Year : 2020

SOIL METAGENOME CHANGES DERIVED FROM EXPERIMENTAL MANIPULATIONS OF ABIOTIC AND BIOTIC CONDITIONS

Rincón Chacón Cristian Camilo

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Department of Ecology and Evolution

SOIL METAGENOME CHANGES DERIVED FROM EXPERIMENTAL MANIPULATIONS OF ABIOTIC AND BIOTIC CONDITIONS

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de Biologie et de médecine
de l'Université de Lausanne

par

Cristian Camilo RINCÓN CHACÓN

Agronome diplômé d'un MSc en sciences moléculaires du vivant
Université de Lausanne, Suisse

Jury
Prof. Stephan Gruber, Président
Prof. Ian Sanders, Directeur de thèse
Prof. Jan Roelof van der Meer, expert
Dr. Klaus Schläppi, expert

Lausanne 2020

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Stephan	**Gruber**
Directeur·trice de thèse	Monsieur	Prof.	Ian	**Sanders**
Expert·e·s	Monsieur	Prof.	Jan Roelof	**van der Meer**
	Monsieur	Dr	Klaus	**Schläppi**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Cristian Camilo Rincon Chacon

Master en sciences moléculaires du vivant, Université de Lausanne

intitulée

Soil metagenome changes derived from experimental manipulations of abiotic and biotic conditions

Lausanne, le 27 avril 2020

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale

Acknowledgements

Funded by a Swiss Confederation scholarship, I started a journey in a new country with different culture, language, weather, a whole adventure. The adventure continued thanks to a Theodor and Gabriela Kummer scholarship. I am indebted to the Swiss education system.

I would like to start by thanking my supervisor Prof. Ian R. Sanders for having welcomed me in his group. His kindness and trust pulled me through this journey. I found in Ian's research the practical application that I want for my work. A particular thanks to the experts Prof. Jan Roelof van der Meer and Dr. Klaus Schläppi for investing their precious time to read and comment on this work. I would also like to thank the president of the jury, Prof. Stephan Gruber for handling the administrative work linked to this dissertation.

Shortly after arriving in Switzerland, I realized how enriching it is to work in a multilingual and multicultural atmosphere. The Sanders group, the DEE, the DMF, the university, WaPiHo and ESN put me in contact with so many people that shared their time to make me experience a great time during these years of PhD. I particularly thank Tania for her company and kindness, Fréd for all he taught me. Also the members of the Sanders group, Romain, Ivan, Jérôme, Lucas, Cindy, Edward, Quim, Chanz, Ricardo, Isabel, Camilo and Alia for the time spent (scientifically or otherwise). Special thanks to Jérémy Bonvin and Consolée Alleti who helped me with lab work and more.

I feel very fortunate because while leaving my family behind in Colombia I was welcomed in new families. Manon, Philippe, Zoé, Margot and Maya Daeppen became my adoptive family in Switzerland. Also, on the other side of the Röstigraben, Danielle, June, Sonia, René, Marcus, Karl, Eva and many other family members and friends made me feel like at home, all my thanks to them.

My family in Colombia, all of them, from the distance, always encouraging me through the hard moments and celebrating with me the good ones. All my love to them. Last but not least to Salomé for being there for me.

Resumé

Les micro-organismes ont évolué pour occuper presque toutes les niches écologiques imaginables et sont donc impliqués dans un large éventail d'interactions écologiques autant entre eux comme avec des formes de vie supérieures. Les communautés microbiennes du sol jouent un rôle clé dans de nombreux processus écologiques, notamment les cycles biogéochimiques des nutriments, la dégradation des xénobiotiques, la formation d'agrégats du sol et la structuration des communautés végétales. Ces processus étant fondamentaux pour le fonctionnement des écosystèmes terrestres, le pourquoi les communautés microbiennes du sol ont suscité un grand intérêt de la communauté scientifique.

Au fil des années, diverses approches indépendantes de la culture en laboratoire, comme le séquençage non ciblé de tout l'ADN extrait d'un échantillon donné (appelé métagénomique) ont été développées et ont généralement été utilisées pour identifier les espèces microbiennes et les variables environnementales qui déterminent leur abondance et distribution.

Cependant, les études utilisant une méthodologie métagénomique se sont, pour la plupart, limitées à une approche descriptive. Un prix élevé ainsi que la puissance de computationnelle et l'expertise bioinformatique nécessaires pour traiter ce type de données ont limité des expériences répliquées et reproductibles. Désormais, avec du séquençage plus abordable et avec des outils bioinformatiques plus efficaces, des expériences de manipulation des métagénomes peuvent être établies permettant de tester des questions de recherche spécifiques sur les facteurs affectant la diversité et la composition métagénomique du sol.

L'objectif de ce travail était d'étudier comment la manipulation expérimentale par simulation du réchauffement climatique ou inoculation avec des microbes bénéfiques affecte le métagénome du sol. Les résultats ont fourni des informations précieuses concernant les réponses des métagénomes du sol aux changements environnementaux. Cela permettra, à l'avenir, d'étudier les métagénomes du sol et la manière dont son fonctionnement et sa résilience pourraient être exploités pour potentiellement améliorer la productivité des écosystèmes et des agroécosystèmes.

General abstract

Microorganisms have evolved to occupy almost every conceivable ecological niche and therefore are involved in a wide range of ecological interactions with each other and with higher forms of life. Soil microbial communities play key roles in many ecological processes including biogeochemical cycling of nutrients, breakdown of xenobiotics, formation of soil aggregates and structuring of plant communities. As these processes are fundamental to the functioning of terrestrial ecosystems, soil microbial communities have attracted great interest from the scientific community.

