lee G, Lee SH, Kim KM et Ryu CM (2017). "Foliar Application of the Leaf-Colonizing Yeast Pseudozyma Churashimaensis Elicits Systemic Defense of Pepper against Bacterial and Viral Pathogens". In : *Scientific Reports* 7, p. 39432. doi : 10.1038/srep39432.

Lefebvre M, Theil S, Ma Y et Candresse T (2019). "The VirAnnot Pipeline : A Resource for Automated Viral Diversity Estimation and Operational Taxonomy Units (OTU) Assignation for Virome Sequencing Data". In : *Phytobiomes Journal*. doi : 10.1094/PBIOMES-07-19-0037-A.

Legeay J, Husson C, Cordier T, Vacher C, Marcais B et Buée M (2019). "Comparison and Validation of Oomycetes Metabarcoding Primers for Phytophthora High Throughput Sequencing". In : *Journal of Plant Pathology* 101.3, p. 743–748. doi : 10.1007/s42161-019-00276-9.

Legendre P et De Caceres M (2013). "Beta Diversity as the Variance of Community Data : Dissimilarity Coefficients and Partitioning". In : *Ecology Letters* 16.8. Sous la dir. de H Morlon, p. 951–963. doi : 10.1111/ele.12141.

Legendre P et Legendre L (2012). *Numerical Ecology*. Third English edition. Developments in Environmental Modelling 24. Amsterdam : Elsevier. 990 p. isbn : 978-0-444-53868-0.

Lenth R (2018). *Emmeans : Estimated Marginal Means, Aka Least-Squares Means*. R package version 1.3.0.

Li C, Lim KMK, Chng KR et Nagarajan N (2016). "Predicting Microbial Interactions through Computational Approaches". In : *Methods* 102, p. 12–19. doi : 10.1016/j.ymeth.2016.02.019.

Li M, Wei Z, Wang J, Jousset A, Friman V, Xu Y, Shen Q et Pommier T (2019). "Facilitation Promotes Invasions in Plant-associated Microbial Communities". In : *Ecology Letters* 22.1. Sous la dir. de J Kloronomos, p. 149–158. doi : 10.1111/ele.13177.

Lidicker WZ (1979). "A Clarification of Interactions in Ecological Systems". In : *BioScience* 29.8, p. 475–477. doi : 10.2307/1307540.

Likar M, Stres B, Rusjan D, Potisek M et Regvar M (2017). "Ecological and Conventional Viticulture Gives Rise to Distinct Fungal and Bacterial Microbial Communities in Vineyard Soils". In : *Applied Soil Ecology* 113, p. 86–95. doi : 10.1016/j.apsoil.2017.02.007.

Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, Bittner L, Darzi Y, Wang J, Audic S, Berline L, Bontempi G, Cabello AM, Coppola L, Cornejo-Castillo FM, d'Ovidio F, Meester LD, Ferrera I, Garet-Delmas MJ, Guidi L, Lara E, Pesant S, Royo-Llonch M, Salazar G, Sánchez P, Sebastian M, Souffreau C, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Coordinators TO, Gorsky G, Not F, Ogata H, Speich S, Stemmann L, Weissenbach J, Wincker P, Acinas SG, Sunagawa S, Bork P, Sullivan MB, Karsenti E, Bowler C, de Vargas C et Raes J (2015). "Determinants of Community Structure in the Global Plankton Interactome". In : *Science* 348.6237, p. 1262073. doi : 10.1126/science.1262073. pmid : 25999517.

Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjøller R, Kõljalg U, Pennanen T, Rosendahl S, Stenlid J et Kauserud H (2013). "Fungal Community Analysis

by High-Throughput Sequencing of Amplified Markers - a User's Guide". In : *New Phytologist* 199.1, p. 288–299. DOI : 10.1111/nph.12243.

Linder C, Viret O et Spring J (2006). "Viticulture Intégrée et Bio-Organique : Synthèse de Sept Ans d'observations". In : *Revue suisse Vitic. Arboric. Hortic.* 38.4, p. 235–243.

Liu B, Gibbons T, Ghodsi M, Treangen T et Pop M (2011). "Accurate and Fast Estimation of Taxonomic Profiles from Metagenomic Shotgun Sequences". In : *BMC Genomics* 12 Suppl 2, S4. DOI : 10.1186/1471-2164-12-S2-S4. pmid : 21989143.

Liu H, Roeder K et Wasserman L (2010). "Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models". In : *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 2*. (Vancouver, British Columbia, Canada). NIPS'10. USA : Curran Associates Inc., p. 1432–1440.

Lo C et Marculescu R (2017). "MPLasso : Inferring Microbial Association Networks Using Prior Microbial Knowledge". In : *PLOS Computational Biology* 13.12, e1005915. DOI : 10.1371/journal.pcbi.1005915.

Lodge DJ, Fisher PJ et Sutton BC (1996). "Endophytic Fungi of *Manilkara Bidentata* Leaves in Puerto Rico". In : *Mycologia* 88.5, p. 733–738. DOI : 10.1080/00275514.1996.12026710.

Loit K, Adamson K, Bahram M, Puusepp R, Anslan S, Kiiker R, Drenkhan R et Tedersoo L (2019). "Relative Performance of MinION (Oxford Nanopore Technologies) vs. Sequel (Pacific Biosciences) Third-Generation Sequencing Instruments in Identification of Agricultural and Forest Fungal Pathogens". In : *Applied and Environmental Microbiology*, AEM.01368–19. DOI : 10.1128/AEM.01368-19. pmid : 31444199.

López-González RC, Gómez-Cornelio S, De la Rosa-García SC, Garrido E, Oropeza-Mariano O, Heil M et Partida-Martínez LP (2017). "The Age of Lima Bean Leaves Influences the Richness and Diversity of the Endophytic Fungal Community, but Not the Antagonistic Effect of Endophytes against Colletotrichum Lindemuthianum". In : *Fungal Ecology* 26, p. 1–10. DOI : 10.1016/j.funeco.2016.11.004.

Louca S, Parfrey LW et Doebeli M (2016a). "Decoupling Function and Taxonomy in the Global Ocean Microbiome". In : *Science* 353.6305, p. 1272–1277. DOI : 10.1126/science.aaf4507.

Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF et Doebeli M (2016b). "High Taxonomic Variability despite Stable Functional Structure across Microbial Communities". In : *Nature Ecology & Evolution* 1.1, p. 0015. DOI : 10.1038/s41559-016-0015.

Love MI, Huber W et Anders S (2014). "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2". In : *Genome Biology* 15.12, p. 550. DOI : 10.1186/s13059-014-0550-8.

Lozupone C et Knight R (2005). "UniFrac : A New Phylogenetic Method for Comparing Microbial Communities". In : *Applied and Environmental Microbiology* 71.12, p. 8228–8235. DOI : 10.1128/AEM.71.12.8228-8235.2005. pmid : 16332807.

Lucas R, Groeneveld J, Harms H, Johst K, Frank K et Kleinsteuber S (2017). "A Critical Evaluation of Ecological Indices for the Comparative Analysis of Microbial Communities Based on Molecular Datasets". In : *FEMS Microbiology Ecology* 93.1. Sous la dir. de T Lueders, fiw209. DOI : 10.1093/femsec/fiw209.

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, Rio TG del, Edgar RC, Eickhorst T, Ley RE, Hugenholtz P, Tringe SG et Dangl JL (2012). "Defining the Core *Arabidopsis Thaliana* Root Microbiome". In : *Nature* 488.7409, p. 86–90. DOI : 10.1038/nature11237.

Lupatini M, Korthals GW, de Hollander M, Janssens TKS et Kuramae EE (2017). "Soil Microbiome Is More Heterogeneous in Organic Than in Conventional Farming System". In : *Frontiers in Microbiology* 7, p. 2064. DOI : 10.3389/fmicb.2016.02064.

Lutz MC, Lopes CA, Rodriguez ME, Sosa MC et Sangorrín MP (2013). "Efficacy and Putative Mode of Action of Native and Commercial Antagonistic Yeasts against Postharvest Pathogens of Pear". In : *International Journal of Food Microbiology* 164.2, p. 166–172. doi : 10.1016/j.ijfoodmicro.2013.04.005.

Lynch MDJ et Neufeld JD (2015). "Ecology and Exploration of the Rare Biosphere". In : *Nature Reviews Microbiology* 13.4, p. 217–229. doi : 10.1038/nrmicro3400.

Ma A, Lu X, Gray C, Raybould A, Tamaddoni-Nezhad A, Woodward G et Bohan DA (2019). "Ecological Networks Reveal Resilience of Agro-Ecosystems to Changes in Farming Management". In : *Nature Ecology & Evolution* 3.2, p. 260. doi : 10.1038/s41559-018-0757-2.

Mace GM, Norris K et Fitter AH (2012). "Biodiversity and Ecosystem Services : A Multilayered Relationship". In : *Trends in Ecology & Evolution* 27.1, p. 19–26. doi : 10.1016/j.tree.2011.08.006.

Macfadyen S, Gibson R, Polaszek A, Morris RJ, Craze PG, Planqué R, Symondson WO et Memmott J (2009). "Do Differences in Food Web Structure between Organic and Conventional Farms Affect the Ecosystem Service of Pest Control ?" In : *Ecology Letters* 12.3, p. 229–238. doi : 10.1111/j.1461-0248.2008.01279.x.

Maczulak AE (2011). *Encyclopedia of Microbiology*. OCLC : 793464191. New York (N.Y.) : Facts on file. isbn : 978-1-4381-3406-2 978-0-8160-7364-1.

Mahé F, Rognes T, Quince C, de Vargas C et Dunthorn M (2014). "Swarm : Robust and Fast Clustering Method for Amplicon-Based Studies". In : *PeerJ* 2, e593. doi : 10.7717/peerj.593.

— (2015). "Swarm v2 : Highly-Scalable and High-Resolution Amplicon Clustering". In : *PeerJ* 3, e1420. doi : 10.7717/peerj.1420. pmid : 26713226.

Maignien L, DeForce EA, Chafee ME, Eren AM et Simmons SL (2014). "Ecological Succession and Stochastic Variation in the Assembly of Arabidopsis Thaliana Phyllosphere Communities". In : *mBio* 5.1, e00682–13. doi : 10.1128/mBio.00682-13.

Makiola A, Dickie IA, Holdaway RJ, Wood JR, Orwin KH, Lee CK et Glare TR (2018). "Biases in the Metabarcoding of Plant Pathogens Using Rust Fungi as a Model System". In : *MicrobiologyOpen*, e780. doi : 10.1002/mbo3.780.

Margulis L (1991). "Symbiogenesis and Symbionticism". In : *Symbiosis as a Source of Evolutionary Innovation*. Boston : Massachusetts Institute of Technology, p. 1–14. isbn : 978-0-262-51990-8.

Maropola MKA, Ramond JB et Trindade M (2015). "Impact of Metagenomic DNA Extraction Procedures on the Identifiable Endophytic Bacterial Diversity in Sorghum Bicolor (L. Moench)". In : *Journal of Microbiological Methods* 112, p. 104–117. doi : 10.1016/j.mimet.2015.03.012. pmid : 25775938.

Márquez LM, Redman RS, Rodriguez RJ et Roossinck MJ (2007). "A Virus in a Fungus in a Plant : Three-Way Symbiosis Required for Thermal Tolerance". In : *Science* 315.5811, p. 513–515. doi : 10.1126/science.1136237. pmid : 17255511.

Martin F, Msatef Y et Botton B (1983). "Nitrogen Assimilation in Mycorrhizas. I. Purification and Properties of the Nicotinamide Adenine Dinucleotide Phosphate-Specific Glutamate Dehydrogenase of the Ectomycorrhizal Fungus Cenococcum Graniforme". In : *New Phytologist* 93.3, p. 415–422. doi : 10.1111/j.1469-8137.1983.tb03441.x.

Martin M (2011). "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads". In : *EMBnet.journal* 17.1, p. 10. doi : 10.14806/ej.17.1.200.

Martinez ND (1992). "Constant Connectance in Community Food Webs". In : *The American Naturalist* 139.6, p. 1208–1218. doi : 10.1086/285382.

Martín-Fernández JA, Hron K, Templ M, Filzmoser P et Palarea-Albaladejo J (2015). "Bayesian-Multiplicative Treatment of Count Zeros in Compositional Data Sets". In : *Statistical Modelling : An International Journal* 15.2, p. 134–158. doi : 10.1177/1471082X14535524.

Martins G, Lauga B, Miot-Sertier C, Mercier A, Lonvaud A, Soulas ML, Soulas G et Masneuf-Pomarède I (2013). "Characterization of Epiphytic Bacterial Communities from Grapes, Leaves, Bark and Soil of Grapevine Plants Grown, and Their Relations". In : *PLOS ONE* 8.8, e73013. doi : 10.1371/journal.pone.0073013.

Martins G, Vallance J, Mercier A, Albertin W, Stamatopoulos P, Rey P, Lonvaud A et Masneuf-Pomarède I (2014). "Influence of the Farming System on the Epiphytic Yeasts and Yeast-like Fungi Colonizing Grape Berries during the Ripening Process". In : *International Journal of Food Microbiology* 177, p. 21–28. doi : 10.1016/j.ijfoodmicro.2014.02.002.

McCormack PJ, Wildman HG et Jeffries P (1994). "Production of Antibacterial Compounds by Phylloplane-Inhabiting Yeasts and Yeastlike Fungi." In : *Appl. Environ. Microbiol.* 60.3, p. 927–931. pmid : 8161185.

McGrath MJ et Andrews JH (2006). "Temporal Changes in Microscale Colonization of the Phylloplane by Aureobasidium Pullulans". In : *Applied and Environmental Microbiology* 72.9, p. 6234–6241. doi : 10.1128/AEM.00744-06. pmid : 16957250.

— (2007). "Role of Microbial Immigration in the Colonization of Apple Leaves by Aureobasidium Pullulans". In : *Applied and Environmental Microbiology* 73.4, p. 1277–1286. doi : 10.1128/AEM.01620-06. pmid : 17142367.

McLaren MR et Callahan BJ (2018). "In Nature, There Is Only Diversity". In : *mBio* 9.1, e02149–17. doi : 10.1128/mBio.02149-17.

McMurdie PJ et Holmes S (2013). "Phyloseq : An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data". In : *PLOS ONE* 8.4, e61217. doi : 10.1371/journal.pone.0061217.

— (2014). "Waste Not, Want Not : Why Rarefying Microbiome Data Is Inadmissible". In : *PLOS Computational Biology* 10.4, e1003531. doi : 10.1371/journal.pcbi.1003531.

Miura T, Sánchez R, Castañeda LE, Godoy K et Barbosa O (2017). "Is Microbial Terroir Related to Geographic Distance between Vineyards ?" In : *Environmental Microbiology Reports* 9.6, p. 742–749. doi : 10.1111/1758-2229.12589.

Moissl-Eichinger C, Pausan M, Taffner J, Berg G, Bang C et Schmitz RA (2017). "Archaea Are Interactive Components of Complex Microbiomes". In : *Trends in Microbiology*. doi : 10.1016/j.tim.2017.07.004.

Montoya JM, Pimm SL et Solé RV (2006). "Ecological Networks and Their Fragility". In : *Nature* 442.7100, p. 259–264. doi : 10.1038/nature04927.

Morin PJ (2011). *Community Ecology*. 2nd ed. OCLC : ocn700735116. Chichester, West Sussex ; Hoboken, NJ : Wiley. 407 p. isbn : 978-1-4443-3821-8 978-1-4051-2411-9.

Morriën E, Hannula SE, Snoek LB, Helmsing NR, Zweers H, de Hollander M, Soto RL, Bouffaud ML, Buée M, Dimmers W, Duyts H, Geisen S, Girlanda M, Griffiths RI, Jørgensen HB, Jensen J, Plassart P, Redecker D, Schmelz RM, Schmidt O, Thomson BC, Tisserant E, Uroz S, Winding A, Bailey MJ, Bonkowski M, Faber JH, Martin F, Lemanceau P, de Boer W, van Veen JA et van der Putten WH (2017). "Soil Networks Become More Connected and Take up More Carbon as Nature Restoration Progresses". In : *Nature Communications* 8, p. 14349. doi : 10.1038/ncomms14349.

Morrison-Whittle P et Goddard MR (2018). "From Vineyard to Winery : A Source Map of Microbial Diversity Driving Wine Fermentation". In : *Environmental Microbiology* 20.1, p. 75–84. doi : 10.1111/1462-2920.13960.

Morrison-Whittle P, Lee SA et Goddard MR (2017). "Fungal Communities Are Differentially Affected by Conventional and Biodynamic Agricultural Management Approaches in

Vineyard Ecosystems". In : *Agriculture, Ecosystems & Environment* 246, p. 306–313. DOI : 10.1016/j.agee.2017.05.022.

Motooka D, Fujimoto K, Tanaka R, Yaguchi T, Gotoh K, Maeda Y, Furuta Y, Kurakawa T, Goto N, Yasunaga T, Narazaki M, Kumanogoh A, Horii T, Iida T, Takeda K et Nakamura S (2017). "Fungal ITS1 Deep-Sequencing Strategies to Reconstruct the Composition of a 26-Species Community and Evaluation of the Gut Mycobiota of Healthy Japanese Individuals". In : *Frontiers in Microbiology* 8, p. 238. DOI : 10.3389/fmicb.2017.00238.

Mounier J, Monnet C, Vallaeys T, Arditi R, Sarthou AS, Helias A et Irlinger F (2008). "Microbial Interactions within a Cheese Microbial Community". In : *Applied and Environmental Microbiology* 74.1, p. 172–181. DOI : 10.1128/AEM.01338-07.

Müller T et Ruppel S (2014). "Progress in Cultivation-Independent Phyllosphere Microbiology". In : *FEMS Microbiology Ecology* 87.1, p. 2–17. DOI : 10.1111/1574-6941.12198.

Munck IA et Bonello P (2018). "Modern Approaches for Early Detection of Forest Pathogens Are Sorely Needed in the United States". In : *Forest Pathology*, e12445. DOI : 10.1111/efp.12445.

Muneret L, Mitchell M, Seufert V, Aviron S, Djoudi EA, Pétillon J, Plantegenest M, Thiéry D et Rusch A (2018). "Evidence That Organic Farming Promotes Pest Control". In : *Nature Sustainability* 1.7, p. 361–368. DOI : 10.1038/s41893-018-0102-4.

Murall C, Abbate J, Puelma Touzel M, Allen-Vercoe E, Alizon S, Froissart R et McCann K (2017). "Invasions of Host-Associated Microbiome Networks". In : *Advances in Ecological Research*. T. 57. Elsevier, p. 201–281. ISBN : 978-0-12-813328-6. DOI : 10.1016/bs.aecr.2016.11.002.

Muschelli J (2019). *Rscopus : Scopus Database 'API' Interface*.

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD et Buckler ES (2011). "Genetic Structure and Domestication History of the Grape". In : *Proceedings of the National Academy of Sciences* 108.9, p. 3530–3535. DOI : 10.1073/pnas.1009363108.

Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ, Huntley J, Ackermann GL, Berg-Lyons D, Holmes S, Caporaso JG et Knight R (2013). "Advancing Our Understanding of the Human Microbiome Using QIIME". In : *Methods in Enzymology*. T. 531. Elsevier, p. 371–444. ISBN : 978-0-12-407863-5. DOI : 10.1016/B978-0-12-407863-5.00019-8.

Naveed M, Mitter B, Reichenauer TG, Wieczorek K et Sessitsch A (2014). "Increased Drought Stress Resilience of Maize through Endophytic Colonization by Burkholderia Phytofirmans PsJN and Enterobacter Sp. FD17". In : *Environmental and Experimental Botany* 97, p. 30–39. DOI : 10.1016/j.envexpbot.2013.09.014.

Nearing JT, Douglas GM, Comeau AM et Langille MGI (2018). "Denoising the Denoisers : An Independent Evaluation of Microbiome Sequence Error-Correction Approaches". In : *PeerJ* 6, e5364. DOI : 10.7717/peerj.5364.

Németh MZ, Pintye A, Horváth ÁN, Vági P, Kovács GM, Gorfer M et Kiss L (2019). "Green Fluorescent Protein Transformation Sheds More Light on a Widespread Mycoparasitic Interaction". In : *Phytopathology* 109.8, p. 1404–1416. DOI : 10.1094/PHYTO-01-19-0013-R.

Nezhad AS et Geitmann A (2013). "The Cellular Mechanics of an Invasive Lifestyle". In : *Journal of Experimental Botany* 64.15, p. 4709–4728. DOI : 10.1093/jxb/ert254.

Nguyen NH, Smith D, Peay K et Kennedy P (2015). "Parsing Ecological Signal from Noise in next Generation Amplicon Sequencing". In : *New Phytologist* 205.4, p. 1389–1393. DOI : 10.1111/nph.12923.

Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS et Kennedy PG (2016). "FUNGuild : An Open Annotation Tool for Parsing Fungal Community Datasets by Ecological Guild". In : *Fungal Ecology* 20, p. 241–248. doi : 10.1016/j.funeco.2015.06.006.

Nichols RV, Vollmers C, Newsom LA, Wang Y, Heintzman PD, Leighton M, Green RE et Shapiro B (2018). "Minimizing Polymerase Biases in Metabarcoding". In : *Molecular Ecology Resources* 18.5, p. 927–939. doi : 10.1111/1755-0998.12895.

Nilsson RH, Abarenkov K, Veldre V, Nylinder S, De Wit P, Brosché S, Alfredsson JF, Ryberg M et Kristiansson E (2010). "An Open Source Chimera Checker for the Fungal ITS Region : COMPUTER PROGRAM NOTE". In : *Molecular Ecology Resources* 10.6, p. 1076–1081. doi : 10.1111/j.1755-0998.2010.02850.x.

Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P et Tedersoo L (2019). "Mycobiome Diversity : High-Throughput Sequencing and Identification of Fungi". In : *Nature Reviews Microbiology* 17.2, p. 95–109. doi : 10.1038/s41579-018-0116-y.

Nilsson RH, Sánchez-García M, Ryberg MK, Abarenkov K, Wurzbacher C et Kristiansson E (2017). "Read Quality-Based Trimming of the Distal Ends of Public Fungal DNA Sequences Is Nowhere near Satisfactory". In : *MycoKeys* 26, p. 13–24. doi : 10.3897/mycokeys.26.14591.

Nilsson RH, Tedersoo L, Ryberg M, Kristiansson E, Hartmann M, Unterseher M, Porter TM, Bengtsson-Palme J, Walker DM, de Sousa F, Gamper HA, Larsson E, Larsson KH, Kõljalg U, Edgar RC et Abarenkov K (2015). "A Comprehensive, Automatically Updated Fungal ITS Sequence Dataset for Reference-Based Chimera Control in Environmental Sequencing Efforts". In : *Microbes and environments* 30.2, p. 145–150. doi : 10.1264/jsme2.ME14121.

Nuss DL (2005). "Hypovirulence : Mycoviruses at the Fungal–Plant Interface". In : *Nature Reviews Microbiology* 3.8, p. 632–642. doi : 10.1038/nrmicro1206.

Odum E (1950). "Bird Populations of the Highlands (North Carolina) Plateau in Relation to Plant Succession and Avian Invasion." In : *Ecology* 31, p. 587–605.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E et Wagner H (2018). *Vegan : Community Ecology Package*.

Öpik M, Metsis M, Daniell TJ, Zobel M et Moora M (2009). "Large-Scale Parallel 454 Sequencing Reveals Host Ecological Group Specificity of Arbuscular Mycorrhizal Fungi in a Boreonemoral Forest". In : *New Phytologist* 184.2, p. 424–437. doi : 10.1111/j.1469-8137.2009.02920.x.

Osono T (2008). "Endophytic and Epiphytic Phyllosphere Fungi of Camellia Japonica : Seasonal and Leaf Age-Dependent Variations". In : *Mycologia* 100.3, p. 387–391. doi : 10.3852/07-110R1. pmid : 18751546.

Ottesen AR, González Peña A, White JR, Pettengill JB, Li C, Allard S, Rideout S, Allard M, Hill T, Evans P, Strain E, Musser S, Knight R et Brown E (2013). "Baseline Survey of the Anatomical Microbial Ecology of an Important Food Plant : Solanum Lycopersicum (Tomato)". In : *BMC Microbiology* 13, p. 114. doi : 10.1186/1471-2180-13-114.

Ovaskainen O, Tikhonov G, Norberg A, Blanchet FG, Duan L, Dunson D, Roslin T et Abrego N (2017). "How to Make More out of Community Data ? A Conceptual Framework and Its Implementation as Models and Software". In : *Ecology Letters* 20.5, p. 561–576. doi : 10.1111/ele.12757.

Palmer JM, Jusino MA, Banik MT et Lindner DL (2018). "Non-Biological Synthetic Spike-in Controls and the AMPtk Software Pipeline Improve Mycobiome Data". In : *PeerJ* 6, e4925. doi : 10.7717/peerj.4925.

Pancher M, Ceol M, Corneo PE, Longa CMO, Yousaf S, Pertot I et Campisano A (2012). "Fungal Endophytic Communities in Grapevines (Vitis Vinifera L.) Respond to Crop Management". In : *Applied and Environmental Microbiology* 78.12, p. 4308–4317. doi : 10.1128/AEM.07655-11.

Pandey B, Naidu RA et Grove GG (2018). "Detection and Analysis of Mycovirus-Related RNA Viruses from Grape Powdery Mildew Fungus Erysiphe Necator". In : *Archives of Virology*, p. 1–12. doi : 10.1007/s00705-018-3714-0.

Panstruga R et Kuhn H (2019). "Mutual Interplay between Phytopathogenic Powdery Mildew Fungi and Other Microorganisms". In : *Molecular Plant Pathology* 20.4, p. 463–470. doi : 10.1111/mpp.12771.

Parratt SR et Laine AL (2016). "The Role of Hyperparasitism in Microbial Pathogen Ecology and Evolution". In : *The ISME Journal* 10.8, p. 1815–1822. doi : 10.1038/ismej.2015.247.

Pauvert C, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I, Vallance J et Vacher C (2019a). "Bioinformatics Matters : The Accuracy of Plant and Soil Fungal Community Data Is Highly Dependent on the Metabarcoding Pipeline". In : *Fungal Ecology* 41, p. 23–33. doi : 10.1016/j.funeco.2019.03.005.

Pauvert C, Vallance J, Deliere L, Buee M et Vacher C (2019b). "Microbial Networks Inferred from Metabarcoding Data Lack Replicability : Consequences for next-Generation Biomonitoring". In : *bioRxiv*, p. 642199. doi : 10.1101/642199.

Pearson RC et Goheen AC, éds. (1988). *Compendium of Grape Diseases*. Disease Compendium Series of the American Phytopathological Society. St. Paul, Minn., USA : APS Press. 93 p. isbn : 978-0-89054-088-6.

Peay KG, Kennedy PG et Talbot JM (2016). "Dimensions of Biodiversity in the Earth Mycobiome". In : *Nature Reviews Microbiology* 14.7, p. 434–447. doi : 10.1038/nrmicro.2016.59.

Pecl GT, Araújo MB, Bell JD, Blanchard J, Bonebrake TC, Chen IC, Clark TD, Colwell RK, Danielsen F, Evengård B, Falconi L, Ferrier S, Frusher S, Garcia RA, Griffis RB, Hobday AJ, Janion-Scheepers C, Jarzyna MA, Jennings S, Lenoir J, Linnetved HI, Martin VY, McCormack PC, McDonald J, Mitchell NJ, Mustonen T, Pandolfi JM, Pettorelli N, Popova E, Robinson SA, Scheffers BR, Shaw JD, Sorte CJB, Strugnell JM, Sunday JM, Tuanmu MN, Vergés A, Villanueva C, Wernberg T, Wapstra E et Williams SE (2017). "Biodiversity Redistribution under Climate Change : Impacts on Ecosystems and Human Well-Being". In : *Science* 355.6332, eaai9214. doi : 10.1126/science.aai9214. pmid : 28360268.

Pedersen TL (2018). *Ggraph : An Implementation of Grammar of Graphics for Graphs and Networks*.

Pellissier L, Albouy C, Bascompte J, Farwig N, Graham C, Loreau M, Maglianesi MA, Melián CJ, Pitteloud C, Roslin T, Rohr R, Saavedra S, Thuiller W, Woodward G, Zimmermann NE et Gravel D (2018). "Comparing Species Interaction Networks along Environmental Gradients". In : *Biological Reviews* 93.2, p. 785–800. doi : 10.1111/brv.12366.

Pennington T, Reiff JM, Theiss K, Entling MH et Hoffmann C (2018). "Reduced Fungicide Applications Improve Insect Pest Control in Grapevine". In : *BioControl* 63.5, p. 687–695. doi : 10.1007/s10526-018-9896-2.

Perazzolli M, Moretto M, Fontana P, Ferrarini A, Velasco R, Moser C, Delledonne M et Pertot I (2012). "Downy Mildew Resistance Induced by Trichoderma Harzianum T39 in Susceptible Grapevines Partially Mimics Transcriptional Changes of Resistant Genotypes". In : *BMC Genomics* 13.1, p. 660. doi : 10.1186/1471-2164-13-660.

Pertot I, Giovannini O, Benanchi M, Caffi T, Rossi V et Mugnai L (2017). "Combining Biocontrol Agents with Different Mechanisms of Action in a Strategy to Control Botrytis

Cinerea on Grapevine". In : *Crop Protection*. Pesticide Use and Risk Reduction with IPM 97, p. 85–93. DOI : 10.1016/j.cropro.2017.01.010.

Pielou E (1966). "The Measurement of Diversity in Different Types of Biological Collections". In : *Journal of Theoretical Biology* 13, p. 131–144. DOI : 10.1016/0022-5193(66)90013-0.

Pilosof S, Porter MA, Pascual M et Kéfi S (2017). "The Multilayer Nature of Ecological Networks". In : *Nature Ecology & Evolution* 1.4, p. 0101. DOI : 10.1038/s41559-017-0101.

Pinto C, Pinho D, Sousa S, Pinheiro M, Egas C et Gomes AC (2014). "Unravelling the Diversity of Grapevine Microbiome". In : *PLoS One* 9.1, e85622.

Pocock MJO, Evans DM et Memmott J (2012). "The Robustness and Restoration of a Network of Ecological Networks". In : *Science* 335.6071, p. 973–977. DOI : 10.1126/science.1214915.

Poisot T, Canard E, Mouillot D, Mouquet N et Gravel D (2012). "The Dissimilarity of Species Interaction Networks". In : *Ecology Letters* 15.12. Sous la dir. de F Jordan, p. 1353–1361. DOI : 10.1111/ele.12002.

Pollock J, Glendinning L, Wisedchanwet T et Watson M (2018). "The Madness of Microbiome : Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies". In : *Applied and Environmental Microbiology* 84.7, e02627–17. DOI : 10.1128/AEM.02627-17. pmid : 29427429.

Polonelli L et Morace G (1986). "Reevaluation of the Yeast Killer Phenomenon." In : *Journal of Clinical Microbiology* 24.5, p. 866–869. pmid : 3771773.

Porter TM et Hajibabaei M (2018). "Scaling up : A Guide to High-Throughput Genomic Approaches for Biodiversity Analysis". In : *Molecular Ecology* 27.2, p. 313–338. DOI : 10.1111/mec.14478.

Poudel R, Jumpponen A, Schlatter DC, Paulitz TC, Gardener BBM, Kinkel LL et Garrett KA (2016). "Microbiome Networks : A Systems Framework for Identifying Candidate Microbial Assemblages for Disease Management". In : *Phytopathology* 106.10, p. 1083–1096. DOI : 10.1094/PHYTO-02-16-0058-FI.

Pradeu T, Kostyrka G et Dupré J (2016). "Understanding Viruses : Philosophical Investigations". In : *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences* 59, p. 57–63. DOI : 10.1016/j.shpsc.2016.02.008.

Props R, Kerckhof FM, Rubbens P, Vrieze JD, Sanabria EH, Waegeman W, Monsieurs P, Hammes F et Boon N (2017). "Absolute Quantification of Microbial Taxon Abundances". In : *The ISME Journal* 11.2, p. 584–587. DOI : 10.1038/ismej.2016.117.

Provost C et Pedneault K (2016). "The Organic Vineyard as a Balanced Ecosystem : Improved Organic Grape Management and Impacts on Wine Quality". In : *Scientia Horticulturae*. Recent Advances in Organic Horticulture Technology and Management - Part 1 208, p. 43–56. DOI : 10.1016/j.scienta.2016.04.024.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J et Glöckner FO (2012). "The SILVA Ribosomal RNA Gene Database Project : Improved Data Processing and Web-Based Tools". In : *Nucleic Acids Research* 41.D1, p. D590–D596. DOI : 10.1093/nar/gks1219.

Quinn TP, Erb I, Richardson MF et Crowley TM (2018). "Understanding Sequencing Data as Compositions : An Outlook and Review". In : *Bioinformatics*. DOI : 10.1093/bioinformatics/bty175.

R Core Team (2014). *R : A Language and Environment for Statistical Computing*. Vienna, Austria : R Foundation for Statistical Computing.

— (2018). *R : A Language and Environment for Statistical Computing*. Vienna, Austria : R Foundation for Statistical Computing.

Raimundo RLG, Guimarães PR et Evans DM (2018). "Adaptive Networks for Restoration Ecology". In : *Trends in Ecology & Evolution* 33.9, p. 664–675. doi : 10.1016/j.tree.2018.06.002.

Rastogi G, Tech JJ, Coaker GL et Leveau JH (2010). "A PCR-Based Toolbox for the Culture-Independent Quantification of Total Bacterial Abundances in Plant Environments". In : *Journal of Microbiological Methods* 83.2, p. 127–132. doi : 10.1016/j.mimet.2010.08.006.

Rebollar EA, Woodhams DC, LaBumbard B, Kielgast J et Harris RN (2017). "Prevalence and Pathogen Load Estimates for the Fungus Batrachochytrium Dendrobatidis Are Impacted by ITS DNA Copy Number Variation". In : *Diseases of Aquatic Organisms* 123.3, p. 213–226. doi : 10.3354/dao03097.

Remus-Emsermann MNP et Schlechter RO (2018). "Phyllosphere Microbiology : At the Interface between Microbial Individuals and the Plant Host". In : *New Phytologist* 218.4, p. 1327–1333. doi : 10.1111/nph.15054.

Rezki S, Campion C, Iacomi-Vasilescu B, Preveaux A, Toualbia Y, Bonneau S, Briand M, Laurent E, Hunault G, Simoneau P, Jacques MA et Barret M (2016). "Differences in Stability of Seed-Associated Microbial Assemblages in Response to Invasion by Phytopathogenic Microorganisms". In : *PeerJ* 4, e1923. doi : 10.7717/peerj.1923.

Ricciardi A, Blackburn TM, Carlton JT, Dick JTA, Hulme PE, Iacarella JC, Jeschke JM, Liebhold AM, Lockwood JL, MacIsaac HJ, Pyšek P, Richardson DM, Ruiz GM, Simberloff D, Sutherland WJ, Wardle DA et Aldridge DC (2017). "Invasion Science : A Horizon Scan of Emerging Challenges and Opportunities". In : *Trends in Ecology & Evolution* 32.6, p. 464–474. doi : 10.1016/j.tree.2017.03.007.

Richter M et Rosselló-Móra R (2009). "Shifting the Genomic Gold Standard for the Prokaryotic Species Definition". In : *Proceedings of the National Academy of Sciences of the United States of America* 106.45, p. 19126–19131. doi : 10.1073/pnas.0906412106. pmid : 19855009.

Ritpitakphong U, Falquet L, Vimoltust A, Berger A, Métraux JP et L'Haridon F (2016). "The Microbiome of the Leaf Surface of Arabidopsis Protects against a Fungal Pathogen". In : *New Phytologist* 210.3, p. 1033–1043. doi : 10.1111/nph.13808.

Rittenour WR, Park JH, Cox-Ganser JM, Beezhold DH et Green BJ (2012). "Comparison of DNA Extraction Methodologies Used for Assessing Fungal Diversity via ITS Sequencing". In : *Journal of environmental monitoring : JEM* 14.3, p. 766–774. doi : 10.1039/c2em10779a. pmid : 22230933.

Rivers AR, Weber KC, Gardner TG, Liu S et Armstrong SD (2018). "ITSxpress : Software to Rapidly Trim Internally Transcribed Spacer Sequences with Quality Scores for Marker Gene Analysis". In : *F1000Research* 7, p. 1418. doi : 10.12688/f1000research.15704.1.

Robert V, Vu D, Amor ABH, Wiele N van de, Brouwer C, Jabas B, Szoke S, Dridi A, Triki M, Daoud S ben, Chouchen O, Vaas L, Cock A de, Stalpers JA, Stalpers D, Verkley GJM, Groenewald M, Santos FB dos, Stegehuis G, Li W, Wu L, Zhang R, Ma J, Zhou M, Gorjón SP, Eurwilaichitr L, Ingsriswang S, Hansen K, Schoch C, Robbertse B, Irinyi L, Meyer W, Cardinali G, Hawksworth DL, Taylor JW et Crous PW (2013). "MycoBank Gearing up for New Horizons". In : *IMA Fungus* 4.2, p. 371. doi : 10.5598/imafungus.2013.04.02.16.

Robin G, Ambroise C et Robin S (2018). "Incomplete Graphical Model Inference via Latent Tree Aggregation". In : *Statistical Modelling*, p. 1471082X1878628. doi : 10.1177/1471082X18786289.

Rodriguez RJ, Jr JFW, Arnold AE et Redman RS (2009). "Fungal Endophytes : Diversity and Functional Roles". In : *New Phytologist* 182.2, p. 314–330. doi : 10.1111/j.1469-8137.2009.02773.x.

Rognes T, Flouri T, Nichols B, Quince C et Mahé F (2016). "VSEARCH : A Versatile Open Source Tool for Metagenomics". In : *PeerJ* 4, e2584. doi : 10.7717/peerj.2584.

Roossinck MJ (2019). "Viruses in the Phytobiome". In : *Current Opinion in Virology* 37, p. 72–76. doi : 10.1016/j.coviro.2019.06.008.

Rosenberg E et Zilber-Rosenberg I (2018). "The Hologenome Concept of Evolution after 10 Years". In : *Microbiome* 6.1, p. 78. doi : 10.1186/s40168-018-0457-9.

Röttjers L et Faust K (2018). "From Hairballs to Hypotheses–Biological Insights from Microbial Networks". In : *FEMS Microbiology Reviews* 42.6, p. 761–780. doi : 10.1093/femsre/fuy030.

Ruggiero MA, Gordon DP, Orrell TM, Bailly N, Bourgoin T, Brusca RC, Cavalier-Smith T, Guiry MD et Kirk PM (2015). "A Higher Level Classification of All Living Organisms". In : *PLOS ONE* 10.4, e0119248. doi : 10.1371/journal.pone.0119248.

Russell AL, Rebolleda-Gómez M, Shaible TM et Ashman T (2019). "Movers and Shakers : Bumble Bee Foraging Behavior Shapes the Dispersal of Microbes among and within Flowers". In : *Ecosphere* 10.5, e02714. doi : 10.1002/ecs2.2714.

Ryberg M (2015). "Molecular Operational Taxonomic Units as Approximations of Species in the Light of Evolutionary Models and Empirical Data from Fungi". In : *Molecular Ecology* 24.23, p. 5770–5777. doi : 10.1111/mec.13444.

Saiki R, Gelfand D, Stoffel S, Scharf S, Higuchi R, Horn G, Mullis K et Erlich H (1988). "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase". In : *Science* 239.4839, p. 487–491. doi : 10.1126/science.239.4839.487.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ et Walker AW (2014). "Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses". In : *BMC Biology* 12. doi : 10.1186/s12915-014-0087-z. pmid : 25387460.

Sanchez-Azofeifa A, Oki Y, Wilson Fernandes G, Ball RA et Gamon J (2012). "Relationships between Endophyte Diversity and Leaf Optical Properties". In : *Trees* 26.2, p. 291–299. doi : 10.1007/s00468-011-0591-5.

Sapkota R, Knorr K, Jørgensen LN, O'Hanlon KA et Nicolaisen M (2015). "Host Genotype Is an Important Determinant of the Cereal Phyllosphere Mycobiome". In : *New Phytologist* 207.4, p. 1134–1144. doi : 10.1111/nph.13418.

Sawant IS, Wadkar PN, Ghule SB, Rajguru YR, Salunkhe VP et Sawant SD (2017). "Enhanced Biological Control of Powdery Mildew in Vineyards by Integrating a Strain of Trichoderma Afroharzianum with Sulphur". In : *Biological Control* 114, p. 133–143. doi : 10.1016/j.biocontrol.2017.08.011.

Scheffers BR, Meester LD, Bridge TCL, Hoffmann AA, Pandolfi JM, Corlett RT, Butchart SHM, Pearce-Kelly P, Kovacs KM, Dudgeon D, Pacifici M, Rondinini C, Foden WB, Martin TG, Mora C, Bickford D et Watson JEM (2016). "The Broad Footprint of Climate Change from Genes to Biomes to People". In : *Science* 354.6313, aaf7671. doi : 10.1126/science.aaf7671. pmid : 27846577.

Schieber TA, Carpi L, Díaz-Guilera A, Pardalos PM, Masoller C et Ravetti MG (2017). "Quantification of Network Structural Dissimilarities". In : *Nature Communications* 8, ncomms13928. doi : 10.1038/ncomms13928.