Over the years, diverse culture-independent approaches, as the untargeted sequencing of all the DNA extracted from a given sample (referred to as metagenomics) have been developed and have been typically used to identify microbial species and the environmental variables that drive their abundance and distribution.

However, the studies using a metagenomic methodology have mostly been limited to a descriptive approach. The high price, computing power and bioinformatic expertise required to process this kind of data have limited thorough and fully replicated experimental investigations. Now, with more affordable sequencing and more powerful bioinformatic tools, experiments manipulating metagenomes can be established allowing for testing of specific research questions into factors affecting soil metagenomic diversity and composition.

The focus of this work was to investigate how experimental manipulation through either climate warming simulation or inoculation of crops with beneficial microbes affect the soil metagenome. The results yielded valuable information concerning the responses of soil metagenomes to environmental changes. This, in the future will allow investigating soil metagenomes and how its functioning and resilience could be harnessed to potentially improve ecosystems and agroecosystems productivity.

Table of contents

General Introduction and Thesis Outline

Although invisible to the naked eye, microorganisms are an essential component of every ecosystem on earth and represent the vast majority of the genetic and metabolic diversity on the planet, they are often referred to as the 'unseen majority' (Whitman et al., 1998). Microorganisms have evolved to occupy almost every conceivable ecological niche and therefore are involved in a wide range of ecological interactions with each other and with higher forms of life (Gray & Head, 2008). Soil microbial communities play key roles in many ecological processes including biogeochemical cycling of nutrients (Crowther et al., 2019), breakdown of xenobiotics, formation of soil aggregates (Gattinger et al., 2008) and structuring of plant communities (van der Heijden et al., 2008). As these processes are fundamental to the functioning of terrestrial ecosystems, soil microbial communities have attracted great interest from the scientific community. Over the years, diverse culture-independent approaches (based on the extraction of DNA from soil and a subsequent analysis) have been developed and have been typically used to identify microbial species and the environmental variables that drive their abundance and distribution (Gray & Head, 2008).

As the soil is a highly heterogeneous environment, soil microbial communities often exhibit high taxonomic diversity (Louca et al., 2017). Studies assessing the factors affecting microbial community composition have been carried out mainly in terms of taxonomically identifying the microbes making up the community and what factors affect this composition (Widder et al., 2016). Taxonomy DNA-based approaches such as meta-barcoding rely on amplification and then sequencing of a variable region (e.g. 16S rRNA gene for bacteria) of the genome (informative enough to be used for identification) that is flanked by highly conserved sequences that can serve as annealing sites for PCR primers (Bengtsson-Palme, 2017). However, as DNA sequences vary, primers do not have equal affinity for all possible DNA molecules in a sample and consequently there is an amplification bias during PCR. Additionally, even with well optimized primers, the currently used sequencing technologies yield a relatively small region which often limits this approach to genus level of resolution (Knight et al., 2018). In the past few years, advances in sequencing technologies and bioinformatic tools enabled the studies of soil metagenomes (defined as all microbial genes present in a defined environment). One such method to describe the metagenome is the untargeted sequencing of all the DNA extracted from a given sample (formally named shotgun metagenomics, here referred to as metagenomics). This approach can be used to investigate a broad range of taxonomic aspects and functional potential of microbial communities (Quince et al., 2017). These developments have allowed the study of complex data sets, such as soil microbiomes. There is now an improved understanding of the biogeography (Martiny et al., 2006; Sunagawa et al., 2015; Thompson et al., 2017; Ramirez et al., 2018), ecology (Dumbrell et al., 2010; Martiny et al., 2011; Shade et al., 2012) and functionality (Logue et al., 2016; Louca et al., 2018) of microbiomes of diverse environments. Besides providing information about the relative abundance of microbial functional genes, shotgun metagenomic studies can also offer microbial taxonomic information (Knight et al., 2018). Some studies comparing the accuracy of metagenomic and metabarcoding (16S rRNA gene amplicon data) data to depict taxonomic composition of microbial communities found that metagenomic data was either consistent with 16S rRNA gene data (Manichanh et al., 2008) or outperformed it (Shakya et al., 2013; Campanaro et al., 2018). However, the studies using a metagenomic methodology have mostly been limited to a descriptive approach, in which the microbiomes are studied to, for example, identify previously undescribed taxa that can carry specific metabolic processes. Besides being comparatively expensive, the depth of sequencing required to obtain meaningful conclusions from metagenomic experiments requires also a relatively high computing power and bioinformatic expertise (Sczyrba et al., 2017). In the past, these two factors have limited thorough and fully replicated experimental investigations into factors affecting soil bacterial diversity and gene composition of the metagenome. Such experimental manipulations allow for testing of specific research questions yielding valuable information concerning the roles of soil metagenomes and their response to environmental changes, for example, climate warming, introduction of non-native microbes etc.

The focus of this work is to investigate how experimental manipulation through either climate warming simulation or inoculation of crops with beneficial microbes affect the soil metagenome. This gains relevance as soil microbiomes are key players in soil fertility, functioning and resilience, thus, directly affecting ecosystems and agroecosystems productivity.

The use of beneficial microbes was approached by the use of soil inoculants, such as arbuscular mycorrhizal fungi (AMF), has increased in the recent years as they have shown positive effects on plant productivity (Ceballos et al., 2019; Zhang et al., 2019). As an organism without and observed sexual stage, it has ben long debated how this widely distributed organisms adapt to diverse environments. One explanation is a high intra-isolate genetic diversity (c.f. Chapter 3) which we considered in this study (c.f. Chapter 2).

Climate warming simulation: studying soil metagenome changes derived from abiotic factors manipulation.

In a scenario of climate warming, species, including soil microbial communities, may migrate upwards to track its current climate. High elevation communities could face diverse scenarios depending on whether lowland species establish or fail to establish at higher elevations (Alexander et al., 2015). Given that microbial species are rarely restricted by geographical barriers (Finlay, 2002) it is likely that some microbes will easily have the capacity to move to higher elevations. Thus, novel plant - soil microbiome interactions could occur and could subsequently have an effect on ecosystem functioning. Studies have been carried out to analyze vegetation changes induced by warming but changes in soil microbial communities remain relatively less studied. Additionally, studies addressing this topic have been developed within a small geographical region with a defined set of biotic and abiotic conditions which limits the extrapolation of the results to other habitats. Finding patterns in microbial gene diversity across a series of regions could provide valuable information for predicting the responses of microbial community functionality to climate change.

Inoculation of crops with beneficial microbes: investigating soil metagenome changes resulting from introduction of biotic agents.

Another way in which soil metagenomes may be affected is by human introduction of microbes for agricultural purposes. The use of beneficial microbial inoculants in agriculture has gained attention because of the capacity of some microbial taxa or microbial communities to provide ecological services as promotion of plant growth and protection against pathogens (Berg, 2009). One of such group of inoculants are arbuscular mycorrhizal fungi (AMF). These fungi form one of the commonest plant–microbe mutualisms. The large majority of terrestrial plants, including many important crops, form arbuscular mycorrhizas (van der Heijden et al., 2015). Their main beneficial effect is uptake and transfer of low-mobility minerals (mainly phosphorus) from the soil to plants, thus improving plant nutrition and productivity. It has been shown that inoculation with genetically different isolates of the model AMF *Rhizophagus irregularis* can have considerable effects on plant growth (Angelard et al., 2010; Ceballos et al., 2019). Furthermore, AMF have been shown to increase plant yield under field conditions with cassava (Ceballos et al., 2013, 2019) and cereal crops (Zhang et al., 2019). Cassava is considered a food security crop, feeding almost 800 million people in the tropics (Howeler, 2013). Because this plant produces starchy roots, changes in yield imply additional carbohydrate allocation towards the belowground plant biomass, which in turn could have an effect on the microbiome surrounding the roots. How AMF inoculation affects the soil microbial communities has not been thoroughly examined. Also, as the use of such inoculants can bring economic benefits to the farmer, these inoculants are now being used in many parts of the world without knowledge about the potential ecological impacts of such practices (Hart et al., 2017). One of the concerns about the use of these inocula include potentially invasive AMF isolates that may directly or indirectly be detrimental to local AMF diversity (Schwartz et al., 2006). Only until very recently, the impact of introduction of AMF on the root microbial community was studied, using however, a metabarcoding approach (Akyol et al., 2019). The effects of AMF inoculation on soil metagenomes have not been yet addressed and this is, to our knowledge, the first study investigating this topic. This information is essential to define a framework for developing successful application of AM fungi in agriculture.

Tracking inoculated AMF: developing tools to allow the study of soil metagenome changes

In the tropics, the inherently low fertility soils are experiencing increasing anthropogenic pressure and the effects of climate change (Jiao et al., 2019) which in turn limits the productivity of land ecosystems and agroecosystems. As a countermeasure to this situation, the use of microbial inoculants, such as AMF, has increased. It is worth noticing that AMF are naturally present in soils around the world (Öpik et al., 2010) and yet inoculation generates a plant growth response even in the presence of a preestablished AMF community (Janoušková et al., 2013; Niwa et al., 2018). A successful management strategy incorporating AMF inoculation should include methodology that can verify the persistence of the inoculated strain among the native microbial community. Verbruggen et al. (2013) suggested that long-term beneficial effects of AMF inoculation can be achieved with persistent introduced AMF species. If the introduced AMF fails to establish in the long term, inoculation would be necessary with every crop cycle and this can be a limiting factor if such practices are to be applied to promote food security in developing countries. Thus, assessment of the persistence of the inoculated AMF isolate is an important component for environmentally safe and economically sustainable use of these inoculants (Pellegrino et al., 2012).

Thesis outline

In chapter 1, using a metagenomic approach the changes of soil metagenome due to transplantation treatments (climate warming simulation) were evaluated. This project was carried out within a collaborative consortium where many researchers across continents (Europe and the US) studied many aspects of biotic changes resulting from community transplant of turfs from high to low elevation sites (TransPlant project). Such a large scale experimental metagenome study has not yet been conducted and the results provide valuable information for predicting the responses of microbial communities to climate change.

In chapter 2, a metagenome dataset was generated to assess the impact of AMF inoculation and intra-isolate AMF genetic diversity on the taxonomic and metagenomic profile of the soil microbial community. Such a study allowed us to identify the soil bacterial metabolic pathways affected by the inoculation with AMF and whether this is influenced by genetic variation in the fungus. This knowledge contributes to set a baseline to, in the future, study the functional changes in the soil microbiome derived from AMF inoculation.

In chapter 3, I investigated the within fungus genetic variability of AMF species in an attempt to develop isolate-specific molecular markers in order to be able to track inoculum persistence in the field. As the AMF inoculation effects on the metagenome need to be further studied, having such markers would allow researchers to investigate whether the changes of the soil bacterial metagenome composition are related to the persistence of the inoculated fungus.