Schielzeth H et Nakagawa S (2013). "Nested by Design : Model Fitting and Interpretation in a Mixed Model Era". In : *Methods in Ecology and Evolution* 4.1, p. 14–24. doi : 10.1111/j.2041-210x.2012.00251.x.

Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT et Quince C (2015). "Insight into Biases and Sequencing Errors for Amplicon Sequencing with the Illumina MiSeq Platform". In : *Nucleic Acids Research* 43.6, e37–e37. doi : 10.1093/nar/gku1341.

Schlatter DC, Burke I et Paulitz TC (2017). "Succession of Fungal and Oomycete Communities in Glyphosate-Killed Wheat Roots". In : *Phytopathology* 108.5, p. 582–594. doi : 10.1094/PHYTO-06-17-0212-R.

Schloss PD, Gevers D et Westcott SL (2011). "Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies". In : *PLoS ONE* 6.12. Sous la dir. de JA Gilbert, e27310. doi : 10.1371/journal.pone.0027310.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV et Weber CF (2009). "Introducing Mothur : Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities". In : *Applied and Environmental Microbiology* 75.23, p. 7537–7541. doi : 10.1128/AEM.01541-09. pmid : 19801464.

Schmid F, Moser G, Muller H et Berg G (2011). "Functional and Structural Microbial Diversity in Organic and Conventional Viticulture : Organic Farming Benefits Natural Biocontrol Agents". In : *Applied and Environmental Microbiology* 77.6, p. 2188–2191. doi : 10.1128/AEM.02187-10.

Schmidt PA, Bálint M, Greshake B, Bandow C, Römbke J et Schmitt I (2013). "Illumina Metabarcoding of a Soil Fungal Community". In : *Soil Biology and Biochemistry* 65, p. 128–132. doi : 10.1016/j.soilbio.2013.05.014.

Schoch CL et al. (2012). "Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi". In : *Proceedings of the National Academy of Sciences* 109.16, p. 6241–6246. doi : 10.1073/pnas.1117018109. pmid : 22454494.

Selosse MA (2017). *Jamais seul : Ces microbes qui construisent les plantes, les animaux et les civilisations*. OCLC : 1005741125. isbn : 978-2-330-07749-5.

Setati ME, Jacobson D, Andong UC et Bauer F (2012). "The Vineyard Yeast Microbiome, a Mixed Model Microbial Map". In : *PLoS ONE* 7.12. Sous la dir. de JA Gilbert, e52609. doi : 10.1371/journal.pone.0052609.

Seufert V et Ramankutty N (2017). "Many Shades of Gray—The Context-Dependent Performance of Organic Agriculture". In : *Science Advances* 3.3, e1602638. doi : 10.1126/sciadv.1602638.

Shade A, McManus PS et Handelsman J (2013). "Unexpected Diversity during Community Succession in the Apple Flower Microbiome". In : *mBio* 4.2, e00602–12. doi : 10.1128/mBio.00602-12. pmid : 23443006.

Shagin DA, Shagina IA, Zaretsky AR, Barsova EV, Kelmanson IV, Lukyanov S, Chudakov DM et Shugay M (2017). "A High-Throughput Assay for Quantitative Measurement of PCR Errors". In : *Scientific Reports* 7. doi : 10.1038/s41598-017-02727-8. pmid : 28578414.

Simonovici M (2019). *Enquête Pratiques Phytosanitaires En Viticulture En 2016 - Nombre de Traitements et Indicateurs de Fréquence de Traitement*. Ministère de l'Agriculture et de l'Alimentation. doi : http://agreste.agriculture.gouv.fr/spip.php?article15027.

Simpson EA (1949). "Measurement of Species Diversity". In : *Nature* 163, p. 688.

Simpson GL (2016). *Permute : Functions for Generating Restricted Permutations of Data*.

Sipiczki M (2016). "Overwintering of Vineyard Yeasts : Survival of Interacting Yeast Communities in Grapes Mummified on Vines". In : *Frontiers in Microbiology* 7, p. 212. doi : 10.3389/fmicb.2016.00212.

Soffe R, Altenhuber N, Bernach M, Remus-Emsermann MNP et Nock V (2019). "Comparison of Replica Leaf Surface Materials for Phyllosphere Microbiology". In : *PLOS ONE* 14.6, e0218102. doi : 10.1371/journal.pone.0218102.

Soliman SS, Trobacher CP, Tsao R, Greenwood JS et Raizada MN (2013). "A Fungal Endophyte Induces Transcription of Genes Encoding a Redundant Fungicide Pathway in Its Host Plant". In : *BMC Plant Biology* 13, p. 93. doi : 10.1186/1471-2229-13-93.

Somervuo P, Koskela S, Pennanen J, Henrik Nilsson R et Ovaskainen O (2016). "Unbiased Probabilistic Taxonomic Classification for DNA Barcoding". In : *Bioinformatics* 32.19, p. 2920–2927. doi : 10.1093/bioinformatics/btw346.

Somervuo P, Yu DW, Xu CC, Ji Y, Hultman J, Wirta H et Ovaskainen O (2017). "Quantifying Uncertainty of Taxonomic Placement in DNA Barcoding and Metabarcoding". In : *Methods in Ecology and Evolution* 8.4. Sous la dir. de D Warton, p. 398–407. doi : 10.1111/2041-210X.12721.

Sommeria-Klein G, Zinger L, Taberlet P, Coissac E et Chave J (2016). "Inferring Neutral Biodiversity Parameters Using Environmental DNA Data Sets". In : *Scientific Reports* 6. doi : 10.1038/srep35644. pmid : 27762295.

Song Z, Schlatter D, Gohl DM et Kinkel LL (2018). "Run-to-Run Sequencing Variation Can Introduce Taxon-Specific Bias in the Evaluation of Fungal Microbiomes". In : *Phytobiomes Journal* 2.3, p. 165–170. doi : 10.1094/PBIOMES-09-17-0041-R.

Stackebrandt E et Goebel BM (1994). "Taxonomic Note : A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology". In : *International Journal of Systematic and Evolutionary Microbiology,* 44.4, p. 846–849. doi : 10.1099/00207713-44-4-846.

Stanley CE et van der Heijden MGA (2017). "Microbiome-on-a-Chip : New Frontiers in Plant–Microbiota Research". In : *Trends in Microbiology* 25.8, p. 610–613. doi : 10.1016/j.tim.2017.05.001.

Stielow J, Lévesque C, Seifert K, Meyer W, Irinyi L, Smits D, Renfurm R, Verkley G, Groenewald M, Chaduli D, Lomascolo A, Welti S, Lesage-Meessen L, Favel A, Al-Hatmi A, Damm U, Yilmaz N, Houbraken J, Lombard L, Quaedvlieg W, Binder M, Vaas L, Vu D, Yurkov A, Begerow D, Roehl O, Guerreiro M, Fonseca A, Samerpitak K, van Diepeningen A, Dolatabadi S, Moreno L, Casaregola S, Mallet S, Jacques N, Roscini L, Egidi E, Bizet C, Garcia-Hermoso D, Martín M, Deng S, Groenewald J, Boekhout T, de Beer Z, Barnes I, Duong T, Wingfield M, de Hoog G, Crous P, Lewis C, Hambleton S, Moussa T, Al-Zahrani H, Almaghrabi O, Louis-Seize G, Assabgui R, McCormick W, Omer G, Dukik K, Cardinali G, Eberhardt U, de Vries M et Robert V (2015). "One Fungus, Which Genes ? Development and Assessment of Universal Primers for Potential Secondary Fungal DNA Barcodes". In : *Persoonia - Molecular Phylogeny and Evolution of Fungi* 35.1, p. 242–263. doi : 10.3767/003158515X689135.

Suda W, Nagasaki A et Shishido M (2009). "Powdery Mildew-Infection Changes Bacterial Community Composition in the Phyllosphere". In : *Microbes and Environments* 24.3, p. 217–223. doi : 10.1264/jsme2.ME09114.

Sun H, Ma H, Hao M, Pretorius IS et Chen S (2009). "Identification of Yeast Population Dynamics of Spontaneous Fermentation in Beijing Wine Region, China". In : *Annals of Microbiology* 59.1, p. 69. doi : 10.1007/BF03175601.

Swett CL, Bourret T et Gubler WD (2016). "Characterizing the Brown Spot Pathosystem in Late-Harvest Table Grapes (Vitis Vinifera L.) in the California Central Valley". In : *Plant Disease* 100.11, p. 2204–2210. doi : 10.1094/PDIS-11-15-1343-RE.

Taberlet P, Bonin A, Zinger L et Coissac E (2018). *Environmental DNA*. T. 1. Oxford University Press. isbn : 978-0-19-876722-0. doi : 10.1093/oso/9780198767220.001.0001.

Tackmann J, Matias Rodrigues JF et von Mering C (2019). "Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data". In : *Cell Systems* 9.3, 286–296.e8. doi : 10.1016/j.cels.2019.08.002.

Talley SM, Coley PD et Kursar TA (2002). "The Effects of Weather on Fungal Abundance and Richness among 25 Communities in the Intermountain West". In : *BMC Ecology* 2.1, p. 7. doi : 10.1186/1472-6785-2-7.

Tamaddoni-Nezhad A, Bohan D, Raybould A et Muggleton S (2015). "Towards Machine Learning of Predictive Models from Ecological Data". In : *Inductive Logic Programming*. Lecture Notes in Computer Science. Springer, Cham, p. 154–167. isbn : 978-3-319-23707-7 978-3-319-23708-4. doi : 10.1007/978-3-319-23708-4_11.

Tamaddoni-Nezhad A, Milani GA, Raybould A, Muggleton S et Bohan DA (2013). "Chapter Four - Construction and Validation of Food Webs Using Logic-Based Machine Learning and Text Mining". In : *Advances in Ecological Research*. Sous la dir. de G Woodward et DA Bohan. T. 49. Supplement C t. Ecological Networks in an Agricultural World. Academic Press, p. 225–289. doi : 10.1016/B978-0-12-420002-9.00004-4.

Taylor DL, Walters WA, Lennon NJ, Bochicchio J, Krohn A, Caporaso JG et Pennanen T (2016). "Accurate Estimation of Fungal Diversity and Abundance through Improved Lineage-Specific Primers Optimized for Illumina Amplicon Sequencing". In : *Applied and Environmental Microbiology* 82.24. Sous la dir. de D Cullen, p. 7217–7226. doi : 10.1128/AEM.02576-16.

Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson H, Hildebrand F, Bork P et Abarenkov K (2015). "Shotgun Metagenomes and Multiple Primer Pair-Barcode Combinations of Amplicons Reveal Biases in Metabarcoding Analyses of Fungi". In : *MycoKeys* 10, p. 1–43. doi : 10.3897/mycokeys.10.4852.

Tedersoo L et Lindahl B (2016). "Fungal Identification Biases in Microbiome Projects". In : *Environmental Microbiology Reports* 8.5, p. 774–779. doi : 10.1111/1758-2229.12438.

Tedersoo L, Sánchez-Ramírez S, Kõljalg U, Bahram M, Döring M, Schigel D, May T, Ryberg M et Abarenkov K (2018). "High-Level Classification of the Fungi and a Tool for Evolutionary Ecological Analyses". In : *Fungal Diversity* 90.1, p. 135–159. doi : 10.1007/s13225-018-0401-0.

Thompson LR et al. (2017). "A Communal Catalogue Reveals Earth's Multiscale Microbial Diversity". In : *Nature*, nature24621. doi : 10.1038/nature24621.

Thorsen J, Brejnrod A, Mortensen M, Rasmussen MA, Stokholm J, Al-Soud WA, Sørensen S, Bisgaard H et Waage J (2016). "Large-Scale Benchmarking Reveals False Discoveries and Count Transformation Sensitivity in 16S rRNA Gene Amplicon Data Analysis Methods Used in Microbiome Studies". In : *Microbiome* 4.1. doi : 10.1186/s40168-016-0208-8.

Tibshirani R (1996). "Regression Shrinkage and Selection Via the Lasso". In : *Journal of the Royal Statistical Society : Series B (Methodological)* 58.1, p. 267–288. doi : 10.1111/j.2517-6161.1996.tb02080.x.

Tilman D, Cassman KG, Matson PA, Naylor R et Polasky S (2002). "Agricultural Sustainability and Intensive Production Practices". In : *Nature* 418.6898, p. 671–677. doi : 10.1038/nature01014. pmid : 12167873.

Tipton L, Müller CL, Kurtz ZD, Huang L, Kleerup E, Morris A, Bonneau R et Ghedin E (2018). "Fungi Stabilize Connectivity in the Lung and Skin Microbial Ecosystems". In : *Microbiome* 6, p. 12. doi : 10.1186/s40168-017-0393-0.

Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, Fukuda S, Ushio M, Nakaoka S, Onoda Y, Yoshida K, Schlaeppi K, Bai Y, Sugiura R, Ichihashi Y, Minamisawa K et Kiers ET (2018). "Core Microbiomes for Sustainable Agroecosystems". In : *Nature Plants* 4.5, p. 247–257. doi : 10.1038/s41477-018-0139-4.

Toome M, Heinsoo K et Luik A (2010). "Relation between Leaf Rust (Melampsora Epitea) Severity and the Specific Leaf Area in Short Rotation Coppice Willows". In : *European Journal of Plant Pathology* 126.4, p. 583–588. doi : 10.1007/s10658-009-9566-4.

Tuck SL, Winqvist C, Mota F, Ahnström J, Turnbull LA et Bengtsson J (2014). "Land-use Intensity and the Effects of Organic Farming on Biodiversity : A Hierarchical Meta-analysis". In : *Journal of Applied Ecology* 51.3, p. 746–755. doi : 10.1111/1365-2664.12219.

Tylianakis JM, Laliberté E, Nielsen A et Bascompte J (2010). "Conservation of Species Interaction Networks". In : *Biological Conservation* 143.10, p. 2270–2279. doi : 10.1016/j.biocon.2009.12.004.

Tylianakis JM et Morris RJ (2017). "Ecological Networks Across Environmental Gradients". In : *Annual Review of Ecology, Evolution, and Systematics* 48.1, p. 25–48. doi : 10.1146/annurev-ecolsys-110316-022821.

Tylianakis JM, Tscharntke T et Lewis OT (2007). "Habitat Modification Alters the Structure of Tropical Host–Parasitoid Food Webs". In : *Nature* 445.7124, p. 202–205. doi : 10.1038/nature05429.

UNITE Community (2017). *UNITE General FASTA Release*. doi : 10.15156/BIO/587475.

Vacher C, Hampe A, Porté AJ, Sauer U, Compant S et Morris CE (2016a). "The Phyllosphere : Microbial Jungle at the Plant–Climate Interface". In : *Annual Review of Ecology, Evolution, and Systematics* 47.1. doi : 10.1146/annurev-ecolsys-121415-032238.

Vacher C, Tamaddoni-Nezhad A, Kamenova S, Peyrard N, Moalic Y, Sabbadin R, Schwaller L, Chiquet J, Smith MA, Vallance J, Fievet V, Jakuschkin B et Bohan DA (2016b). "Learning Ecological Networks from Next-Generation Sequencing Data". In : *Advances in Ecological Research*. T. 54, p. 1–39. isbn : 978-0-08-100978-9. doi : 10.1016/bs.aecr.2015.10.004.

Van der Heijden MGA, Bardgett RD et Straalen NMV (2008). "The Unseen Majority : Soil Microbes as Drivers of Plant Diversity and Productivity in Terrestrial Ecosystems". In : *Ecology Letters* 11.3, p. 296–310. doi : 10.1111/j.1461-0248.2007.01139.x.

Vandenkoornhuyse P, Quaiser A, Duhamel M, Van AL et Dufresne A (2015). "The Importance of the Microbiome of the Plant Holobiont". In : *New Phytologist* 206.4, p. 1196–1206. doi : 10.1111/nph.13312.

Vandeputte D, Kathagen G, D'hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, Wang J, Tito RY, De Commer L, Darzi Y, Vermeire S, Falony G et Raes J (2017). "Quantitative Microbiome Profiling Links Gut Community Variation to Microbial Load". In : *Nature* 551.7681, p. 507–511. doi : 10.1038/nature24460.

Vannier N, Agler M et Hacquard S (2019). "Microbiota-Mediated Disease Resistance in Plants". In : *PLOS Pathogens* 15.6. Sous la dir. de C Zipfel, e1007740. doi : 10.1371/journal.ppat.1007740.

Varanda CMR, Oliveira M, Materatski P, Landum M, Clara MIE et Félix MdR (2016). "Fungal Endophytic Communities Associated to the Phyllosphere of Grapevine Cultivars under Different Types of Management". In : *Fungal Biology* 120.12, p. 1525–1536. doi : 10.1016/j.funbio.2016.08.002.

Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC et Pati A (2015). "Microbial Species Delineation Using Whole Genome Sequences". In : *Nucleic Acids Research* 43.14, p. 6761–6771. doi : 10.1093/nar/gkv657. pmid : 26150420.

Vayssier-Taussat M, Albina E, Citti C, Cosson JF, Jacques MA, Lebrun MH, Le Loir Y, Ogliastro M, Petit MA, Roumagnac P et Candresse T (2014). "Shifting the Paradigm from Pathogens to Pathobiome : New Concepts in the Light of Meta-Omics". In : *Frontiers in Cellular and Infection Microbiology* 4, p. 29. doi : 10.3389/fcimb.2014.00029.

Vega NM et Gore J (2018). "Simple Organizing Principles in Microbial Communities". In : *Current Opinion in Microbiology*. Antimicrobials * Microbial Systems Biology 45, p. 195–202. doi : 10.1016/j.mib.2018.11.007.

Venturelli OS, Carr AC, Fisher G, Hsu RH, Lau R, Bowen BP, Hromada S, Northen T et Arkin AP (2018). "Deciphering Microbial Interactions in Synthetic Human Gut Microbiome Communities". In : *Molecular Systems Biology* 14.6, e8157. doi : 10.15252/msb.20178157.

Verny L, Sella N, Affeldt S, Singh PP et Isambert H (2017). "Learning Causal Networks with Latent Variables from Multivariate Information in Genomic Data". In : *PLOS Computational Biology* 13.10. Sous la dir. de J Listgarten, e1005662. doi : 10.1371/journal.pcbi.1005662.

Vionnet L, De Vrieze M, Dutartre A, Gfeller A, Lüthi A, L'Haridon F et Weisskopf L (2018). "Microbial Life in the Grapevine : What Can We Expect from the Leaf Microbiome ?" In : *OENO One* 52.3. doi : 10.20870/oeno-one.2018.52.3.2120.

Vogel C, Bodenhausen N, Gruissem W et Vorholt JA (2016). "The Arabidopsis Leaf Transcriptome Reveals Distinct but Also Overlapping Responses to Colonization by Phyllosphere Commensals and Pathogen Infection with Impact on Plant Health". In : *New Phytologist* 212.1, p. 192–207. doi : 10.1111/nph.14036.

Vorholt JA (2012). "Microbial Life in the Phyllosphere". In : *Nature Reviews Microbiology* 10.12, p. 828–840. doi : 10.1038/nrmicro2910.

Vorholt JA, Vogel C, Carlström CI et Müller DB (2017). "Establishing Causality : Opportunities of Synthetic Communities for Plant Microbiome Research". In : *Cell Host & Microbe* 22.2, p. 142–155. doi : 10.1016/j.chom.2017.07.004.

Vu D, Groenewald M, de Vries M, Gehrmann T, Stielow B, Eberhardt U, Al-Hatmi A, Groenewald JZ, Cardinali G, Houbraken J, Boekhout T, Crous PW, Robert V et Verkley GJM (2019). "Large-Scale Generation and Analysis of Filamentous Fungal DNA Barcodes Boosts Coverage for Kingdom Fungi and Reveals Thresholds for Fungal Species and Higher Taxon Delimitation". In : *Studies in Mycology* 92, p. 135–154. doi : 10.1016/j.simyco.2018.05.001.

Wang H, Wei Z, Mei L, Gu J, Yin S, Faust K, Raes J, Deng Y, Wang Y, Shen Q et Yin S (2017). "Combined Use of Network Inference Tools Identifies Ecologically Meaningful Bacterial Associations in a Paddy Soil". In : *Soil Biology and Biochemistry* 105, p. 227–235. doi : 10.1016/j.soilbio.2016.11.029.

Wang Q, Garrity GM, Tiedje JM et Cole JR (2007). "Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy". In : *Applied and Environmental Microbiology* 73.16, p. 5261–5267. doi : 10.1128/AEM.00062-07. pmid : 17586664.

Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC et Hui FK (2015a). "So Many Variables : Joint Modeling in Community Ecology". In : *Trends in Ecology & Evolution* 30.12, p. 766–779. doi : 10.1016/j.tree.2015.09.007.

Warton DI, Foster SD, De'ath G, Stoklosa J et Dunstan PK (2015b). "Model-Based Thinking for Community Ecology". In : *Plant Ecology* 216.5, p. 669–682. doi : 10.1007/s11258-014-0366-3.

Watts SC, Ritchie SC, Inouye M et Holt KE (2019). "FastSpar : Rapid and Scalable Correlation Estimation for Compositional Data". In : *Bioinformatics* 35.6, p. 1064–1066. doi : 10.1093/bioinformatics/bty734.

Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L, Alm EJ, Birmingham A, Cram JA, Fuhrman JA, Raes J, Sun F, Zhou J et Knight R (2016). "Correlation Detection Strategies in Microbial Data Sets Vary Widely in Sensitivity and Precision". In : *The ISME Journal* 10.7, p. 1669–1681. doi : 10.1038/ismej.2015.235.

Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vázquez-Baeza Y, Birmingham A, Hyde ER et Knight R (2017). "Normalization and Microbial Differential Abundance Strategies Depend upon Data Characteristics". In : *Microbiome* 5.1. doi : 10.1186/s40168-017-0237-y.

Westcott SL et Schloss PD (2015). "De Novo Clustering Methods Outperform Reference-Based Methods for Assigning 16S rRNA Gene Sequences to Operational Taxonomic Units". In : *PeerJ* 3, e1487. doi : 10.7717/peerj.1487.

— (2017). "OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units". In : *mSphere* 2.2, e00073–17. doi : 10.1128/mSphereDirect.00073-17.

White JR, Maddox C, White O, Angiuoli SV et Fricke WF (2013). "CloVR-ITS : Automated Internal Transcribed Spacer Amplicon Sequence Analysis Pipeline for the Characterization of Fungal Microbiota". In : *Microbiome* 1, p. 6. doi : 10.1186/2049-2618-1-6.

White T, Bruns T, Lee S et Taylor J (1990). "Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics". In : *PCR Protocols : A Guide to Methods and Applications*, p. 315–322.

Whittaker RH (1975). *Communities and Ecosystems*. 2d ed. New York : Macmillan. 385 p. isbn : 978-0-02-427390-1.

Wickham H (2007). "Reshaping Data with the Reshape Package". In : *Journal of Statistical Software* 21.12, p. 1–20.

— (2009). *Ggplot2 : Elegant Graphics for Data Analysis*. Springer New York. isbn : 978-0-387-98140-6.

— (2011). "The Split-Apply-Combine Strategy for Data Analysis". In : *Journal of Statistical Software* 40.1, p. 1–29.

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, Cordero OX, Brown SP, Momeni B, Shou W, Kettle H, Flint HJ, Haas AF, Laroche B, Kreft JU, Rainey PB, Freilich S, Schuster S, Milferstedt K, van der Meer JR, Grokopf T, Huisman J, Free A, Picioreanu C, Quince C, Klapper I, Labarthe S, Smets BF, Wang H, Isaac Newton Institute Fellows et Soyer OS (2016). "Challenges in Microbial Ecology : Building Predictive Understanding of Community Function and Dynamics". In : *The ISME Journal* 10.11, p. 2557–2568. doi : 10.1038/ismej.2016.45.

Wilke CO (2018). *Cowplot : Streamlined Plot Theme and Plot Annotations for 'Ggplot2'*.

Wurzbacher C, Larsson E, Bengtsson-Palme J, Van den Wyngaert S, Svantesson S, Kristiansson E, Kagami M et Nilsson RH (2018). "Introducing Ribosomal Tandem Repeat Barcoding for Fungi". In : *Molecular Ecology Resources* 19.1, p. 118–127. doi : 10.1111/1755-0998.12944.

Xiao Y, Angulo MT, Friedman J, Waldor MK, Weiss ST et Liu YY (2017). "Mapping the Ecological Networks of Microbial Communities". In : *Nature Communications* 8.1, p. 2042. doi : 10.1038/s41467-017-02090-2.

Yahr R, Schoch CL et Dentinger BTM (2016). "Scaling up Discovery of Hidden Diversity in Fungi : Impacts of Barcoding Approaches". In : *Phil. Trans. R. Soc. B* 371.1702, p. 20150336. doi : 10.1098/rstb.2015.0336. pmid : 27481788.

Yang RH, Su JH, Shang JJ, Wu YY, Li Y, Bao DP et Yao YJ (2018). "Evaluation of the Ribosomal DNA Internal Transcribed Spacer (ITS), Specifically ITS1 and ITS2, for the Analysis of Fungal Diversity by Deep Sequencing". In : *PLOS ONE* 13.10. Sous la dir. de D Cullen, e0206428. doi : 10.1371/journal.pone.0206428.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R et Rosselló-Móra R (2014). "Uniting the Classification of Cultured and Uncultured Bacteria and Archaea Using 16S rRNA Gene Sequences". In : *Nature Reviews Microbiology* 12.9, p. 635–645. doi : 10.1038/nrmicro3330.

Yoon G, Gaynanova I et Müller CL (2019). "Microbial Networks in SPRING - Semi-Parametric Rank-Based Correlation and Partial Correlation Estimation for Quantitative Microbiome Data". In : *Frontiers in Genetics* 10, p. 516. doi : 10.3389/fgene.2019.00516.

Zaneveld JR, McMinds R et Vega Thurber R (2017). “Stress and Stability : Applying the Anna Karenina Principle to Animal Microbiomes”. In : *Nature Microbiology* 2.9, p. 17121. doi : 10.1038/nmicrobiol.2017.121.

Zarraonaindia I, Owens SM, Weisenhorn P, West K, Hampton-Marcell J, Lax S, Bokulich NA, Mills DA, Martin G, Taghavi S, van der Lelie D et Gilbert JA (2015). “The Soil Microbiome Influences Grapevine-Associated Microbiota”. In : *mBio* 6.2, e02527–14. doi : 10.1128/mBio.02527-14.

Zengler K et Zaramela LS (2018). “The Social Network of Microorganisms — How Auxotrophies Shape Complex Communities”. In : *Nature Reviews Microbiology* 16.6, p. 383–390. doi : 10.1038/s41579-018-0004-5.

Zhang J, Kobert K, Flouri T et Stamatakis A (2014). “PEAR : A Fast and Accurate Illumina Paired-End reAd mergeR”. In : *Bioinformatics* 30.5, p. 614–620. doi : 10.1093/bioinformatics/btt593.

Zhang Y, Kastman EK, Guasto JS et Wolfe BE (2018). “Fungal Networks Shape Dynamics of Bacterial Dispersal and Community Assembly in Cheese Rind Microbiomes”. In : *Nature Communications* 9.1, p. 336. doi : 10.1038/s41467-017-02522-z.

Zhang Z, Kong X, Jin D, Yu H, Zhu X, Su X, Wang P, Zhang R, Jia M et Deng Y (2019). “Euonymus Japonicus Phyllosphere Microbiome Is Significantly Changed by Powdery Mildew”. In : *Archives of Microbiology* 201.8, p. 1099–1109. doi : 10.1007/s00203-019-01683-3.

Zilber-Rosenberg I et Rosenberg E (2008). “Role of Microorganisms in the Evolution of Animals and Plants : The Hologenome Theory of Evolution”. In : *FEMS Microbiology Reviews* 32.5, p. 723–735. doi : 10.1111/j.1574-6976.2008.00123.x.

Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, Barba MD, Dickie IA, Dumbrell AJ, Ficetola GF, Fierer N, Fumagalli L, Gilbert MTP, Jarman S, Jumpponen A, Kauserud H, Orlando L, Pansu J, Pawlowski J, Tedersoo L, Thomsen PF, Willerslev E et Taberlet P (2019). “DNA Metabarcoding—Need for Robust Experimental Designs to Draw Sound Ecological Conclusions”. In : *Molecular Ecology* 28.8, p. 1857–1862. doi : 10.1111/mec.15060.

Zott K, Claisse O, Lucas P, Coulon J, Lonvaud-Funel A et Masneuf-Pomarede I (2010). “Characterization of the Yeast Ecosystem in Grape Must and Wine Using Real-Time PCR”. In : *Food Microbiology* 27.5, p. 559–567. doi : 10.1016/j.fm.2010.01.006.

Zurell D, Pollock LJ et Thuiller W (2018). “Do Joint Species Distribution Models Reliably Detect Interspecific Interactions from Co-Occurrence Data in Homogenous Environments ?” In : *Ecography* 41.11, p. 1812–1819. doi : 10.1111/ecog.03315.

Zuur AF, Ieno EN, Walker N, Saveliev AA et Smith GM (2009). *Mixed Effects Models and Extensions in Ecology with R*. Statistics for Biology and Health. New York, NY : Springer New York. isbn : 978-0-387-87457-9 978-0-387-87458-6. doi : 10.1007/978-0-387-87458-6.

VII

Annexes

1 La biosurveillance pour le 21ème siècle : Intégrer le séquençage nouvelle-génération dans l'analyse des réseaux écologiques

Chapter One – Biomonitoring for the 21st Century : Integrating Next-Generation Sequencing Into Ecological Network Analysis

Stéphane A.P. Derocles[1*], David A. Bohan[1], Alex J. Dumbrell[2], James J.N. Kitson[3], François Massol[4], Charlie Pauvert[5], Manuel Plantegenest[6], Corinne Vacher[5], Darren M. Evans[3]

1– Agroécologie, AgroSup Dijon, INRA, University of Bourgogne Franche-Comté, Dijon, France. 2– School of Biological Sciences, University of Essex, Colchester, United Kingdom. 3– School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom. 4– CNRS, UMR 8198 Evo-Eco-Paleo, Université de Lille, SPICI group, Lille, France. 5– BIOGECO, INRA, Univ. Bordeaux, Pessac, France. 6– UMR 1349 IGEPP, INRA, Agrocampus-Ouest, Université de Rennes 1, Rennes Cedex, France.

* Correspondence : Dr. Stéphane A.P Derocles (stephane.derocles@inra.fr)

Ecological network analysis (ENA) provides a mechanistic framework for describing complex species interactions, quantifying ecosystem services, and examining the impacts of environmental change on ecosystems. In this chapter, we highlight the importance and potential of ENA in future biomonitoring programs, as current biomonitoring indicators (e.g. species richness, population abundances of targeted species) are mostly descriptive and unable to characterize the mechanisms that underpin ecosystem functioning. Measuring the robustness of multilayer networks in the long term is one way of integrating ecological metrics more generally into biomonitoring schemes to better measure biodiversity and ecosystem functioning. Ecological networks are nevertheless difficult and labour-intensive to construct using conventional approaches, especially when building multilayer networks in poorly studied ecosystems (i.e. many tropical regions). Next-generation sequencing (NGS) provides unprecedented opportunities to rapidly build highly resolved species interaction networks across multiple trophic levels, but are yet to be fully exploited. We highlight the impediments to ecologists wishing to build DNA-based ecological networks and discuss some possible solutions. Machine learning and better data sharing between ecologists represent very important areas for advances in NGS-based networks. The future of network ecology is very exciting as all the tools necessary to build highly resolved multilayer networks are now within ecologists reach.

Keywords : Ecological network, Food web, Next-generation sequencing, Molecular approach, Metabarcoding, Machine-learning, Ecosystem services, Environmental changes, Robustness, Biomonitoring

Available at https://doi.org/10.1016/bs.aecr.2017.12.001

2 Les effets maternels et l'environnement façonnent les communautés fongiques des graines chez les chênes

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license.

Maternal effects and environmental filtering shape seed fungal communities in oak trees

Tania Fort[1], Charlie Pauvert[1], Amy E. Zanne[2], Otso Ovaskainen[3,4], Thomas Caignard[1], Matthieu Barret[5], Stéphane Compant[6], Arndt Hampe[1], Sylvain Delzon[7] and Corinne Vacher[1]*

1. BIOGECO, INRA, Univ. Bordeaux, 33615 Pessac, France

2. George Washington University, Biological Sciences Department, 800 22nd St. , Washington DC 20052, USA

3. Organismal and Evolutionary Biology Research Programme, P.O. Box 65, 00014 University of Helsinki, Finland.

4. Center for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway

5. IRHS, Agrocampus-Ouest, INRA, Université d'Angers, SFR4207 QuaSaV, 49071 Beaucouzé, France.

6. AIT Austrian Institute of Technology GmbH, Center for Health & Bioresources, Bioresources Unit, Konrad-Lorenz Straße 24, 3430 Tulln, Austria

7. Univ. Bordeaux, INRA, BIOGECO, Pessac France.

*Author for correspondence: Corinne Vacher

Tel: +33(0)5 40 00 88 99

E-mail: corinne.vacher@inra.fr

Address: UMR BIOGECO, Université de Bordeaux, Allée Geoffroy St-Hilaire, Bât. B2, 33615 Pessac, France

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Summary

- Trees, as foundation species, play a pivotal role in the species interaction networks that constitute forest ecosystems. From the seed stage, they interact with microbial communities that affect their growth, health and fitness. Despite their eco-evolutionary importance, the processes shaping seed microbial communities in natural forests have received little attention.
- To unravel these processes, we analyzed the microbial communities of seeds collected in populations of sessile oak (*Quercus petraea*) growing along elevation gradients. We focused on the fungal communities as this group includes seed pathogens. Ecological processes shaping the communities were quantified using joint species distribution models.
- Fungi were present in all seed tissues, including the embryo. Fungal communities differed significantly among oak populations along the elevation gradients, and among mother trees within the same population. These maternal effects remained significant after seed fall, despite colonization by fungal species on the ground. Associations between tree pathogens and their antagonists were detected in the seeds.
- Our results demonstrate that both maternal effects and environmental filtering shape seed microbial communities of sessile oak. They provide a starting point for future research aimed at identifying the seed extended phenotypic traits that influence seed dispersal and germination, and seedling survival and growth across environments.

Keywords

Microbial ecology; disease ecology; community genetics; ecological networks; joint species distribution models; endophyte; host-parasite interaction; vertical transmission

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Introduction

Seeds are colonized by a wide range of microorganisms that positively and negatively influence plant growth, health and fitness. They associate with many endophytic and epiphytic microorganisms that foster the growth of seedlings and protect them against natural enemies (Baker & Smith, 1966; Nelson, 2004; Links, *et al.* 2014). For example, the seed-borne endophytic fungus *Epichloë festucae* limits pest attacks and pathogen development by producing alkaloids (Clay & Schardl, 2002; Faeth, 2002). Seeds are also colonized by pathogens that reduce germination rate (Nelson, 2017) and seedling survival (Kremer, 1987; Gilbert, 2002). For example, the fungal pathogen *Ciboria batschiana*, the causal agent of acorn black rot, can damage up to 80% of acorns when conditions are wet (Prochazkova *et al.*, 2005). By affecting the early stages of the plant life cycle, seed pathogens have significant impacts on natural ecosystems, including forests in which they shape tree species diversity and spatial structure (Janzen, 1970; Gilbert, 2002). Understanding how pathogens and other microorganisms are acquired and transmitted to the seeds of foundation tree species (i.e. tree species that structure and stabilize forest communities and ecosystem processes; Whitham *et al.*, 2006), and how these microorganisms associate to regulate seed germination and seedling development, is thus crucial to our ability to predict and manage the regeneration of forest ecosystems.

Up to now, processes of seed microbiota assembly have received more attention in crop plants than in natural ecosystems. In crops, microorganisms are transmitted by the mother plant to its seeds at the floral and during early seed development stages, through vascular tissues or contact between vegetative and reproductive organs (Maude, 1996). They can also be transmitted from the pollen of the father plant, insect vectors, or bioaerosols (Escobar Rodríguez *et al.*, 2018; Frank *et al.*, 2017). Once seeds fall on the ground, epiphytic microbial communities coalesce with microbial communities of litter and upper soil (Rillig *et al.*, 2015). Germinating seeds then release molecules that attract soil microbes, surrounding themselves with a microbiologically active soil area called the spermosphere (Nelson, 2004; Schiltz *et al.*, 2015). The emergence of the plant radicle creates cracks in the seed tegument, enabling microbes to colonize internal tissues (Nelson *et al.*, 2018). These events lead to intense biotic interactions among microorganisms (Nelson, 2004) and drastic changes in seed microbiota composition and function (Ofek *et al.*, 2011; Yang *et al.*, 2017; Torres-Cortés *et al.*, 2018). Recent studies suggest that seed colonization by soil microorganisms

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

represents the most influential microbial acquisition for seedling growth and health (Nelson *et al.*, 2018).

The seed microbiota is therefore a dynamic entity that is shaped, like all ecological communities, by four fundamental processes: selection by the abiotic environment (environmental filtering) and biotic interactions (biotic filtering), dispersal, ecological drift and evolutionary diversification (Vellend, 2010; Nemergut *et al.*, 2013; Ovaskainen *et al.*, 2017; Zhou & Ning, 2017). Unraveling these assembly processes is particularly important because they govern seed and seedling extended phenotypes (defined as the diversity and composition of associated communities; Whitham *et al.*, 2006), as well as the performance and eventually fitness of the plant host (Compant *et al.*, 2010; Truyens *et al.*, 2015; Brader *et al.*, 2017). Some processes are deterministic and depend on taxon-specific functional traits (Minard *et al.*, 2019), such as the response to selection. Other processes are partly or purely stochastic (Zhou & Ning, 2017; Rezki *et al.*, 2018) and generate divergences among communities occupying identical environments (Chase & Myers, 2011). The dispersal process is particular in the case of seed microbial communities because seeds are mobile. Microorganisms are recruited from a sequence of species pools: first from the microbiota of the mother tree's aboveground organs, and later from surrounding microbial sources such as bioaerosols, litter and soil (Barret *et al.*, 2016). The microorganisms that are directly transferred from the vascular system of mother plant to the seedlings through seeds are termed vertically-transmitted (Truyens *et al.*, 2015), while the others are called horizontally-acquired. Vertically-transmitted microorganisms can trigger maternal effects (defined as the causal influence of the maternal genotype or phenotype on the offspring phenotype; Wolf & Wade, 2009) in seed and seedling extended phenotypes (Vivas *et al.*, 2015).

To gain insight into processes of seed microbiome assembly in natural ecosystems, we analyzed the microbial turnover among acorns of sessile oak (*Quercus petraea*) populations growing along elevation gradients. We collected acorns in the canopy and on the ground beneath individual trees, and used Hierarchical Models of Species Communities (HMSC; Ovaskainen *et al.*, 2017) to quantify the ecological processes shaping seed microbial communities at each spatial level (oak population, mother tree and seed) and to generate hypotheses of interactions among microorganisms. We focused on fungal communities as they include pathogens that are detrimental to seed survival (such as *Ciboria batschiana*; Prochazkova *et al.*, 2005). After having described and visualized the fungal communities associated to acorn tissues, we tested the following

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

hypotheses: (**H1**) acorn fungal communities are shaped by environmental filtering and maternal effects, (**H2**) the maternal effects tend to disappear after acorn fall because of horizontal acquisition of fungi from the ground, (**H3**) the response of fungal species to variations in the acorn environment depend on their lifestyle, and (**H4**) biotic interactions among fungal species play a role in the protection of acorns against pathogens.

Materials and methods

Sampling design

Acorns were collected on October 21st and 22nd 2015 in four populations of sessile oak located in the Pyrenees Mountains (France). Two populations (Ade and Bager) were at an elevation of ~400 m a.s.l. and the other two (Gedre-Bas and Gabas) at ~1200 m a.s.l. The sampling date was chosen as close as possible to the fruiting peak at both elevations (see Caignard *et al.*, 2017 for a complete description of the sampling sites and fruiting phenology). In each site, acorns were collected from three trees randomly selected among the dominant adults. For each tree, we collected four acorns from the canopy using a slingshot and four acorns from the ground beneath the crown (within a distance of 2 m from the trunk). We also collected the biotic microenvironment of canopy acorns, defined as all tree tissues present in a cylinder of 4 cm diameter and 6 cm length around the acorn (including the acorn cupule, the cupules of other acorns, the twig to which the acorn was attached, the leaf petioles and the base of leaves) and the biotic microenvironment of ground acorns, defined as all substrates beneath the acorn within a cylinder of 4 cm diameter and 1 cm depth (including dead oak leaves, dead leaves of other plant species, acorn caps, twigs, pieces of bark, granules of soil, mosses, lichens or herbs). Each sample was collected aseptically, using new plastic gloves and scissors cleaned with 96% ethanol to minimize contamination. Samples were stored in individual plastic vials in a cooler with ice until they could be stored at -80°C.

Ten additional acorns were harvested from five mother trees (two in Gedre-Bas, two in Gabas and one in Ade). For each tree, one acorn was harvested in the canopy and one acorn was harvested on the ground. These acorns were surface-sterilized and dissected to characterize the endophytic fungal communities associated with acorn internal tissues. Surface-sterilization was completed using a three-step process: immersion for 3 min in a 70% ethanol solution, immersion for 2 min in

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

a 3% calcium hypochlorite solution and rinsing with DNAway and sterilized water. After drying the acorns on sterilized filter papers, the fruit walls and the embryos were collected using a sterilized nutcracker and pliers and stored at -80°C.