References

Akyol TY, Niwa R, Hirakawa H, Maruyama H, Sato T, Suzuki T, Fukunaga A, Sato T, Yoshida S, Tawaraya K, et al. 2019. Impact of Introduction of Arbuscular Mycorrhizal Fungi on the Root Microbial Community in Agricultural Fields. Microbes and Environments 34: 23–32.

Alexander JM, Diez JM, Levine JM. 2015. Novel competitors shape species' responses to climate change. Nature 525: 515–518.

Angelard C, Colard A, Niculita-Hirzel H, Croll D, Sanders IR. 2010. Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. Current Biology 20: 1216–1221.

Bengtsson-Palme J. 2017. Strategies for Taxonomic and Functional Annotation of Metagenomes. Elsevier Inc.

Berg G. 2009. Plant-microbe interactions promoting plant growth and health: Perspectives for controlled use of microorganisms in agriculture. Applied Microbiology and Biotechnology 84: 11–18.

Campanaro S, Treu L, Kougias PG, Zhu X, Angelidaki I. 2018. Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. Scientific Reports 8: 1–12.

Ceballos I, Mateus ID, Peña R, Peña-Quemba DC, Masso C, Vanlauwe B, Rodriguez A, Sanders IR. 2019. Using variation in arbuscular mycorrhizal fungi to drive the productivity of the food security crop cassava. : 1–21.

Ceballos I, Ruiz M, Fernández C, Peña R, Rodríguez A, Sanders IR. 2013. The In Vitro Mass-Produced Model Mycorrhizal Fungus, *Rhizophagus irregularis*, Significantly Increases Yields of the Globally Important Food Security Crop Cassava. PLoS ONE 8.

Crowther TW, van den Hoogen J, Wan J, Mayes MA, Keiser AD, Mo L, Averill C, Maynard DS. 2019. The global soil community and its influence on biogeochemistry. Science 365.

Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH. 2010. Relative roles of niche and neutral processes in structuring a soil microbial community. ISME Journal 4: 337–345.

Finlay BJ. 2002. Global dispersal of free-living microbial eukaryote species. Science 296: 1061–1063.

Gattinger A, Palojärvi A, Schloter M. 2008. Soil Microbial Communities and Related Functions. In: Perspectives for Agroecosystem Management. Elsevier, 279–292.

Gray ND, Head IM. 2008. Microbial Ecology. In: Encyclopedia of Ecology. Hoboken, NJ, USA: Elsevier, 2357–2368.

Hart MM, Antunes PM, Abbott LK. 2017. Unknown risks to soil biodiversity from commercial fungal inoculants. Nature Ecology and Evolution 1: 1.

van der Heijden MGA, Bardgett RD, Van Straalen NM. 2008. The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology Letters 11: 296–310.

van der Heijden MGA, Martin FM, Selosse MA, Sanders IR. 2015. Mycorrhizal ecology and evolution: The past, the present, and the future. New Phytologist 205: 1406–1423.

Howeler RH. 2013. Save and grow : cassava : a guide to sustainable production intensification. Rome: Food and Agriculture Organization of the United Nations.

Janoušková M, Krak K, Wagg C, Štorchová H, Caklová P, Vosátka M. 2013. Effects of inoculum additions in the presence of a preestablished arbuscular mycorrhizal fungal community. Applied and Environmental Microbiology 79: 6507–6515.

Jiao S, Chen W, Wei G. 2019. Resilience and assemblage of soil microbiome in response to chemical contamination combined with plant growth (I Cann, Ed.). Applied and Environmental Microbiology 85: 1–16.

Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D, et al. 2018. Best practices for analysing microbiomes. Nature Reviews Microbiology 16: 1–13.

Logue JB, Findlay SEG, Comte J. 2016. Microbial Responses to Environmental Changes (JB Logue, SEG Findlay, and J Comte, Eds.). Frontiers Media SA.

Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF, Doebeli M. 2017. High taxonomic variability despite stable functional structure across microbial communities. Nature Ecology & Evolution 1: 1–12.

Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, Ackermann M, Hahn AS, Srivastava DS, Crowe SA, et al. 2018. Function and functional redundancy in microbial systems. Nature Ecology & Evolution 2: 936–943.

Manichanh C, Chapple CE, Frangeul L, Gloux K, Guigo R, Dore J. 2008. A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. Nucleic Acids Research 36: 5180–5188.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman J a, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, et al. 2006. Microbial biogeography: putting microorganisms on the map. Nature reviews. Microbiology 4: 102–112.

Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. 2011. Drivers of bacterial β-diversity depend on spatial scale. Proceedings of the National Academy of Sciences of the United States of America 108: 7850–7854.

Niwa R, Koyama T, Sato T, Adachi K, Tawaraya K, Sato S, Hirakawa H, Yoshida S, Ezawa T. 2018. Dissection of niche competition between introduced and indigenous arbuscular mycorrhizal fungi with respect to soybean yield responses. Scientific Reports 8: 2–5.

Öpik M, Vanatoa A, Vanatoa E, Moora M, Davison J, Kalwij JM, Reier Ü, Zobel M. 2010. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). New Phytologist 188: 223–241.

Pellegrino E, Turrini A, Gamper HA, Cafà G, Bonari E, Young JPW, Giovannetti M. 2012. Establishment, persistence and effectiveness of arbuscular mycorrhizal fungal inoculants in the field revealed using molecular genetic tracing and measurement of yield components. New Phytologist 194: 810–822.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology 35: 833–844.

Ramirez KS, Knight CG, de Hollander M, Brearley FQ, Constantinides B, Cotton A, Creer S, Crowther TW, Davison J, Delgado-Baquerizo M, et al. 2018. Detecting macroecological patterns in bacterial communities across independent studies of global soils. Nature Microbiology 3: 189–196.

Schwartz MW, Hoeksema JD, Gehring CA, Johnson NC, Klironomos JN, Abbott LK, Pringle A. 2006. The promise and the potential consequences of the global transport of mycorrhizal fungal inoculum. Ecology Letters 9: 501–515.

Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, et al. 2017. Critical Assessment of Metagenome Interpretation - A benchmark of metagenomics software. Nature Methods 14: 1063–1071.

Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JBH, et al. 2012. Fundamentals of microbial community resistance and resilience. Frontiers in Microbiology 3: 1–19.

Shakya M, Quince C, Campbell JH, Yang ZK, Schadt CW, Podar M. 2013. Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. Environmental Microbiology 15: 1882–1899.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, et al. 2015. Structure and function of the global ocean microbiome. Science 348: 1261359.

Thompson LR, Sanders JG, Mcdonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, et al. 2017. A communal catalogue reveals Earth ' s multiscale microbial diversity.

Verbruggen E, van der Heijden MGA, Rillig MC, Kiers ET. 2013. Mycorrhizal fungal establishment in agricultural soils: Factors determining inoculation success. New Phytologist 197: 1104–1109.

Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: The unseen majority. Proceedings of the National Academy of Sciences of the United States of America 95: 6578–6583.

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, Cordero OX, Brown SP, Momeni B, Shou W, et al. 2016. Challenges in microbial ecology: Building predictive understanding of community function and dynamics. ISME Journal 10: 2557–2568.

Zhang S, Lehmann A, Zheng W, You Z, Rillig MC. 2019. Arbuscular mycorrhizal fungi increase grain yields: a meta-analysis. New Phytologist 222: 543–555.

.

Chapter 1: Metagenomic profiling of soil microbial communities following climate warming simulation. A “TransPlant” experiment.

Cristian Rincón[1], Tom W. N. Walker[1,2], Chelsea Chisholm[1,2], Jake M. Alexander[1,2] and Ian R. Sanders[1].

[1]Department of Ecology and Evolution; University of Lausanne; Lausanne, Switzerland. [2]Department of Environmental Systems Science; ETH Zurich; Zurich, Switzerland.

Abstract

The key role played by soil microbial communities in some of the most important ecosystem processes is undeniable. However, despite their importance for carbon and nutrient cycling, soil fertility and structuring of plant communities, the understanding of how soil microbial communities and their functioning will be affected by climate warming is still very limited.

Transplantation of intact turfs moved to lower elevation in a mountain gradient was used to expose microbial communities to a warmer climate in combination with a new neighboring community. This allowed the investigation of the net effect of both direct (altered climatic conditions) and indirect (altered biotic interactions) effects of climate warming on community responses.

Then, using a metagenomic approach, we investigated whether transplantation showed consistent effects on the potential metabolic capabilities of the soil microbial community across different geographical regions. Differential gene abundance analyses suggested that the transplantation had little effect over the gene composition of the microbial communities. However, this trend was not consistent for all the regions. Elevation arose as a component potentially driving the gene composition of the observed communities which suggests, in turn, that the main constraining factors affecting the metagenome likely remain linked to the climatic variables.

Later, information related to nutrient cycling and plant communities performance will be integrated to the results presented here.

Introduction

The key role played by soil microbial communities in some of the most important ecosystem processes is undeniable. However, despite their importance for carbon (Bahram *et al.*, 2018) and nutrient cycling (Crowther *et al.*, 2019), soil fertility (Luo *et al.*, 2016) and structuring plant communities (Wagg *et al.*, 2014), the understanding of how soil microbial communities and their functioning will be affected by climate change is still very limited (de Vries & Griffiths, 2018).

Alpine ecosystems are particularly vulnerable to climate warming given the particularities such as the short plant growth season (Donhauser & Frey, 2018). In a climate warming scenario, species may migrate upwards to track its current climate. This is likely to affect plant and animal species but also soil microbial communities. Alpine soil microbial communities could face a range of different scenarios depending on whether lowland species establish or fail to establish at higher elevations (Alexander *et al.*, 2015).

Given that microbial species are rarely restricted by geographical barriers (Finlay, 2002) the movement of microbes to higher elevations as the climate warms could result in novel plant-soil microbiome and soil microbe-microbe interactions that could ultimately affect community structure and function. Novel interactions among soil organisms can modify carbon fluxes, mineral nutrient cycles (van der Putten *et al.*, 2016) and modify host specific interactions with pathogens (Klironomos, 2002) or mutualists (Wagg *et al.*, 2011).

Open top chambers represent one of the most used methodologies available for investigating climate warming effects on communities. These greenhouse-like structures can increase mean daily air temperature (Yang *et al.*, 2018). However, such chambers can limit colonization processes, thus restricting the potential novel interactions arising from migrations as well as comprising changes in other potentially confounding factors (Alexander *et al.*, 2015). Another approach that overcomes some of these limitations is community transplant experiments. In these experiments, intact turfs of whole plant communities are moved to a lower elevation exposing them to a warmer climate in combination with a new neighboring community. This allows the investigation of the net effect of both direct (altered climatic conditions) and indirect (altered biotic interactions) effects of climate change on community responses (Alexander *et al.*, 2015).