DNA extraction, amplification and sequencing of fungal communities

Samples were ground into a homogeneous powder using liquid nitrogen. Mortars and pestles were cleaned using DNAway and autoclaved for 20 min at 121°C between each sample. Approximately 45 mg of powder from each sample was transferred to a microplate under a laminar flow hood. Microplates were stored at -80°C until DNA extraction. Total DNA was extracted using DNeasy Plant Mini Kits (Qiagen, USA) according to the manufacturer's protocol except that DNA extracts were eluted twice with 50µL of elution buffer (10mM Tris-Cl, 0.5mM EDTA; pH 9.0).

The ITS1 region of the nuclear ribosomal internal transcribed spacer, considered the universal barcode marker for fungi (Schoch *et al.*, 2012), was amplified using the ITS1F (5′-CTTGGTCATTTAGAGGAAGTAA-3′; Gardes & Bruns, 1993) and ITS2 (5′-GCTGCGTTCTTCATCGATGC-3′; White, Bruns, Lee, & Taylor, 1990) primers. The reaction mixture (12.5 µL of final volume) consisted of 1.25 µL of template DNA, 2.5µL of 1µM of each of the forward and reverse primers and 6.25µL of 2X KAPA HiFi HotStart ReadyMix (Kapa Biosystems) containing 0.6 mM of each dNTP, 5 mM $MgCl_2$, and 2 units of Kapa Taq DNA polymerase. PCR cycling reactions were conducted on a Veriti 96-well Thermal Cycler (Applied Biosystems) using the following conditions: initial denaturation at 95 °C for 3 min followed by 20 cycles at 98°C for 20 s, 55°C for 45 s, 72°C for 15 s with final extension of 72 °C for 1 min. ITS1 amplification was confirmed by electrophoresis on a 2% agarose gel. Templates that were not successfully amplified using this protocol were amplified again after DNA dilution to 20ng/µL or 10ng/µL. Two marine fungal species (*Yamadazyma barbieri* and *Candida oceani*) were used as positive controls as they were unlikely to be found in our samples. One positive control included DNA of a strain of *C. oceani*, and the other included an equimolar mixture of the DNA of both species. A first negative control was represented by 1 mL of water washes of 4 empty plastic vials opened during the sampling campaign and washed with sterile water. The PCR mix was used as a second negative control.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

PCR products were diluted five times in PCR-grade water and used as a DNA template for a second PCR performed using the tailed primers ITS1F_PlaGe (5'-CTTTCCCTACACGACGCTCTTCCGATCTCTTGGTCATTTAGAGGAAGTAA-3') and ITS2_PlaGe (5'-GGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCGTTCTTCATCGATGC-3') designed by the Get-PlaGe sequencing facility (Toulouse, France). The second PCR was performed twice (once in 12.5 µL and once in 25µL of final volume) using the same reaction mixture as the first PCR. The PCR conditions were as follows: initial denaturation at 95 °C for 3 min followed by 10 cycles at 98°C for 20 s, 60°C for 15 s, 72°C for 15 s and a final extension step of 72°C for 1 min. PCR products were purified (CleanPCR, MokaScience), multiplex identifiers and sequencing adapters were added, and library sequencing on an Illumina MiSeq platform (v3 chemistry, 2x250 bp) and sequence demultiplexing (with exact index search) were performed at the Get-PlaGe sequencing facility (Toulouse, France).

Bioinformatic analyses

Paired-end sequences were joined using PEAR v0.9.10 (Zhang *et al.*, 2014). Only pairs with a minimum overlap of 50 bp and without any uncalled bases were kept. Assembled sequences were filtered using DADA2 v1.4.0 (Callahan *et al.*, 2016). Only sequences with less than one expected error and longer than 100 bp were retained in the dataset. Amplicon sequence variants (ASVs) were inferred using DADA2 and chimeric sequences were removed using the consensus method of the *removeBimeras* function. Taxonomic assignments were performed using the RDP classifier (Wang *et al.*, 2007) implemented in DADA2 and trained with the UNITE database 7.2 (UNITE Community 2017), with an 80% confidence threshold. The ASV table was then imported in R using the phyloseq package v1.26.0 (McMurdie & Holmes, 2013) and filtered. Only ASVs assigned to a fungal phylum were kept. Positive and negative controls were used to remove contaminants (Galan *et al.*, 2016). The cross-contamination threshold (T_{CC}) was defined as the maximal number of sequences of each ASV found in negative or positive control samples. The false-assignment threshold (T_{FA}) was defined as the highest sequence count of a positive control strain in a non-control sample, divided by the total number of sequences of the strain in the whole run and multiplied by the total number of sequences of each ASV. ASVs were removed from all samples where they harbored fewer sequences than either threshold (T_{FA} or T_{CC}).

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

SNP genotyping and maternity exclusion analyses

Acorns collected on the ground were genotyped to confirm that they belonged to the mother tree above them. Genotyping was performed using 39 polymorphic single nucleotide polymorphism (SNP) markers (Gerzabek *et al.*, 2017). DNA was diluted to a final concentration of 15 to 20 ng/µL and sequenced using the iPLEX Gold Genotyping kit (Agena, San Diego, CA, USA) at the Genome Transcriptome Facility of Bordeaux (PGTB, Bordeaux, France) according to the manufacturer's instructions (see Chancerel *et al.*, 2013 for more details). Two samples of aboveground tissues of each tree were genotyped and compared to estimate the typing error rate of false calls during genotyping. Loci with poor performance during the clustering procedure (call rates <60%) were excluded, resulting in a final set of 28 loci. Acorn genotypes were compared to the genotype of their putative mother tree. Considering the low error rate of these SNPs (Gerzabek *et al.*, 2017), we took a deliberately conservative approach and assumed that if tree and acorn shared no alleles for at least one locus, the mother-offspring relationship was not confirmed and the acorn sample was removed from the dataset.

Confocal microscopy

Sixteen additional acorns were collected in autumn 2017 from the ground of the oak forest of Bellebat (44°43'36.4"N 0°13'22.5"W, Southwest of France) to visualize fungal colonization outside and inside acorns. Acorns were cut in half with secators and fixed overnight at 4°C in a paraformaldehyde solution (4% w/v in PBS, pH 7.2). Samples were then rinsed three times with PBS, immersed in 15 mL PBS containing 50 µg ml^{-1} of wheat germ agglutinin (WGA)-AlexaFluor488 conjugate (Life Technologies, USA), incubated 2 hours at 37°C, and rinsed again three times with PBS. The samples were observed under a confocal microscope (Olympus Fluoview FV1000 with multiline laser FV5-LAMAR-2 and HeNe(G) laser FV10-LAHEG230-2). Pictures were taken with X, Y, Z coordinates at 405, 488, 594 nm and with 10X, 20X, 40X or 60X objectives. Images were merged (RGB) using Image J software (Schneider *et al.* 2012). Pictures were created using Z Project Stacks (Campisano *et al.* 2014), then cropped, and the light/contrast balance was improved (Glassner *et al.* 2015). Images presented in this publication correspond to the average colonization level observed.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Statistical analyses

Comparison of acorn fungal communities among mother trees and environments

To test hypotheses H1 and H2, we analyzed the effects of mother tree, local environment (oak population), and microenvironment (acorn position in the canopy *versus* on the ground) on acorn fungal community richness and composition. Fungal richness was defined as the total number of ASVs per acorn sample and was modelled using generalized linear mixed models (GLMMs) with a negative binomial distribution and a log-link function. The first model had oak population, acorn position and their interaction as fixed effects and the mother tree as a random effect. We then tested the effect of the mother tree on fungal richness for acorns in the canopy and acorns on the ground separately. The two models had the mother tree as fixed effect and the population as random effect. The natural logarithm of the total number of sequences per sample (sequencing depth) was introduced as an offset in all models.

We then analyzed the effects of the three same factors on fungal community composition by using permutational multivariate analyses of variance (PERMANOVAs) with 9999 permutations. Compositional dissimilarities among acorn samples were estimated using quantitative and binary versions of the Jaccard index (Jaccard, 1901) and visualized with principal coordinate analyses (PCoA). We first tested the effects of oak population, acorn position and their interaction on compositional dissimilarities among samples, by constraining or not constraining the permutations by elevation. We then assessed the effect of mother tree on community composition for acorns in the canopy and on the ground separately. The two models had the mother tree as a fixed effect and permutations were constrained by population. The natural logarithm of the total number of sequences per sample (sequencing depth) was introduced as the first effect in all models.

In addition, we investigated whether changes in fungal community composition after acorn fall were due to either the substitution of canopy-associated fungal species by ground-associated fungal species or gain of ground-associated fungal species without loss of canopy-associated fungal species, by partitioning Jaccard binary dissimilarities among acorns of the same mother tree using the betapart package v1.5.1 (Baselga *et al.,* 2018). The proportion of fungal species of acorns on the ground also found in acorns in the canopy was calculated for each mother tree.

Quantification of maternal effects, environmental filtering and biotic interactions

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Hierarchical Models of Species Communities (HMSC; Ovaskainen *et al.*, 2017) were then used to quantify processes of community assembly. The models assumed that variation in fungal community composition among acorns (i.e. the **Y** matrix in the HMSC framework) was accounted for by four ecological predictors (*Mother mycobiota*, *Microenv mycobiota*, *Acorn position* and *Site elevation*) introduced as fixed effects in the **X** matrix. The **Y** matrix represented ASV sequence counts in all acorn samples, out of which we included only the ASVs that were present in five or more acorns. *Mother mycobiota* and *Microenv mycobiota* represented fungal communities in the canopy of mothers trees and in the microenvironment of acorns, respectively. They were ASV-specific predictors and thus the **X** matrix was different for every ASV of the **Y** matrix. *Mother mycobiota* was calculated for all acorn samples as the average relative abundance of the focal ASV in the twigs and leaves of mother trees, and was included at the tree level to model vertical transmission of fungi from the mother tree to its acorns. *Microenv mycobiota* was calculated as the residuals of the regression of the relative abundance of the focal ASV in the microenvironment of each acorn over its relative abundance in the canopy of the mother tree, and was included at the sample level to model horizontal acquisition of fungi from the materials surrounding each acorn. *Site elevation* and *Acorn position* represented filtering of fungal communities by climate and microclimate, respectively. *Site elevation* was included at the site level to model selection exerted by site-level abiotic factors, such as average air temperature, on acorn fungal communities. *Acorn position* (canopy vs. ground) was included at the sample level to model selection exerted by microclimate, such as higher humidity on the ground, on acorn fungal communities. *Mother mycobiota* and *Microenv mycobiota* were included in interaction with *Acorn position* to test the hypothesis that their contribution to fungal communities differ between acorns in the canopy and acorns on the ground (H2). We also introduced the log-transformed sequencing depth of each sample (*Sequencing depth*) as a fixed effect in the **X** matrix, to take into account methodological biases influencing ASV sequence counts. Random effects at each hierarchical level (oak population, mother tree and acorn sample) were also introduced to model variations in ASV sequence counts that can neither be attributed to the four ecological predictors nor sequencing depth. In addition, we tested the hypothesis that fungal lifestyle modulates fungal ASV responses to environmental variations (H3), by including the trophic mode (saprotroph, plant pathogen or other) and the degree of specialization toward acorns of each ASV in the **T** matrix. The putative trophic mode of each ASV was determined using the FUNGuild database (Nguyen *et al.*, 2016).

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Their degree of specialization toward acorns was defined as the log-ratio of ASV relative abundance in acorns *versus* other sample types (i.e. branches, leaves, litter and upper soil), calculated using DESeq2 (Love *et al.*, 2014). To account for the zero-inflated nature of the data, we applied a hurdle modelling approach. We first fitted a probit model on ASV presence-absence data and then fitted a linear model on sequence count data conditional on presence, in which counts were log-transformed and scaled to zero mean and unit variance for each ASV and absences masked as non-available data. We fitted both models with default prior distributions (Ovaskainen *et al.*, 2017). For each of the four MCMC chains, we sampled the posterior for 1,500,000 iterations, out of which we excluded the first 500,000 as burn-in and thinned the remaining iterations by 1000, thus producing a total of 4,000 posterior samples. We examined MCMC convergence through the distributions of potential scale reduction factors (PSRF) of the model parameters. To examine model fit, we applied a two-fold cross validation across the samples and evaluated predictive performance by AUC for the presence-absence model and R^2 for the linear model. Finally, residual correlations among fungal ASV sequence counts at the acorn level were interpreted as hypothetical biotic interactions among fungal strains (see Ovaskainen *et al.*, 2017). We examined associations among ASVs assigned at the species level to test hypothesis H4.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Results

All acorn tissues, including the fruit wall, seed coat, and embryo were colonized by fungi, with a dense colonization under the endocarp (Fig. 1A-E). Ascomycota represented 91.1% and 89.4% of the sequences of canopy and ground acorns, respectively. *Dothideomycetes*, *Leotiomycetes* and *Sordariomycetes* were the three main classes of ascomycetes present (Fig. S1). Among the ten most abundant species associated to whole acorns, five are known as plant pathogens (*Gnomoniopsis paraclavulata, Taphrina carpini*, *Epicoccum nigrum*, *Mycosphaerella tassiana* and *Polyscytalum algarvense*) and two are known as antagonists of other microorganisms (*Cladosporium delicatulum, Cylindrium elongatum*) (Table 1; Tables S2 and S3 for all ASVs and subset of ASVs used in HMSC models, respectively). The ubiquitous fungi *Curvibasidium cygneicolum* and *Epicoccum nigrum* were dominant in the internal tissues of acorns, including the embryo (Table S1).

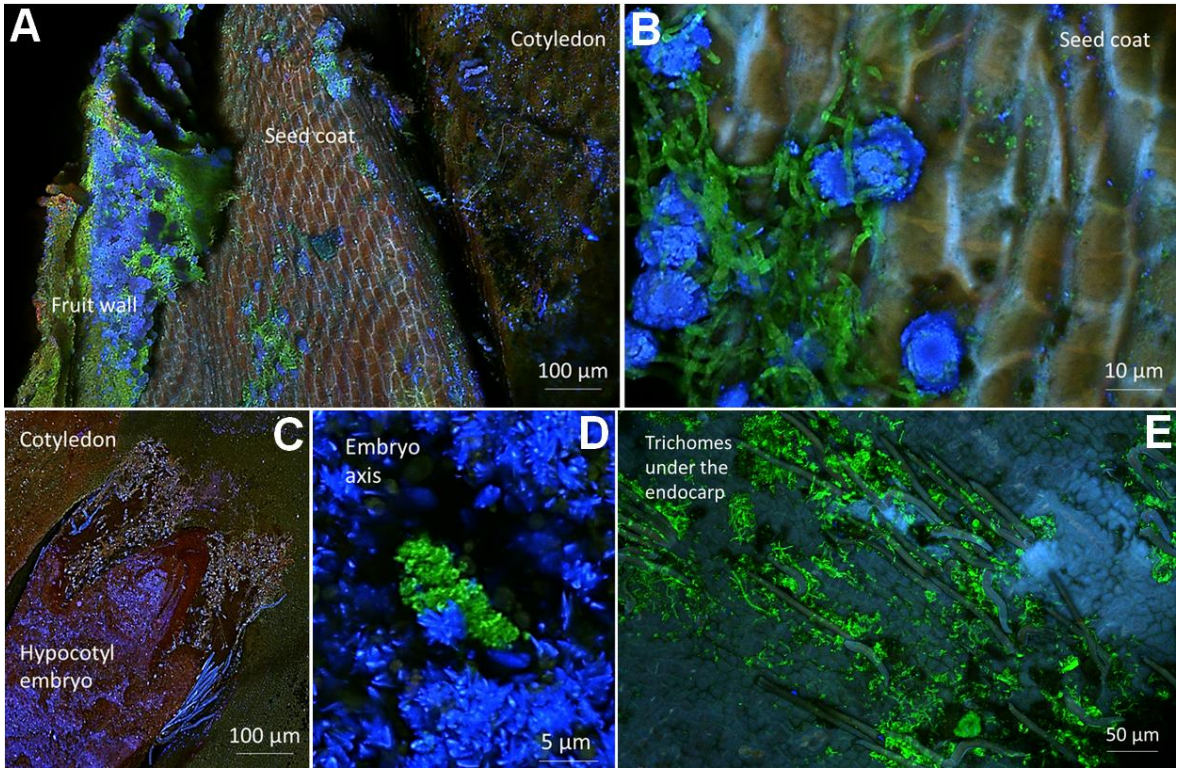


Figure 1. Fungi in cross sections of seeds of sessile oak collected on the ground. Fungi (green fluorescent) were revealed by confocal microscopy and WGA-ALEXA fluor488 staining. (A) Fruit wall and seed coat. (B) Zoom-in view of panel A. (C) Embryo and cotyledon. (D) Zoom-in view of panel C. (E) Internal surface of the endocarp.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table 1. Most abundant fungal species associated with seeds of sessile oak and their microenvironment. Seeds were either collected in the tree canopy or on the ground. Materials from the seed microenvironment (twigs and leaves, or litter and upper soil) were also collected. The fungal community of all 4 sample types was analyzed using a metabarcoding approach. Only Amplicon sequence variants (ASV) assigned to Ascomycota (A) or Basidiomycota (B) with the UNITE database were kept. Average relative abundances of all ASVs were computed for each sample type, after merging ASVs assigned to the same fungal species. Only ASVs identified at the species level are shown in the table.

Fungal species	**Average relative abundance per sample type (%)**					**Lifestyle**	**Reference(s)**
	All seeds	Seeds in the canopy	Canopy samples (twigs, leaves)	Seeds on the ground	Ground samples (litter, upper soil)		
Gnomoniopsis paraclavulata (A)	13.4	5.3	1.5	21.9	0	Pathogen isolated in leaves, buds, cupules and shoots of *Castenea sativa*. Commonly isolated from overwintered leaves of *Quercus* sp.	Tosi *et al.* (2014) ; Sogonov *et al.* (2008)
Stromatoseptoria castaneicola (A)	2.7	5.2	0.9	0	0.2	Causes leaf spots on *Castenea sativa*.	Quaedvlieg *et al.* (2013)
Taphrina carpini (A)	2.4	3.4	0.7	1.4	0.3	Common pathogen encountered on *Fagaceae* leaves.	Bacigálová (1991); Inácio *et al.* (2004); Cordier *et al.* (2012)
Epicoccum nigrum (A)	2.3	0.4	0.1	4.2	0.2	Ubiquitous fungus found in soil, leaves and seeds described as primary saprotroph and plant pathogen.	Ahumada-Rudolph *et al.* (2014)
Mycosphaerella tassiana (A)	1.7	1.5	0.1	1.9	0.3	Common pathogen found in the phyllosphere including that of oak.	Schubert *et al.* (2007); Jakuschkin *et al.* (2016)
Curvibasidium cygneicollum (B)	1.5	1.2	0.2	1.9	0.1	Endophyte of fruits, leaves, trunk and soil behaving as a phytopathogen or a mycoparasite. The species is insensitive to the mycocins produce by *Filobasidium* and *Cystofilobasidium*.	Sampaio *et al.* (2004); Mašínová *et al.* (2017) ; Sampaio *et al.* (2004)
Cylindrium elongatum (A)	1.5	1.4	0.1	1.5	4.3	Bacterial and fungal antagonist found on oak leaves.	Reyes-Estebanez (2011) ; Duarte *et al.* (2015)
Polyscytalum algarvense (A)	1.5	1.9	0.1	1	5.3	Necrotroph fungi found on *Eucalyptus* leaves.	Cheewangkoon et al. (2009); Crous (2018)
Fusarium pseudensiforme (A)	1.3	1.4	0	1.2	0	Found on bark of trees.	Nalim *et al.* (2011)
Cladosporium delicatulum (A)	1.3	1.3	0.2	1.3	0.3	Found in cereal seeds, mycoparasite of *Taphrina spp* and *Magnaporthe oryzae*	Amanelah Baharvandi & Zafari (2015) ; Chaibub *et al.* (2016)

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Acorn fungal communities are shaped by environmental filtering and maternal effects

Acorn fungal community richness (Table 2) and composition (Tables 3 and S2) differed significantly among oak populations along elevation gradients, and among mother trees within the same population, confirming the hypothesis that both environmental filtering and maternal effects shape acorn fungal communities (H1). PCoAs suggested that fungal communities were more similar between populations at the same elevation (Fig. 2). However, the population effect remained significant when PERMANOVAs were constrained by elevation (Pseudo-F = 1.24, P < 0.01), implying that selection exerted by abiotic conditions along the gradients was not the only process triggering variation in fungal community composition among oak populations.