Studies have been carried out to analyze vegetation changes due to climate warming but, at present, soil microbial community responses to such warming remain relatively understudied. Researchers in biogeography of soil microbial community patterns have studied the effects of several environmental factors. Bacterial community composition is known to change in response to elevated CO_2 (Weber *et al.*, 2011; Hayden *et al.*, 2012) where Acidobacteria was shown to decrease in abundance while Actinobacteria and Bacteroidetes increased. Research studying the effect of an elevation gradient (which incorporates a number of confounding environmental factors), showed that patterns of plant and bacterial diversity in response to elevation gradients were fundamentally different. Bacterial taxon richness and phylogenetic diversity decreased monotonically from low to high elevations while plants followed a unimodal pattern (Bryant *et al.*, 2009). However, in another study bacterial diversity showed no significant trend in response to elevation changes. This was in direct contrast to the significant diversity decrease with increased elevation observed in plant and animal taxa across the same montane gradient (Fierer *et al.*, 2011). It has also been shown that both elevation and microtopography (i.e. ridges, depressions, south-facing, and north-facing slopes) play a role in the structuring of soil microbial communities. Bacterial alpha diversity was only affected by micro-topography while elevation did not affect richness or evenness of the bacterial communities (Frindte *et al.*, 2019). Sheik *et al.* (2011) found that warming, using infrared heaters, increased total microbial abundance but decreased bacterial diversity. Likewise, it was observed that warming (induced either via buried resistance heating cables or infrared heaters) increased the abundance of bacterial taxa associated with oligotrophic strategies, such as Acidobacteria (DeAngelis *et al.*, 2015) and Alphaproteobacteria (Hayden *et al.*, 2012).

The abovementioned studies were conducted using a meta-barcoding approach describing changes in taxonomic abundance and composition of soil microbial communities based on 16S rRNA gene sequences. Therefore, these investigations considered the taxonomic composition of the soil bacterial communities but did not consider the metabolic capabilities of the bacterial groups that could potentially alter how such microbiomes influence important ecosystem processes in the soil (de Vries & Griffiths, 2018). The reason for so many studies adopting a meta-barcoding approach was because methodological constraints prevented a detailed exploration of the soil bacterial metagenome. With the development of improved molecular techniques, higher computational power and better bioinformatic tools, research has made a shift towards the study of the potential functional traits of soil microbial communities by sequencing the metagenome of soil microbial communities and elucidating the probable metabolic traits existing within the microbiome (Knight *et al.*, 2018). As it will be seen in chapter 2, the apparent lack of a taxonomic response of the soil microbiome to an environmental perturbation can mask underlying changes in the metabolic capabilities of the genes making up the microbiomes metagenome.

Given the importance of soil microbial communities mediating biogeochemical processes, exploring the functional biogeography of the soil microbiome is key for improving accuracy in global biogeochemical model predictions (Cavicchioli *et al.*, 2019; Crowther *et al.*, 2019). A few studies have investigated the changes in abundance and diversity of genes comprising the soil microbial metagenome along elevation gradients (Zhang *et al.*, 2013; Gao *et al.*, 2014; Yang *et al.*, 2014; Shen *et al.*, 2016; Qi *et al.*, 2017) and some have studied gene abundance changes of the microbiome in other climate change experiments (Zhao *et al.*, 2014; Yue *et al.*, 2015). Notably, these studies were all carried out in Chinese mountain ranges, thus, considering a relatively small geographical scale. This limits the extrapolation of these results to other habitats around the globe. In these studies, changes in functional gene diversity were observed with elevation and implicated a decrease in genes involved in C- and N-cycling with warming. This is seemingly inconsistent with observations from other areas across the northern hemisphere (Yue *et al.*, 2015).

So far, the combination of an experimental manipulation of the soil microbial communities through turf transplantation, and its replication across multiple regions, has not been addressed and could provide valuable information for predicting the responses of microbial gene composition to climate change.

Here, we used a metagenomic approach to characterize whether there were changes in the potential metabolic capabilities of the soil microbial community in a climate warming scenario. As the samples used in this study originated from several mountain ranges (central Europe, Scandinavia, USA), this would allow us to characterize patterns in microbial gene diversity across a large geographical scale and to see whether any effects of elevation gradients on the soil metagenome showed consistent patterns across different geographical regions. This work was part of a larger project that took advantage of a network of soil transplant experiments across elevation gradients in multiple regions. The larger project aimed to investigate ecosystem responses to climate change, more specifically, our collaborators aimed to establish broad-scale patterns of the impacts of novel plant-soil interactions on ecosystem processes.

Materials and methods

Field sites

This study was performed in a series of experimental plots set up by an international group of collaborators in which whole plant and soil communities were exposed to new climatic conditions by transplanting turfs along altitudinal gradients. This collaborative project is known as the TransPlant Project. The TransPlant Project aims to improve the understanding of the ecosystem responses to climate warming by simulating the effects of species movement from lower altitudes to higher altitudes. There were three treatments per locality. First, communities were transplanted from high to low elevation (here denoted “HL”). Second, communities were transplanted from the low site back into the same low site, (here denoted “LL” and third, communities were transplanted from the high site back into the high site, referred as “HH”. Field site characteristics of the sample plots are summarized in Table S1. A total of 10 localities were sampled. The localities are denoted as follows: Calanda (CAL) and Lavey (LAV) in Switzerland; Lautaret (LAU) and Villard-Reculas (VR) in France; Granau-Hochaml (GH) in Germany; Arizona (ARI), Montana (MON) and the Rocky Mountain Biological Laboratory (RMBL) in the U.S. and Skjaelligehaugen (SKJ) and Ulvehaugen (ULV) in Norway.