Table 2. Generalized linear mixed-effects models (GLMM) of fungal community richness of seeds of sessile oak. Richness was defined as the number of amplicon sequence variants (ASV) per sample. The total number of sequences per sample (sequencing depth, SD) was introduced as an offset in all models. The effects of tree population (T), seed position (canopy *versus* ground, P) and their interaction were tested on the whole seed dataset while the effect of mother tree was tested separately on canopy seeds and ground seeds.

	d.f.	chi-square	*P*-value
All seeds			
Tree population (T)	3	12.5	**<.01**
Seed position (P)	1	12.6	**<.001**
T × P	3	5.74	0.12
Seeds in the canopy			
Mother Tree	10	30.6	**<.01**
Seeds on the ground			
Mother Tree	11	28.3	**<.01**

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table 3. Permutational multivariate analyses of variance (PERMANOVA) of compositional dissimilarities among fungal communities of seeds of sessile oak. Dissimilarities among seeds were estimated using the binary Jaccard distance. The total number of sequences per sample (sequencing depth, SD) was log-transformed and introduced as the first explanatory variable in all models. The effects of tree population (T), seed position (canopy *versus* ground, P) and their interaction were tested on the whole seed dataset while the effect of mother tree was tested separately on canopy seeds and ground seeds.

	d.f.	F-value	*P*-value	R^2
All seeds				
Log(SD)	1	1.3	**<.01**	0.01585
Tree population (T)	3	2.4	**<.001**	0.08711
Seed position (P)	1	1.8	**<.001**	0.02226
S x P	3	1.2	**<.001**	0.04494
Seeds in the canopy				
Log(SD)	1	1.3	0.191	0.03052
Mother tree	10	1.4	**0.010**	0.32742
Seeds on the ground				
Log(SD)	1	1.1	0.319	0.02834
Mother tree	11	1.2	**0.025**	0.34149

PCoAs and HMSC models gave different estimates of the relative contribution of environmental filtering and maternals effects. PCoAs suggested that mother tree identity had a lower influence on fungal community composition than elevation (Fig. 2), whereas HMSC models indicated the opposite trend (Table 4). The model of ASV presence-absence explained (in units of AUC, averaged over the ASVs) 80% of the variation for the model fitted to all data and 69% of the variation based on the two-fold cross-validation approach. According to this model, elevation was a minor direct driver of fungal community composition (only 2% of the explained variance). In contrast, the average relative abundance of a fungal ASV in the tissues of a mother tree (*Mother mycobiota*) was the second most important predictor of the occurrence of this ASV in an acorn from this tree, in interaction with the acorn position (39% of the explained variance). *Mother mycobiota* was the unique predictor of occurrence for several ASVs belonging to the orders *Helotiales*, *Venturiales* and *Xylariales* (Fig. S2). A similar ranking of the predictors was obtained for the model of ASV sequence counts (Table 4), except that sequencing depth and random effects had a much larger influence on ASV sequence counts than on ASV presence-absence (17% *versus*

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1% and 21% *versus* 6%, respectively). This model only explained 34% of the variation for the model fitted to all data and 18% of the variation based on the two-fold cross-validation approach.

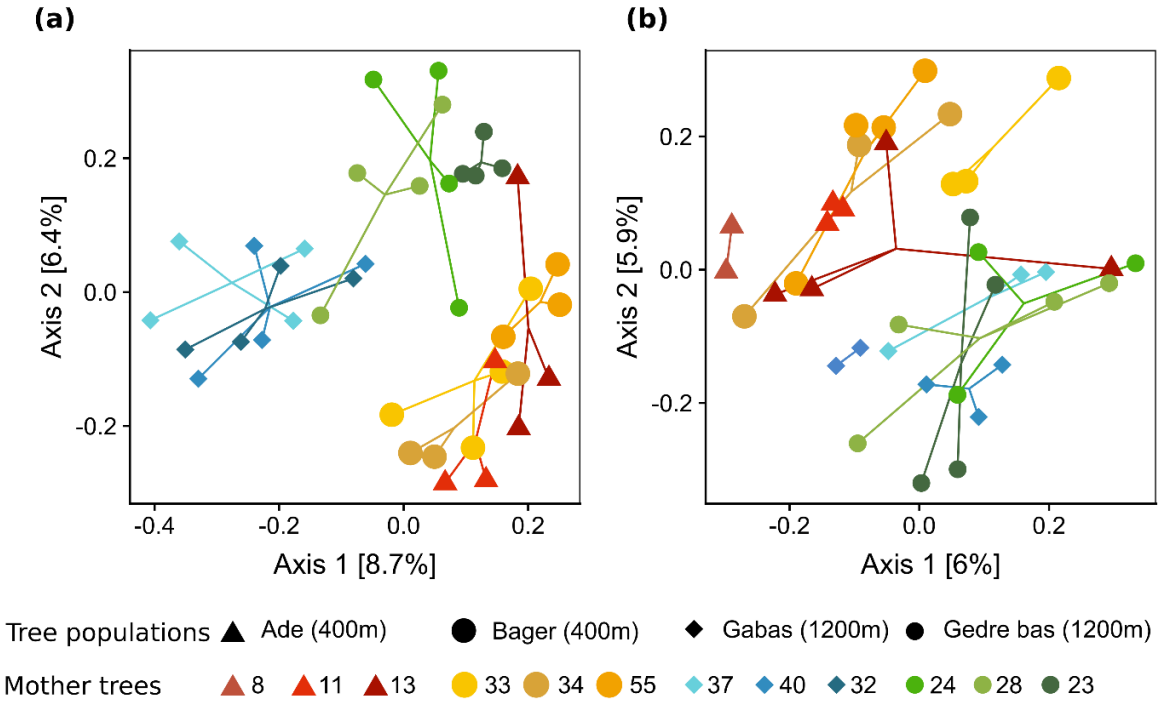


Figure 2. Compositional dissimilarities among fungal communities of seeds collected (A) in the canopy and (B) on the ground in four populations of sessile oak. Dissimilarities among seeds were estimated using binary Jaccard distance and represented with a PCoA plot. Fungal community composition differed significantly among tree populations and among mother trees (Table 3).

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table 4. Partitioning of the variance in fungal community composition of seeds of sessile oak. Four fixed effects were included in the HMSC models to explain variations in the presence-absence (PA) or the sequence count (SC) of a focal fungal ASV among seeds: *Sequencing depth* (total number of sequences per sample), *Seed position* (canopy or ground), *Microenv. mycobiota* (relative abundance of the focal ASV in the seed biotic microenvironment), *Mother mycobiota* (average relative abundance of the focal ASV in the mother tree aboveground tissues), and *Elevation*. Random effects were included at each spatial scale (seed, tree and population). Results of variance partitioning are given as percentages (%) of total explained variance.

Predictor			Explained variance (%)	
Type	**Level**	**Name**	**PA model**	**SC model**
Fixed	Seed	Sequencing depth	1	17
Fixed	Seed	Seed position (P)	4	8
Fixed	Seed	Microenv. mycobiota * P	47	23
Fixed	Tree	Mother mycobiota * P	39	25
Fixed	Population	Elevation	2	6
Random	Seed	Seed	3	9
Random	Tree	Tree	1	6
Random	Population	Population	2	6

Maternal effects persist after acorn fall despite horizontal acquisition of fungi from the ground

Fungal community richness (Table 2) and composition (Table 3 and S2) differed significantly between acorns in the canopy and acorns on the ground. For instance, *Gnomoniopsis paraclavulata* was four times more abundant in acorns on the ground than in acorns in the canopy, while *Stromatospheria castaneicola* was only present in canopy acorns (Table 1). Fungal richness increased and composition shifted toward that of ground materials after acorn fall (Fig. 3), confirming the horizontal acquisition of fungi from the ground. Partitioning of Jaccard beta-diversity indicated that these temporal changes in community composition were mainly driven by turnover (replacement of fungal species rather than net gains or losses in species number; Table S4). HMSC models confirmed the large influence of horizontal transmission on acorn fungal communities. The relative abundance of a fungal ASV in the microenvironment of an acorn (*Microenv mycobiota*), in interaction with the acorn position, was generally the best predictor of ASV occurrence (47% of the explained variance), especially for the *Capnodiales*, *Dothideales* and *Taphrinales* orders (Fig. S2).

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

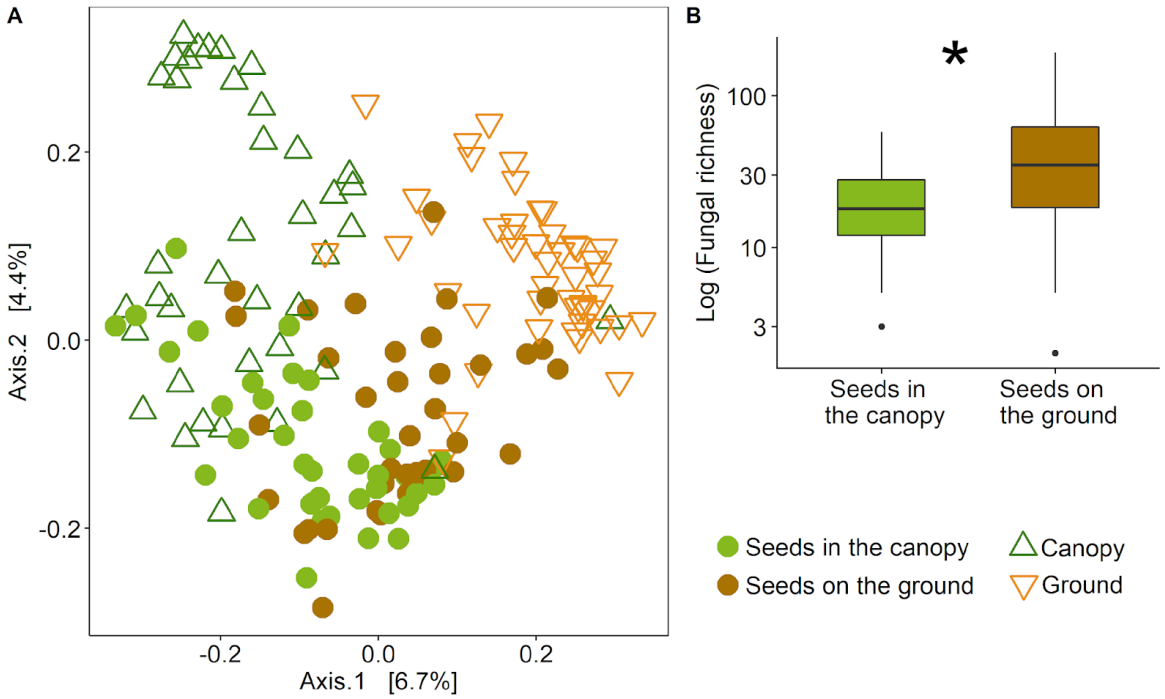


Figure 3. Fungal community composition and richness of seeds of sessile oak collected in the canopy and on the ground. (A) PCoA plot of compositional dissimilarities among fungal communities associated to seeds, canopy (leaves and twigs) and ground materials (litter and upper soil). Dissimilarities among samples were estimated using binary Jaccard distance. Fungal community composition differed significantly among the four sample types (PERMANOVA; $F=4.3$, $P<0.001$). (B) Richness (log-transformed) of seed fungal communities, defined as the number of ASVs per sample. Richness was significantly higher in seeds on the ground (Table 2).

Despite the turnover in fungal community composition after acorn fall and in contrast with our hypothesis (H2), the maternal effects were significant for acorns on the ground (Table 3), indicating that some maternal species were retained after acorn fall. Overall, acorns on the ground shared 10 to 40% of their fungal community with acorns in the canopy of the same mother tree, and 21 to 50% with mother tree tissues (Table S5). On average, 38% of fungal ASVs of acorns on the ground were present in both acorns in the canopy and mother tree tissues. Fungal species most often retained after acorn fall were *Taphrina* sp., *Cladosporium delicatulum*, a mycoparasite of *Taphrina* sp. (Baharvandi & Zafari, 2015), and the ubiquitous *Epicoccum nigrum*.

The response of fungal species to variations in the acorn environment depend on their lifestyle

HMSC models showed that fungal lifestyle influenced the response of acorn fungal communities to environmental variations, in accordance with our hypothesis (H3). High elevation selected for

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

saprotroph species and seed specialists, increasing their proportion in acorn fungal communities (Table S3). Acorn fall favored pathogen species and saprotrophs. Their proportion increased after acorn fall while their abundance (conditional on presence) was not altered (Table S3). These findings suggest that vertically-transmitted pathogens did not increase in abundance after acorn fall, and that acorns were colonized by pathogen and saprotroph species of the ground.

Biotic interactions among fungal species might play a role in the protection of acorns against pathogens

Fungal colonization was very dense on acorn external surfaces (Fig. 1A) but also on internal surface of the fruit wall (Fig. 1E), indicating that fungal colonizers might enter into contact and compete for space and eventually resources. In contrast to this expectation, residual co-occurrence patterns of HMSC models at the acorn level revealed only positive associations between fungal ASVs. Among fungal ASVs associated with each other, 14 could be assigned at the species level (Fig. 5). Six of them are described as plant pathogens and two of them, *Mycosphaerella tassiana* and *Taphrina carpini*, have already been found in association with oak (Table S6). Four of the species are described as mycoparasites, including *Cladosporium delicatulum*. The positive association between *Taphrina carpini* and *Cladosporium delicatulum* (Fig. 5), which are both found in the embryo and the fruit wall (Table S3), might therefore represent a vertically-transmitted host-parasite interaction. This interaction might play a role in oak disease regulation, in accordance with our hypothesis (H4).

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

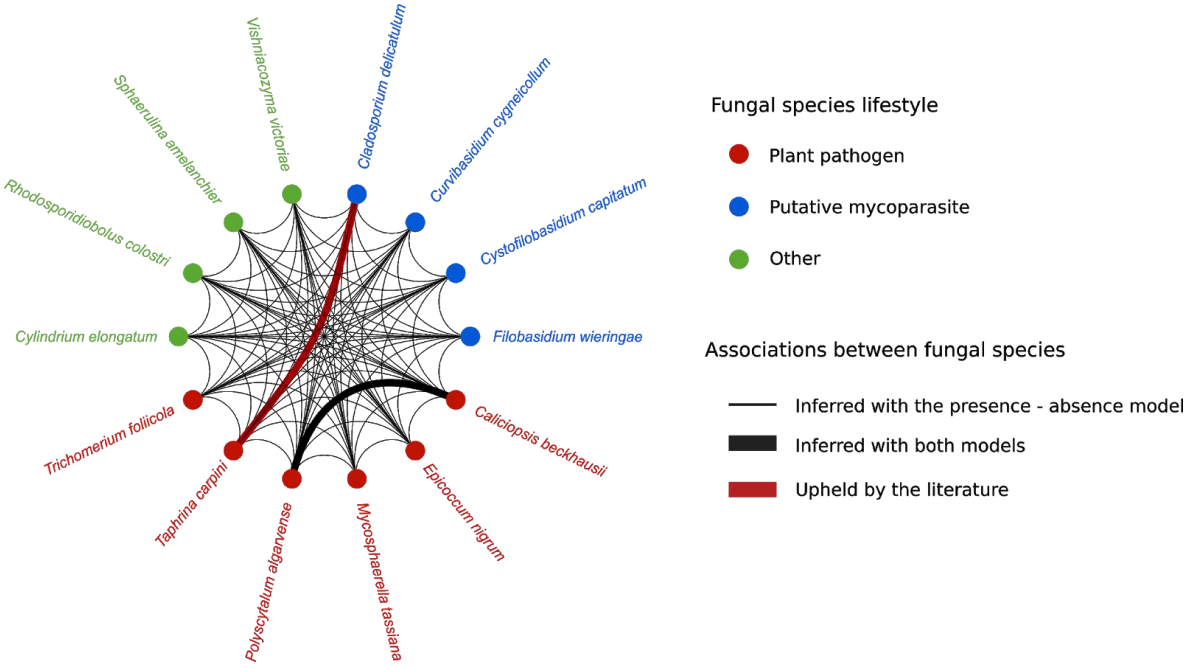


Figure 4. Network of fungal associations estimated by HMSC models. Associations were all positive. Network nodes correspond to fungal ASVs assigned to the species level and their color correspond to putative functions according to FUNGuild and the literature (Table S6). Network links indicate associations with at least 95% posterior probability estimated by the presence-absence model (thin black plain line) or by both the presence-absence model and the sequence count model (thick black plain line). The positive association between *Taphrina carpini* and *Cladosporium delicatulum* is indicated in red because it is described in the literature as a host-parasite interaction (Table S6).

Discussion

Seed microbial communities of trees have received little attention so far despite their potential influence on forest dynamics and evolution. In this study, we compared fungal communities of seeds among four populations of sessile oak (*Quercus petraea*), a dominant tree species of deciduous temperate European forests (McShea *et al.*, 2006). The populations were located along elevation gradients in the Pyrenees Mountains (France). They differed by up to 800 m in elevation. Our analyses revealed that the richness and composition of seed fungal communities differed significantly among oak populations along gradients, suggesting that abiotic filtering was a major driver of community assembly. However, Hierarchical Models of Species Communities (HMSC; Ovaskainen et al. 2017) revealed that the direct influence of elevation on seed fungal communities was small. Elevation accounted for only 2% of the explained variance in fungal species presence-absence, and 6% of the variance in species abundance. The apparently large effects of elevation

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

suggested by multivariate analyses might therefore be indirect effects. HMSC models estimated that fungal community composition in the microenvironment of each seed (i.e. twigs and leaves surrounding seeds in the canopy, or ground materials surrounding dropped seeds), explained 47% of the variance in species presence-absence, and was by far the strongest driver of seed fungal communities. These findings suggest that environmental filtering act directly on the fungal communities of leaves, litter and soil (as shown by Cordier et al., 2012; Coince et al., 2014; Vacher et al., 2016), to then indirectly shape seed communities.

Seed fungal communities also differed among mother trees within the same population and these differences remained significant after seed fall. Such maternal effects have already been found for the abundance of some fungal genera in seeds (Johnston-Monje & Raizada, 2011; Pinciroli *et al.*, 2013). Here we showed that maternal effects extend to the whole fungal community of seeds through to seed fall. HMSC indicated that the fungal community composition of mother tree tissues (twigs and leaves) had a major, direct influence on that of seeds. It accounted for 39% of the explained variance in fungal species presence-absence, and 25% of the variance in species abundance. Recent findings by Vivas *et al.*, (2017) on *Eucalyptus* trees indicate that these maternal effects can persist in seedlings and influence their growth and resistance to pathogens. Together, these results suggest that maternal effects in seed and seedling extended phenotypes could be a major driver of forest regeneration success.

In addition, our confocal microscopy analyses revealed, for the first time, the presence of fungal aggregates within embryos of acorns of sessile oak, as well as a dense fungal colonization on internal surfaces of fruit walls. These endophytic fungal populations contained foliar pathogens of *Fagaceae* tree species, such as *Mycosphaerella tassiana* and *Taphrina carpini* (Schubert *et al.*, 2007; Bacigálová, 1991), ubiquitous fungal species, such as *Epicoccum nigrum* (Andrews & Harris, 2000), and endophytic yeasts previously described in other fruits, such as *Curvibasidium cygneicollum* (Sampaio *et al.*, 2004; Mašínová *et al.*, 2017). Network inference analyses revealed a positive association between the foliar pathogen *Taphrina carpini* and the mycoparasite *Cladosporium delicatulum*, suggesting that mother trees do not only transmit pathogens but also pathogen antagonists. Our results hence confirm that fungal pathogens use seeds for their own dispersion, and that the fungal pathogen's parasites can follow them using the same dispersion mode (Ewald, 1989; Feldman *et al.*, 2008). Unraveling the genetic architecture of these tripartite

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

interactions, involving tree seedlings, seed-borne pathogens and their hyperparasites, could improve our understanding of forest ecosystem dynamics and evolution.

Finally, our analyses confirmed that seed fall corresponds to a major transition in seed fungal communities. Our results showed that fungal community richness significantly increased and that composition shifted toward that of ground materials after seed fall, confirming that seeds on the ground are rapidly colonized by the species present in the surrounding microenvironment (Crist, 2009, Qin *et al.,* 2016; Truyens *et al.,* 2015; Klaedtke *et al.,* 2016). For instance, the species *Gnomoniopsis paraclavulata*, that was previously found in association with oak litter (U'Ren *et al.,* 2016), drastically increased in abundance after acorn fall. Our analyses also suggested that seed fall triggers a replacement of canopy-inherited species by ground-derived species, rather than an addition of species associated to ground materials. This replacement was however only partial. Seeds on the ground shared up to 50% of their fungal colonizers with the twigs and leaves of their mother tree. The mechanisms of community filtration during vertical transmission (Vannier *et al.*, 2018) will have to be investigated in future studies.

Conclusions

Our study revealed that acorns of sessile oak harbor diversified fungal communities in their internal tissues, including the embryo, and on their surfaces. These communities were shaped by maternal effects, environmental filtering and biotic interactions. Maternal effects persisted after seed fall, despite seed colonization by soil and litter fungi. Environmental filtering did not shape directly seed fungal communities, but rather influenced communities in the microenvironments surrounding seeds. Biotic interactions included several host-parasite interactions between tree pathogens and their antagonists, one of which was likely to be vertically-transmitted. Future research will have to investigate the maternal and environmental drivers of the rate of vertical transmission of microorganisms (e.g. Cavazos *et al.,* 2018; Sneck *et al.,* 2017; Leff *et al.,* 2017), and assess the role of these microorganisms on seed survival and germination, seedling growth and health (e.g. Vivas *et al.,* 2017; Leroy *et al.,* 2019), and ultimately tree fitness. The influence of vertically-transmitted microorganisms on seed and seedling secondary metabolites (Chen *et al.,* 2018; Shazad *et al.,* 2018), and their cascading effects on tree biotic interactions (e.g. Peris *et al.,* 2018) will also have to be investigated. Previous research on oak trees showed for instance a

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

significant relationship between a fungus-like pathogen associated with acorns and the abundance of several oak-dependant bird species, including seed dispersers (Monahan & Koenig, 2006). A combination of germination experiments in controlled conditions (e.g. Leroy *et al.*, 2019), seed microbiome analyses in common gardens (Vivas *et al.*, 2015), and seed microorganisms manipulation across environments (Gundel *et al.*, 2017) will be required to integrate seed microbial ecology into predictive models of forest dynamics and evolution.

Acknowledgements

We thank Albin Lobo (University of Copenhagen, Denmark) and Jean-Marc Louvet (Biogeco, INRA-University of Bordeaux, France) for their assistance during sampling campaigns. We are also very grateful to Martine-Martin Clotté, Céline Lalanne, Patrick Leger, Adline Delcamp (Biogeco, INRA-University of Bordeaux, France) and Jessica Vallance (SAVE, INRA-Bordeaux Sciences Agro, France) for providing materials and reagents for processing the samples and advice on molecular biology protocols. We thank Samuel Pennarun and all members of the Genotoul sequencing facility (Get-PlaGe, Toulouse, France) for performing MiSeq sequencing, and the Genotoul bioinformatics facility (Bioinfo Genotoul, Toulouse, France) for providing computing and storage resources. We also thank Erwan Guichoux and all members of the Genome Transcriptome facility of Bordeaux (PGTB, Cestas, France) for performing the SNP genotyping. Thanks to Franck Deniel and Gaëtan Burgaud (ESIAB, Université de Bretagne Occidentale, France) for giving us marine fungal strains used as positive controls. We would like to acknowledge George Washington University for its contribution to the funding of molecular biology analyses. We also thank the INRA Meta-Omics of Microbial Ecosystems (MEM) metaprogramme (Learn-biocontrol project), the INRA Ecosystem Services (EcoServ) metaprogramme (IBISC project), the LABEX COTE (MICROMIC project; ANR-10-LABX-45), the LABEX CEBA (ANR-10-LABX-25-01) and the Aquitaine Region (Athene project, n°2016-1R20301-00007218) for additional financial support throughout the project. TF's PhD grant was funded by the University of Bordeaux. OO was supported by the Academy of Finland (grant 309581), the Erkko foundation, and the Research Council of Norway (SFF-III grant 223257).

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Author contributions

TF participated in the sampling campaigns, processed all samples, extracted and amplified DNA, performed the statistical analyses, interpreted the results and wrote the first draft of the manuscript. CP helped process the samples, performed the bioinformatic analyses and provided statistical analysis scripts. AEZ assisted with supervision and writing. OO conducted the HMSC analysis and helped interpreting the results. TC organized the sampling campaigns and provided environmental data. MB provided advice on molecular biology protocols, bioinformatics and statistical analyses. SC took the microscopy pictures and interpreted the results. AH provided advice and analysis tools for SNP genotyping. SD, AEZ and CV had the original idea for the study. CV participated in the sampling campaigns, coordinated and supervised all stages of the work, and made a major contribution to the writing of the manuscript. All authors revised the manuscript and approved the final version.

Data availability

All raw sequences obtained from the sequencing of acorns and their biotic microenvironment are available from the National Center for Biotechnology Information Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the accession number PRJNA551388. The code and ASV tables are available as an archive at https://doi.org/10.15454/SM6OCR.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

References

Ahumada-Rudolph R, Cajas-Madriaga D, Rudolph A, Reinoso R, Torres C, Silva M, Becerra J. 2014. Variation of sterols and fatty acids as an adaptive response to changes in temperature, salinity and pH of a marine fungus *Epicoccum nigrum* isolated from the Patagonian Fjords. *Revista de biología marina y oceanografía* **49**: 293–305.

Amanelah Baharvandi H, Zafari D. 2015. Identification of *Cladosporium delicatulum* as a mycoparasite of *Taphrina pruni*. *Archives of Phytopathology and Plant Protection* **48**: 688–697.

Andrews JH, Harris RF. 2000. The ecology and biogeography of microorganisms on plant surfaces. *Annual Review of Phytopathology* **38**: 145–180.

Bacigálová K. 1991. New localities of *Taphrina carpini* (Rostr.) Johans, on *Carpinus betulus* in Slovakia. *Czech Mycology* **46**: 296–302.

Baker KF, Smith SH. 1966. Dynamics of seed transmission of plant pathogens. *Annual Review of Phytopathology* **4**: 311 – 332.

Barret M, Guimbaud JF, Darrasse A, Jacques MA. 2016. Plant microbiota affects seed transmission of phytopathogenic microorganisms. *Molecular plant pathology* **17**: 791–795.

Baselga A, Orme CDL. 2012. Betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution* **3**: 808–812.

Brader G, Compant S, Vescio K, Mitter B, Trognitz F, Ma L-J, Sessitsch A. 2017. Ecology and genomic insights into plant-pathogenic and plant-nonpathogenic endophytes. *Annual Review of Phytopathology* **55**: 61–83.

Caignard T, Kremer A, Firmat C, Nicolas M, Venner S, Delzon S. 2017. Increasing spring temperatures favor oak seed production in temperate areas. *Scientific Reports* **7**: 1–8.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**: 581–583.

Campisano A, Ometto L, Compant S, Pancher M, Antonielli L, Yousaf S, Varotto C, Anfora G, Pertot I, Sessitsch A, et al. 2014. Interkingdom transfer of the acne-causing agent, propionibacterium acnes, from human to grapevine. *Molecular Biology and Evolution* **31**: 1059–1065.

Cavazos BR, Bohner TF, Donald ML, Sneck ME, Shadow A, Omacini M, Rudgers JA, Miller TEX. 2018. Testing the roles of vertical transmission and drought stress in the prevalence of heritable fungal endophytes in annual grass populations. *New Phytologist*: 1075–1084.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Chaibub AA, de Carvalho JCB, de Sousa Silva C, Collevatti RG, Gonçalves FJ, de Carvalho Barros Côrtes MV, de Filippi MCC, de Faria FP, Lopes DCB, de Araújo LG. **2016**. Defence responses in rice plants in prior and simultaneous applications of *Cladosporium* sp. during leaf blast suppression. *Environmental Science and Pollution Research* **23**: 21554–21564.

Chancerel E, Lamy JB, Lesur I, Noirot C, Klopp C, Ehrenmann F, Boury C, Provost G Le, Label P, Lalanne C, et al. 2013. High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biology* **11**: 50.

Chase JM, Myers JA. **2011**. Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society* **366**: 2351–2363.

Cheewangkoon R, Groenewald JZ, Summerell BA, Hyde KD, To-Anun C, Crous PW. **2009**. Myrtaceae, a cache of fungal biodiversity. *Persoonia: Molecular Phylogeny and Evolution of Fungi* **23**: 55–85.

Chen H, Wu H, Yan B, Zhao H, Liu F, Zhang H, Sheng Q, Miao F, Liang Z. **2018**. Core microbiome of medicinal plant *Salvia miltiorrhiza* seed : A rich reservoir of beneficial microbes for secondary metabolism? *Molecular Sciences* **19**: 672–687.

Chomnunti P, Bhat DJ, Jones EBG. 2012. Trichomeriaceae, a new sooty mould family of Chaetothyriales. *Fungal Diversity* **56**:63-76

Clay K, Schardl C. **2002**. Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. *The American Naturalist* **160**: S99–S127.

Compant S, Van Der Heijden MGA, Sessitsch A. **2010**. Climate change effects on beneficial plant-microorganism interactions. *FEMS Microbiology Ecology* **73**: 197–214.

Coince A, Cordier T, Lengellé J, Defossez E, Vacher C, Robin C, Buée M, Marçais B. **2014**. Leaf and root-associated fungal assemblages do not follow similar elevational diversity patterns. *PLoS ONE* **9**.

Cordier T, Robin C, Capdevielle X, Fabreguettes O, Desprez-Loustau ML, Vacher C. **2012**. The composition of phyllosphere fungal assemblages of European beech (*Fagus sylvatica*) varies significantly along an elevation gradient. *New Phytologist* **196**: 510–519.

Crist T, Friese F. **2009**. The Impact of Fungi on Soil Seeds : Implications for plants and granivores in a semiarid shrub-steppe. *Ecological Society of America* **74**: 2231–2239.

Crous PW, Schumacher RK, Wingfield MJ, Akulov A, Denman S, Roux J, Braun U, Burgess

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

TI, Carnegie AJ, Váczy KZ, et al. 2018. New and Interesting Fungi. *Fungal Systematics and Evolution* **1**: 169–215.

Desjardins AE. **2003**. Gibberella from A(venaceae) to Z(eae). *Annual Review of Phytopathology* **41**: 177–198.

Douglas Boyette C, Lynn Walker H. **1986**. Evaluation of *Fusarium lateritium* as a Biological Herbicide for Controlling Velvetleaf (Abutilon theophrasti) and Prickly Sida (Sida spinosa). *Weed Science Society of America* **34**: 106–109.

Duarte S, Bärlocher F, Trabulo J, Cássio F. **2015**. Stream-dwelling fungal decomposer communities along a gradient of eutrophication unraveled by 454 pyrosequencing. *Fungal Diversity* **70**: 127–148.

Escobar Rodriguez C, Mitter B, Barret M, Sessitsch A, Compant S. **2018**. Commentary: seed bacterial inhabitants and their routes of colonization. *Plant and Soil* **422**: 129–134.

Ewald PW. **1989**. Transmission modes and evolution of the parasitism-mutualism continuum. *Annals of the New York Academy of Sciences* **503**: 295–306.

Faeth SH. **2002**. Fungal Endophytes: Common Host Plant Symbionts but Uncommon Mutualists. *Integrative and Comparative Biology* **42**: 360–368.

Feldman TS, Brien HEO, Arnold AE. **2008**. Moths that vector a plant pathogen also transport endophytic fungi and mycoparasitic antagonists. *Microbial Ecology* **56**: 742–750.

Frank A, Saldierna Guzmán J, Shay J. **2017**. Transmission of bacterial endophytes. *Microorganisms* **5**: 70.

Galan M, Razzauti M, Bard E, Bernard M, Brouat C, Charbonnel N, Dehne-Garcia A, Loiseau A, Tatard C, Tamisier L, et al. 2016. 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife. *mSystems* **1**: e00032-16.

Gardes M, Bruns TD. **1993**. ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology* **2**: 113–118.

Garrido-benavent I, Pérez-ortega S. **2015**. Unravelling the diversity of European Caliciopsis (Coryneliaceae, Ascomycota): *Caliciopsis valentina* sp. nov. and *C. beckhausii* comb. nov., with a worldwide key to Caliciopsis. *Mycological Progress* **14**: 1–11.

Gerzabek G, Oddou-Muratorio S, Hampe A. **2017**. Temporal change and determinants of maternal reproductive success in an expanding oak forest stand. *Journal of Ecology* **105**: 39–48.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Gilbert GS. **2002**. Evolutionary ecology of plant diseases in natural ecosystems. *Annual Review of Phytopathology* **40**: 13–43.

Glassner H, Zchori-Fein E, Compant S, Sessitsch A, Katzir N, Portnoy V, Yaron S. **2015**. Characterization of endophytic bacteria from cucurbit fruits with potential benefits to agriculture in melons (Cucumis melo L.). *FEMS Microbiology Ecology* **91**: 1–13.

Glushakova AM, Kachalkin A V. 2017. Endophytic yeasts in *Malus domestica* and *Pyrus communis* fruits under anthropogenic impact. *Microbiology* **86**: 128–135.

Golubev WI, Tomashevskaya MA. **2009**. Characterization of Mycocin Secreted by *Rhodotorula colostri* (Castelli) Lodder. *Biology bulletin* **36**: 373–378.

Gundel PE, Rudgers JA, Whitney KD. 2017. Vertically transmitted symbionts as mechanisms of transgenerational effects. *American Journal of Botany* 104: 787–792.

Jaccard P. **1901**. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* **37**: 547–579.

Inácio J, Rodrigues MG, Sobral P, Fonseca Á. **2004**. Characterisation and classification of phylloplane yeasts from Portugal related to the genus *Taphrina* and description of five novel *Lalaria* species. *FEMS Yeast Research* **4**: 541–555.

Isaeva O V., Glushakova AM, Yurkov AM, Chernov IY. **2009**. The yeast *Candida railenensis* in the fruits of English oak (Quercus robur L.). *Microbiology* **78**: 355–359.

Jakuschkin B, Fievet V, Schwaller L, Fort T, Robin C, Vacher C. **2016**. Deciphering the pathobiome: Intra- and interkingdom interactions involving the pathogen *Erysiphe alphitoides*. *Microbial Ecology* **72**: 870–880.

Janzen DH. **1970**. Herbivores and the Number of Tree Species in Tropical Forests. *The American Naturalist* **104**: 501–528.

Johnston-Monje D, Raizada MN. **2011**. Conservation and diversity of seed associated endophytes in Zea across boundaries of evolution, ethnography and ecology. *PLoS ONE* **6**:e20396.

Klaedtke S, Jacques MA, Raggi L, Préveaux A, Bonneau S, Negri V, Chable V, Barret M. **2016**. Terroir is a key driver of seed-associated microbial assemblages. *Environmental microbiology* **18**: 1792–1804.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Leff JW, Lynch RC, Kane NC, Fierer N. 2017. Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower, *Helianthus annuus*. *New Phytologist* **214**: 412–423.

Leroy C, Maes AQ, Louisanna E, Séjalon-Delmas N. 2019. How significant are endophytic fungi in bromeliad seeds and seedlings ? Effects on germination, survival and performance of two epiphytic plant species. *Fungal Ecology* **39**: 296–306.

Love MI, Huber W, Anders S, Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome biology* **15**: 521–550.

Mašínová T, Bahnmann BD, Větrovský T, Tomšovský M, Merunková K, Baldrian P. 2017. Drivers of yeast community composition in the litter and soil of a temperate forest. *FEMS microbiology ecology* **93**: 1–10.

Maude RB. 1996. *Seedborne Diseases and Their Control: Principles and Practice*.

McMurdie PJ, Holmes S. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**: e61217.

Mcshea WJ, Healy WM, Devers P, Fearer T, Frank H, Stauffer D, Waldon J. 2019. Forestry matters : Decline of oaks will impact wildlife in hardwood forests. *Journal of Wildlife management* **71**: 1717–1728.

Minard G, Tikhonov G, Ovaskainen O, Saastamoinen M. Variation in Melitaea cinxia gut microbiota is phylogenetically highly structured but 2only mildly driven by host plant microbiota, sex or parasitism. *bioRxiv*, in press.

Monahan WB, Koenig WD. 2005. Estimating the potential effects of sudden oak death on oak-dependent birds. *Biological Conservation* **127**: 146–157.

Mueller, G.M., Bills, G.F., Foster MSB. 2004. *Biodiversity of Fungi : Inventory and monitoring methods*.

Nalim FA, Samuels GJ, Wijesundera RL, Geiser DM. 2011. New species from the *Fusarium solani* species complex derived from perithecia and soil in the old world tropics. *Mycologia* **103**: 1302–1330.

Nelson EB. 2004. Microbial dynamics and interactions in the spermosphere. *Annual Review of Phytopathology* **42**: 271–309.

Nelson EB. 2017. The seed microbiome: origins, interactions, and impacts. *Plant and Soil* **422**: 7–34.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Nelson EB, Simoneau P, Barret M, Mitter B, Compant S. **2018**. The soil, the seed, the microbes and the plant. *Plant Soil* **422**: 1–5.

Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, Knelman JE, Darcy JL, Lynch RC, Wickey P, *et al.* 2013. Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews* **77**: 342–356.

Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG. **2016**. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology* **20**: 241–248.

Oberwinkler F, Bandoni R, Blanz P. **1983**. Cystofilobasidium : a New Genus in the *Filobasidiaceae* *. *Systematic and applied microbiology* **4**: 114–122.

Ofek M, Hadar Y, Minz D. **2011**. Colonization of cucumber seeds by bacteria during germination. *Environmental Microbiology* **13**: 2794–2807.

Ovaskainen O, Tikhonov G, Norberg A, Guillaume Blanchet F, Duan L, Dunson D, Roslin T, Abrego N. **2017**. How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters* **20**: 561–576.

Palou L, Guardado A, Monstesinos-Herrero C. **2010**. First report of *Penicillium* spp. and *Pilidiella granati* causing postharvest fruit rot of pomegranate in Spain. *New disease reports* **22**: 21.

Peris JE, Rodríguez A, Peña L, Fedriani JM. **2017**. Fungal infestation boosts fruit aroma and fruit removal by mammals and birds. *Scientific Reports*: 1–9.

Pertot I, Giovannini O, Benanchi M, Caf T, Rossi V, Mugnai L. **2017**. Combining biocontrol agents with different mechanisms of action in a strategy to control *Botrytis cinerea* on grapevine. **97**: 85–93.

Pinciroli M, Gribaldo A, Vidal A, Bezus R, Sisterna M. **2013**. Mycobiota evolution during storage of paddy, brown and milled rice in different genotypes. *Summa phytopathologica* **39**: 157–161.

Prochazkova Z, Sikorova A, Peskova V. **2005**. Preliminary observations on the occurrence of *Ciboria batschiana* (Zopf) Buchwald in the Czech Republic. *Working Papers of the Finnish Forest Research Institute* **11**: 13–18.

Qin Y, Pan X, Yuan Z. **2016**. Seed endophytic microbiota in a coastal plant and phytobeneficial properties of the fungus *Cladosporium cladosporioides*. *Fungal Ecology* **24**: 53–60.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Quaedvlieg W, Verkley GJM, Shin HD, Barreto RW, Alfenas AC, Swart WJ, Groenewald JZ, Crous PW. **2013**. Sizing up *Septoria*. *Studies in Mycology* **75**: 307–390.

Reyes-Estebanez M. **2011**. Antimicrobial and nematicidal screening of anamorphic fungi isolated from plant debris of tropical areas in Mexico. *African Journal of Microbiology Research* **5**: 1083–1089.

Rezki S, Campion C, Simoneau P, Jacques MA, Shade A, Barret M. **2018**. Assembly of seed-associated microbial communities within and across successive plant generations. *Plant and Soil* **422**: 67–79.

Rillig MC, Antonovics J, Caruso T, Lehmann A, Powell JR, Veresoglou SD, Verbruggen E. **2015**. Interchange of entire communities: Microbial community coalescence. *Trends in Ecology and Evolution* **30**: 470–476.

Sampaio JP, Golubev WI, Fell JW, Gadanho M, Golubev NW. **2004**. *Curvibasidium cygneicollum* gen. nov., sp. nov. and *Curvibasidium pallidicorallinum* sp. nov., novel taxa in the microbotryomycetidae (Urediniomycetes), and their relationship with *Rhodotorula fujisanensis* and *Rhodotorula nothofagi*. *International Journal of Systematic and Evolutionary Microbiology* **54**: 1401–1407.

Santiago IF, Rosa CA, Rosa LH. **2016**. Endophytic symbiont yeasts associated with the Antarctic angiosperms Deschampsia antarctica and Colobanthus quitensis. *Polar Biology* **40**:177-183

Schiltz S, Gaillard I, Pawlicki-Jullian N, Thiombiano B, Mesnard F, Gontier E. **2015**. A review: what is the spermosphere and how can it be studied? *Journal of Applied Microbiology* **119**: 1467–1481.

Schneider C, Rasband W, Eliceiri K. **2012**. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671–675.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, et al. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* **109**: 6241–6246.

Schubert K, Groenewald JZ, Braun U, Dijksterhuis J, Starink M, Hill CF, Zalar P, De Hoog GS, Crous PW. **2007**. Biodiversity in the *Cladosporium herbarum* complex (*Davidiellaceae*, *Capnodiales*), with standardisation of methods for *Cladosporium* taxonomy and diagnostics. *Studies in Mycology* **58**: 105–156.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Shahzad R, Khan AL, Bilal S, Asaf S, Lee I. **2018**. What Is There in Seeds? Vertically Transmitted Endophytic Resources for Sustainable Improvement in Plant Growth. *Frontiers in Plant Science* **9**.