The project within which this work was developed aims to investigate the community and ecosystem consequences of novel plant-soil interactions following climate change. To do so, the project comprised the transplant experiments in which analyses were performed on plant community structure and diversity, DNA-based meta-barcoding of soil microbial communities and measurements of ecosystem carbon and nutrient cycling.

Soil sampling

From each turf, at each site and in each treatment, five soil cores were taken at random within the plot and pooled to account for potentially high small-scale heterogeneity in the soil community. Soil cores were 10 mm in diameter, to minimize disturbance to the community, and 50 mm in depth, to focus sampling solely on the rhizosphere. Soil cores were mixed and stored in RNA*later*™, to stabilize and protect nucleic acids, then shipped to the University of Lausanne and stored at -80 °C until DNA extraction. Up to 8 replicates were taken per treatment per locality (Table S1).

DNA extraction, library preparation and sequencing.

The RNA*later*™ was removed from the soil samples by centrifugation. Briefly, an equal volume of ice cold 1x PBS was added to the sample to reduce the density of the solution. Tubes were then centrifuged at 14.000g for 5 min and the supernatant discarded. Subsequently, total soil DNA was extracted using a Fast DNA Spin Kit for soil (MP Biomedical) following the protocol of the manufacturer. Next, the DNA extractions were equimolarly pooled per treatment to end up with 3 samples per locality (HH, HL, and LL), for the 10 localities, thus totaling 30 libraries. Sequencing libraries were prepared using a TruSeq DNA kit (Illumina). Libraries were validated and quantified using the fluorometric protocol of Promega®. Finished libraries were processed with Fragment Analyzer to assess their quality. Libraries were sequenced with Illumina HiSeq 4000 150 base pair, paired-end multiplexing 10 libraries per lane.

Data processing

Raw data of the 30 libraries (3 treatments: HH, HL and LL from each of the 10 localities) were initially checked with FastQC v0.11.4 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After library quality check, the adaptor sequences were trimmed using TagCleaner (Schmieder *et al.*, 2010). Reads were then quality-filtered (`min_qual_mean 20`) and trimmed using Prinseq-lite version

0.20.4 (Schmieder & Edwards, 2011). Low quality 3'-ends were trimmed and reads containing uncalled bases (N) removed. Only reads of at least 120 bp long were kept for further analyses.

Metagenome co-assemblies were constructed by locality (10 co-assemblies, each with 3 samples) using Megahit v1.1.4 (Li *et al.*, 2015) (`--k-list 33,47,63,77,93,117, --min-contig-len 750`). Metrics of the assemblies were assessed with metaQUAST v5.0.2 (Mikheenko *et al.*, 2016). Resulting contigs were processed with cd-hit v4.6.8 (Fu *et al.*, 2012) to reduce redundancy by clustering reads using a similarity threshold of 95% over at least 90% of the alignment. The resulting file was annotated in parallel by performing an ORF prediction with prodigal v2.6.3 (Hyatt *et al.*, 2010) and, subsequently, blast-like annotation against the NCBI RefSeq non-redundant (NR) sequence database of proteins (Pruitt *et al.*, 2007) and eggNOG v4.5 (Huerta-Cepas *et al.*, 2016) using diamond v0.9.18 (Buchfink *et al.*, 2014) with an e-value threshold of 0.001.

Quality filtered reads were mapped back to the annotated contigs to obtain a count matrix of the abundance of each contig with BBMap v37.82 (Bushnell B., https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/) using default parameters. These matrices, containing the abundance per sample of each annotated contig, were used as input data for the R packages and will be referred to hereon as the gene catalogues. Count matrices were also built for KEGG orthologs (KO) (Kanehisa *et al.*, 2016). Gene-encoding nucleotide sequences were defined as ‘‘genes’’ in this study (Sunagawa *et al.*, 2015).

Statistical analyses

Climatic data for the locations, based on GPS coordinates, was obtained by rasterizing and stacking the layers of the 19 bioclimatic variables available from the CHELSA (climatologies at high-resolution for the earth's land surface areas) database (Karger *et al.*, 2017). The values of the CHELSA database, corresponding to the 10 evaluated localities, are presented in Table S2. A PCA was performed on the environmental data to assess the degree of overlapping of the climatic variables of the samples. Samples were grouped either by region or by locality (Figure 1) using vegan function `ordihull` (Oksanen, 2013).

Given the differences in “year range” (time elapsed between transplantation and sampling) and vegetation across the localities (Table S2), linear mixed-effects models (LMM) were applied to determine the existence of a treatment effect (fixed effect) accounting for variance inflicted in the response variable (abundance of each gene) by the “locality” and “elevation” factors (random effects). The model including the random effect was compared to the null model using a likelihood ratio test (LRT). The genes for which the LRT showed a significant improvement of the model fit ($p < 0.05$) were classified using the clusters of orthologous groups (COGs) system (Tatusov, 2000). The 20 most frequent COGs were plotted adding the remaining to an “others” category.

The R package DESeq2 (Love *et al.*, 2014) was used to identify differentially abundant genes between treatments. First, the gene abundance matrix was filtered to remove entries with less than 25 mapped reads. Abundance of a genes was considered significant if the absolute value of the $\log_2$ fold change was higher than 4 and the FDR Benjamini-Hochberg adjusted p-value lower than 0.01. DESeq2 was also used to estimate the differential abundance of KOs. To contextualize the KO assignations a series of marker genes for several metabolic processes (Louca *et al.*, 2017; Salazar *et al.*, 2019) were contrasted for the three pair-wise comparisons between treatment (i.e. HH vs HL, HH vs LL and HL vs LL).