Sneck ME, Rudgers JA, Young CA, Miller TEX. 2017. Variation in the prevalence and Transmission of heritable symbionts across host populations in heterogeneous environments. Microbial Ecology 74: 640–653.

Sogonov M V., Castlebury LA, Rossman AY, Mejía LC, White JF. **2008**. Leaf-inhabiting genera of the *Gnomoniaceae*, *Diaporthales*. *Studies in Mycology* **62**: 1–77.

Thambugala KM, Hyde KD, Tanaka K, Tian Q, Wanasinghe DN, Ariyawansa HA, Jayasiri SC, Boonmee S, Camporesi E, Hashimoto A, *et al.* 2015. Towards a natural classification and backbone tree for Lophiostomataceae, Floricolaceae, and Amorosiaceae fam. nov. *Fungal Diversity* **74**: 199–266.

Torres-Cortés G, Bonneau S, Bouchez O, Genthon C, Briand M, Jacques M-A, Barret M. **2018**. Functional microbial features driving community assembly during seed germination and emergence. *Frontiers in Plant Science* **9**: 902.

Truyens S, Weyens N, Cuypers A, Vangronsveld J. **2015**. Bacterial seed endophytes: Genera, vertical transmission and interaction with plants. *Environmental Microbiology Reports* **7**: 40–50.

U'Ren JM, Arnold AE. **2016**. Diversity, taxonomic composition, and functional aspects of fungal communities in living, senesced, and fallen leaves at five sites across North America. *PeerJ* **4**: e2768.

Vacher C, Hampe A, Porté AJ, Sauer U, Compant S, Morris CE. **2016**. The Phyllosphere: Microbial jungle at the plant–climate interface. *Annual Review of Ecology, Evolution, and Systematics* **47**: 1–24.

Vannier N, Mony C, Bittebiere A, Michon-coudouel S, Biget M, Vandenkoornhuyse P. **2018**. A microorganisms' journey between plant generations. *Microbiome* **6**: 1–11.

Vellend M. **2010**. Conceptual synthesis in community ecology. *The quarterly review of biology* **85**: 183–206.

Vivas M, Kemler M, Slippers B. 2015. Maternal effects on tree phenotypes : Considering the microbiome. *Trends in Plant Science* 20: 541–544.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Vivas M, Kemler M, Mphahlele MM, Wingfield MJ, Slippers B. **2017**. Maternal effects on phenotype, resistance and the structuring of fungal communities in *Eucalyptus grandis*. *Environmental and Experimental Botany* **140**: 120–127.

Wang Q, Garrity GM, Tiedje JM, Cole JR. **2007**. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**: 5261–5267.

Whitham TG, Bailey JK, Schweitzer| JA, Shuster SM, Bangert RK, LeRoy CJ, Lonsdorf E V., Allan GJ, DiFazio SP, Potts BM, et al. 2006. A framework for community and ecosystem genetics : from genes to ecosystems. *Nature Reviews Genetics* **7**: 510–523.

Wolf JB, Wade MJ. **2009**. What are maternal effects (and what are they not)? **364**: 1107–1115.

Yang L, Danzberger J, Schöler A, Schröder P, Schloter M, Radl V. **2017**. Dominant Groups of Potentially Active Bacteria Shared by Barley Seeds become Less Abundant in Root Associated Microbiome. *Frontiers in Plant Science* **8**: 1–12.

Zhang J, Kobert K, Flouri T, Stamatakis A. **2014**. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**: 614–620.

Zhou J, Ning D. **2017**. Stochastic community assembly : Does it matter in microbial ecology? *Microbiology and Molecular Biology Reviews* **81**: 2-17.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Supplementary Materials

Table S1. Most abundant fungal species found as endophytes in seeds of sessile oak. Seeds were collected in the tree canopy or on the ground and were surface sterilized. The fruit wall and embryo were separated and their fungal community was analyzed using a metabarcoding approach. Only Amplicon sequence variants (ASV) assigned to Ascomycota (A) or Basidiomycota (B) with the UNITE database were kept. Average relative abundances of all ASVs were computed for each sample type, after merging ASVs assigned to the same fungal species. Only ASVs identified at the species level are shown in the table.

Fungal species	**Average relative abundance (%)**			**Lifestyle**	**Reference(s)**
	All internal tissues	Embryos	Fruit walls		
Curvibasidium cygneicollum (B)	8.7	6.8	10.4	Endophyte of fruits, leaves, trunk and soil behaving as a phytopathogen or a mycoparasite. The species is insensitive to the mycocins produce by *Filobasidium* and *Cystofilobasidium*.	Sampaio *et al.* (2004); Mašínová *et al.* (2017)
Epicoccum nigrum (A)	6.1	7.2	5	Ubiquitous fungus found in soil, leaves and seeds described as primary saprotroph and plant pathogen.	Ahumada-Rudolph *et al.* (2014)
Mycosphaerella tassiana (A)	4.7	1	8.2	Common pathogen found in the phyllosphere including that of oak.	Schubert *et al.* (2007); Jakuschkin *et al.* (2016)
Gibberella baccata (A)	4.5	4.5	4.5	Worldwide pathogen found in the phyllosphere and rarely found in seeds or soils.	Douglas Boyette & Lynn Walker (1986); Desjardins, (2003)
Cladosporium delicatulum (A)	3.2	3.5	3	Found in cereal seeds, mycoparasite of *Taphrina spp* and *Magnaporthe oryzae*.	Amanelah Baharvandi & Zafari (2015) ; Chaibub *et al.* (2016)
Caliciopsis beckhausii (A)	1.9	1.7	2	Most species of the genus are saprotrophs or parasites growing on bark of *Quercus* sp.	Garrido-Benavent & Perez-Ortega (2015)
Rhodosporidiobolus colostri (B)	1.5	0.7	2.2	Ubiquitious yeast behaving as a fungal antagonist.	Golubev and Tomashevskaya (2009)
Angustimassarina acerina (A)	1.4	2.4	0.5	Found on twigs of *Acer platanoides* and in bioaerosols.	Thambugala (2015); Banchi *et al.* (2018)
Penicillium paczoskii (A)	1.1	1.1	1.1	Pathogen causing post-harvest fruit rots recovered from harvested seeds.	Palou *et al.* (2010)
Taphrina carpini (A)	1	1	1.1	Common pathogen encountered on *Fagaceae* leaves.	Bacigálová (1991); Inácio *et al.* (2004); Cordier *et al.* (2012)

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table S2. Permutational multivariate analyses of variance (PERMANOVA) of compositional dissimilarities among fungal communities of seeds of sessile oak. Dissimilarities among seeds were estimated using the Jaccard quantitative distance. The total number of sequences per sample (sequencing depth, SD) was log-transformed and introduced as the first explanatory variable in all models. The effects of tree population (T), seed position (canopy *versus* ground, P) and their interaction were tested on the whole acorn dataset while the effect of mother tree was tested separately on canopy seeds and ground seeds.

	d.f.	F-value	*P*-value	R^2
All seeds				
Log(SD)	1	1.3	0.018	0.01630
Tree population (T)	3	2.1	**<.001**	0.07967
Seed position (P)	1	1.8	**<.001**	0.02240
T x P	3	1.2	**<.01**	0.04619
Seeds in the canopy				
Log(SD)	1	1.0	0.626	0.02588
Mother tree	10	1.2	**0.011**	0.30870
Seeds on the ground				
Log(SD)	1	1.1	0.238	0.02935
Mother tree	11	1.2	0.150	0.34904

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table S3. Response of fungal ASVs to site elevation and seed position depending on their lifestyle, in the HMSC presence-absence (PA) and sequence count (SC) models. In the PA model, values are the posterior probabilities that the community weighted mean specialization, or the proportion of pathogens or saprotrophs is higher in the reference modality of the ecological factor. In the SC model, values are the posterior probabilities that the community weighted mean specialization, or the abundance of pathogens or saprotrophs (conditional on presence) is higher in the reference modality of the ecological factor. Probabilities higher than 0.9 are in bold.

		PA model			SC model		
Ecological factor	Reference modality	Specialization	Pathogen	Saprotroph	Specialization	Pathogen	Saprotroph
Elevation	High	**0.96875**	0.57875	**0.97825**	0.71475	0.84475	0.75225
Seed position	Ground	0.78575	**0.92525**	**0.9225**	0.38525	0.42925	0.33575

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table S4. Nestedness and turnover components of compositional dissimilarities between fungal communities of seeds in the canopy and seeds on the ground. Compositional dissimilarities among fungal communities were estimated for each mother tree using the binary Jaccard dissimilarity index and partitioned with the beta.temp function of the betapart package (Baselga et al. 2018). A mother tree at the Bager site was excluded from the analysis because it had only two seeds on the ground.

		Dissimilarity between ground seeds and canopy seeds		
Tree population	**Mother tree**	**Turnover**	**Nestedness**	**Total**
Ade	33	0.75	0.06	0.81
	34	0.78	0.09	0.86
	55	0.85	0.08	0.92
Bager	11	0.66	0.09	0.75
	13	0.54	0.35	0.90
Gabas	32	0.62	0.17	0.79
	37	0.84	0.03	0.87
	40	0.76	0.08	0.84
Gedre bas	23	0.77	0.13	0.90
	24	0.91	0.01	0.92
	28	0.84	0.05	0.89

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table S5. Number and percentage of fungal ASVs of seeds on the ground also found in seeds in the canopy and mother tree aboveground tissues. A mother tree at the Bager site was excluded from the analysis because it had only two acorns on the ground.

Site	Mother tree	Total number of ASVs in seeds on the ground	Number (%) of ASVs in seeds on the ground shared with	
			Seeds in the canopy	Mother tree aboveground tissues
Ade	33	35	14 (40%)	17 (48%)
	34	142	26 (18%)	30 (21%)
	55	109	11 (10%)	44 (40%)
Bager	11	121	39 (32%)	59 (48%)
	13	247	27 (10%)	58 (23%)
Gabas	32	101	26 (25%)	51 (50%)
	37	67	18 (26%)	30 (44%)
	40	169	37 (21%)	56 (33%)
Gedrebas	23	105	12 (11%)	30 (28%)
	24	42	6 (14%)	11 (26%)
	28	102	16 (15%)	25 (24%)

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table S6. Fungal ASVs positively associated with another other ASV at the seed level according to HMSC models. Putative functions were inferred using FUNGuild and literature search. Only ASVs assigned at the species level are shown. The network of positive associations is represented on Figure 4.

Fungal species	Guild inference		Lifestyle	References
	FUNGuild	Literature		
Caliciopsis beckhausii	Pathotroph	Plant pathogen	Most species of the genus are saprotrophs or parasites growing on bark of *Quercus* sp.	Garrido-Benavent & Perez-Ortega (2015)
Cladosporium delicatulum	Unknown	Seed endophyte Putative mycoparasite	Found in cereal seeds, mycoparasite of *Taphrina spp* and *Magnaporthe oryzae*.	Amanelah Baharvandi & Zafari (2015) ; Chaibub *et al.* (2016).
Curvibasidium cygneicollum	Unknown	Generalist ubiquitous Putative mycoparasite	Endophyte of fruits, leaves, trunk and soil behaving as a phytopathogen or a mycoparasite. The species is insensitive to the mycocins produce by *Filobasidium* and *Cystofilobasidium*.	Sampaio *et al.* (2004); Mašínová *et al.* (2017)
Cylindrium elongatum	Unknown	Fungal antagonist Ubiquitous	Bacterial and fungal antagonist found on oak leaves.	Reyes-Estebanez (2011) ; Duarte *et al.* (2015)
Cystofilobasidium capitatum	Unknown	Seed endophyte Putative mycoparasite	Ubiquitious yeast which colonize the cotyledon after germination of english oak. The genus *Cystofilobasidium* have hyphal structures with haustoria only in the dikaryotic phase suggesting mycoparasites capacities.	Oberwinkler *et al.* (1983); Isaeva *et al.* (2009); Bills, Muller & Foster (2004)
Epicoccum nigrum	Pathotroph	Plant pathogen	Ubiquitous fungus found in soil, leaves and seeds described as primary saprotroph and plant pathogen.	Ahumada-Rudolph *et al.* (2014)
Filobasidium wieringae	Saprotroph	Generalist ubiquitous Putative mycoparasite	Yeast found in the phyllosphere. The genus *Filobasidium* have hyphal structures with haustoria only in the dikaryotic phase suggesting mycoparasites capacities.	Glushakova & Kachalkin (2017); Bills, Muller & Foster (2004)
Mycosphaerella tassiana	Pathotroph	Plant pathogen	Common pathogen found in the phyllosphere including that of oak.	Schubert *et al.* (2007); Jakuschkin *et al.* (2016)
Polyscytalum algarvense	Pathotroph	Plant pathogen	Necrotroph fungi found on *Eucalyptus* leaves.	Cheewangkoon *et al.* (2009); Crous (2018)
Rhodosporidiobolus colostri	Unknown	Fungal antagonist Ubiquitous	Ubiquitious yeast behaving as a fungal antagonist.	(Golubev & Tomashevskaya, 2009)
Sphaerulina amelanchier	Symbiotroph	Leaf and litter endophyte	Found in leaf litter of *Amelanchier*, *Betula*, *Castanea* and *Quercus*.	Quaedvlieg *et al.* (2013)
Taphrina carpini	Pathotroph	Plant pathogen	Common pathogen encountered on *Fagaceae* leaves.	Bacigálová (1991); Inácio *et al.* (2004); Cordier *et al.* (2012)
Trichomerium foliicola	Symbiotroph	Leaf epiphyte	Foliar epiphytes. *Trichomerium* is a genus apparently gaining their nutrients from insect exudates.	Chomnunti *et al.* (2012)
Vishniacozyma victoriae	Unknown	Generalist ubiquitous Fungal antagonist	Microbial antagonist of *Penicillium expansum* and *Botrytis cinerea*. Ubiquitous yeast able to tolerate stressful environments.	Pertot *et al.* (2017); Santiago *et al.* (2016)

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license.

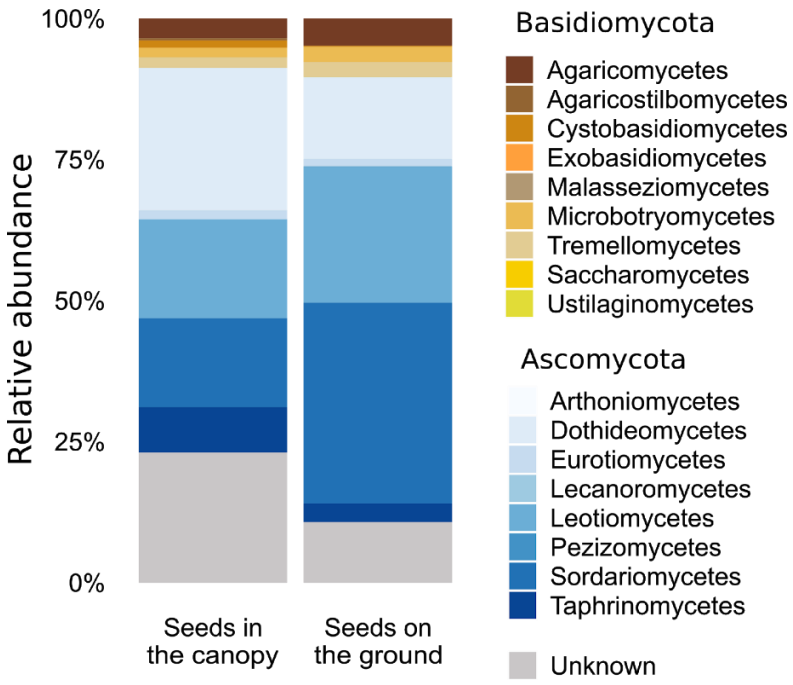


Figure S1. Taxonomic composition of fungal communities of seeds of sessile oak. The barplot indicates the average sequence percentage assigned to each fungal class.

bioRxiv preprint first posted online Jul. 3, 2019; doi: http://dx.doi.org/10.1101/691121. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license.

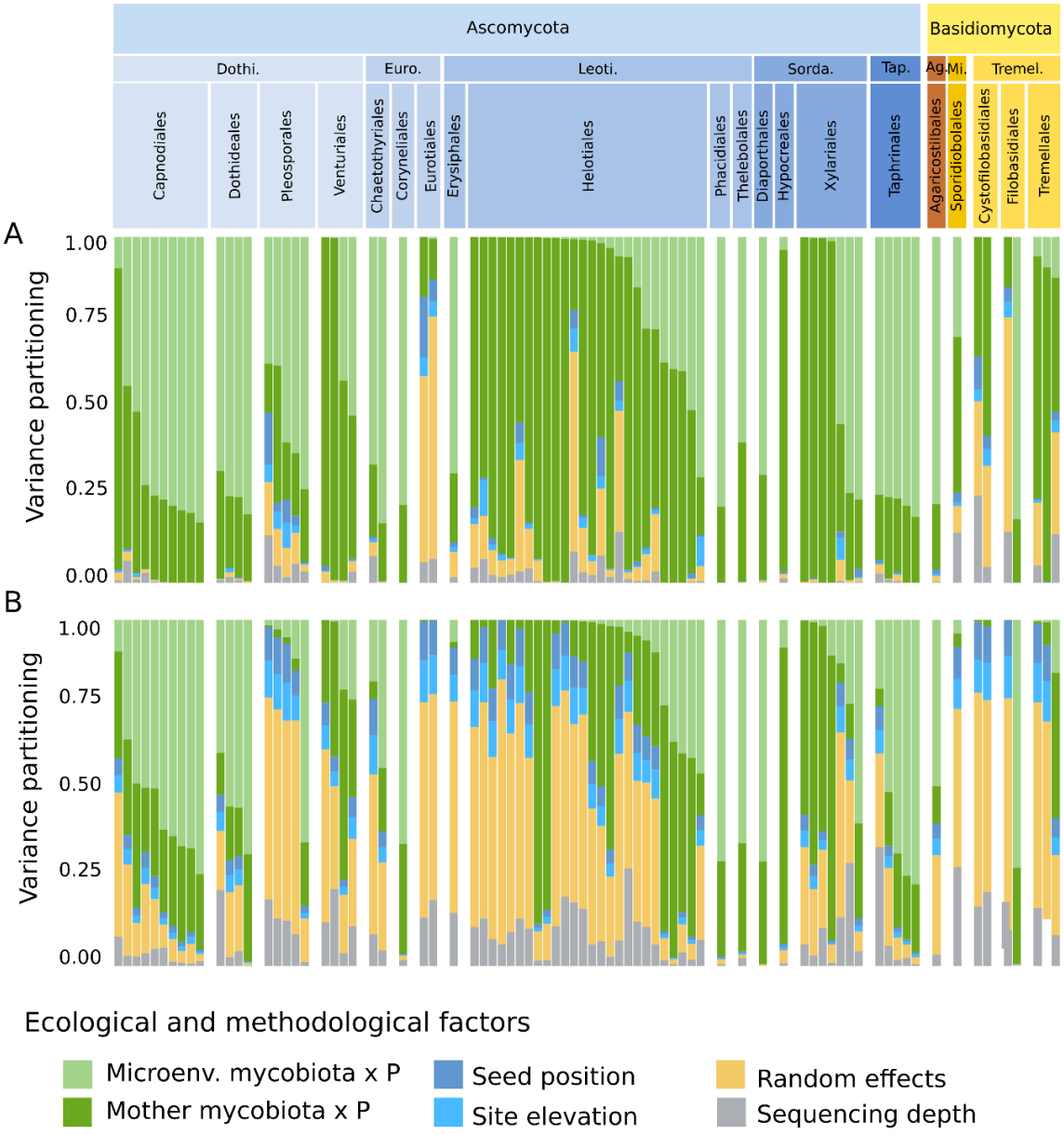


Figure S2. Partitioning of the variance in the composition of seed fungal communities of sessile oak. Four fixed effects were included in the HMSC models to explain variations in (A) the presence-absence or (B) the sequence count of a focal fungal ASV among seed samples: *Sequencing depth* (total number of sequences per sample), *Seed position* (P, canopy or ground), *Microenv. mycobiota* (relative abundance of the focal ASV in the seed microenvironment), *Mother mycobiota* (relative abundance of the focal ASV in the mother tree aboveground tissues), and *Site elevation*. *Mother mycobiota* and *Microenv. mycobiota* were introduced in interaction with P. Random effects were included at each spatial scale (seed, tree and population). Results of variance partitioning are given as percentages (%) of total explained variance. ASV are ranked by fungal phylum, class and order (Dothi: *Dothideomycetes*, Euro: *Eurotiomyecetes*, Leo: *Leotiomyecetes*, Sorda: *Sordariomycetes*, Tap: *Taphrinomycetes*, Ag: *Agaricostilbmoyecetes*, Mi: *Microbotryomycetes*, Tremel: *Tremellomycetes*). Only ASVs assigned at the order level are shown.