To explore the dissimilarities of the communities between treatments and their relation with the climatic data, a canonical ordination (Redundancy Analysis, RDA) was built on a Bray-Curtis dissimilarity matrix using the environmental data from CHELSA as explanatory variables. Variance inflation factors (VIF) of the variables were estimated. These measure the proportion by which the variance of a regression coefficient is inflated in the presence of other explanatory variables. Strong linear dependencies (correlations) were found among the explanatory variables in the RDA model. To reduce the complexity of the model, while considering the correlations, forward selection in packfor’s `forward.sel` was applied to explore a potential reduction of the number of explanatory variables (Borcard *et al.*, 2011) and a new RDA was constructed with only the retained variables after forward selection.

To determine the degree of similarity, in terms of gene composition, between HL samples to either the HH or LL samples the quotient of the beta diversity between HL and LL and the beta diversity between HL and HH was computed (βdiv HL vs LL ÷ βdiv HL vs HH, hereon called dissimilarity ratio). With this dissimilarity ratio (which was log transformed), a result above 0 indicates that HL is closer in its gene composition to HH at a given locality. A result below 0 indicates that HL is closer in its gene composition to LL. Different beta diversity indices were used to estimate the dissimilarity ratio as each one weights differently several parameters of the gene composition in each sample. The dissimilarity ratio based on the Bray-Curtis index was correlated to the CHELSA variables. Correlation coefficients higher than 0.35 were plotted.

Results

A total of 1098 million reads were obtained with an average of 33 million reads per pooled sample. On average, 93% of the reads in each sample were considered of sufficiently high quality after filtering and trimming (Table S3).

After co-assembly, an average of 390K contigs were obtained among sites. Samples from VR and LAU (France) had a much larger number of contigs (650K contigs) than most of the other samples. Metrics generated by MetaQUAST showed a similar contig length distribution for each of the localities. The total number of nucleotides integrated in the assembly showed the same trend with VR and LAU (France) exhibiting higher values. Samples from MON (US) presented a smaller than average largest contig (29 kb compared to an average of 73 kb). However, the assembly in its totality was used in further analyses as neither the N50 value, nor the total length of the assembly, were lower than the average (Table S4).

On average, the number of predicted ORFs represented 766K, with VR and LAU (France) exhibiting twice this amount. This matched the trend seen in the amount of contigs obtained after assembly. Annotation revealed that 86.29% of the ORFs retrieved a hit to a protein in the NR protein database (Table S3).

The climate data ordination showed PC1 and PC2 explaining 67.7% of the variance observed. The distance between localities reflected how dissimilar their climatic conditions were (Figure 1a). The US and Scandinavia-Alps presented distinct climatic conditions and Scandinavia and the Alps presented an overlapping of their climatic space. Likewise, the distance between treatments indicated the dissimilarity of climate variables and experimental conditions (i.e. elevation and year range) between the high and low elevation plots within a locality (Figure 1b). Permutational analysis of variance (`adonis`) showed significant differences ($P < 0.001$) among regions and among localities.

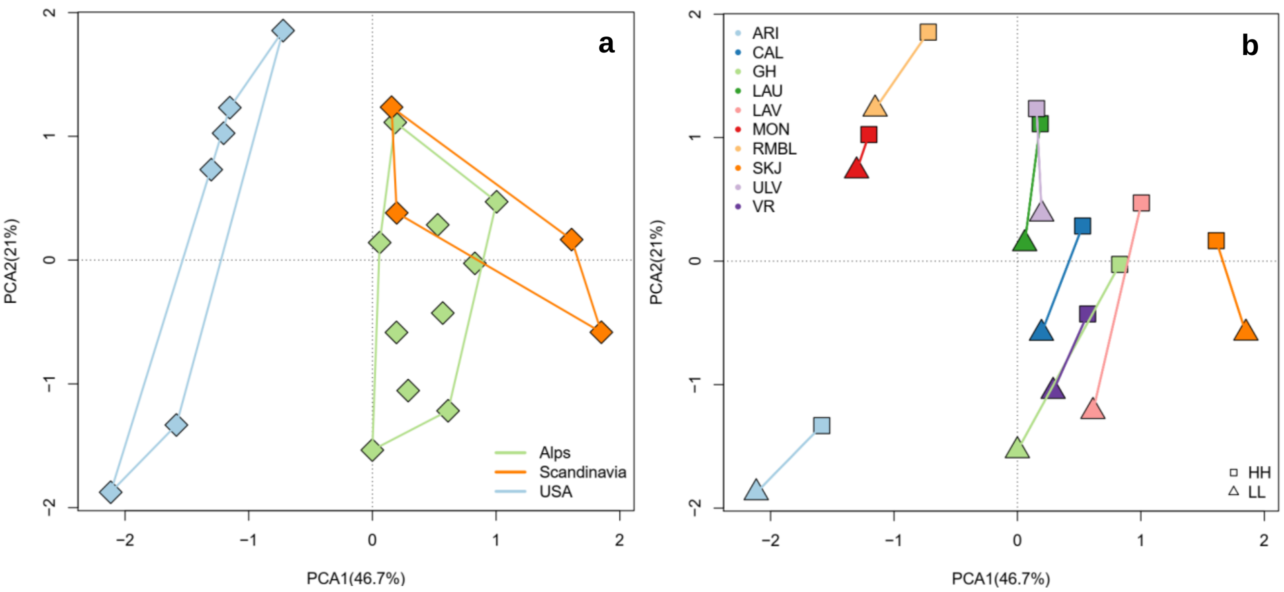
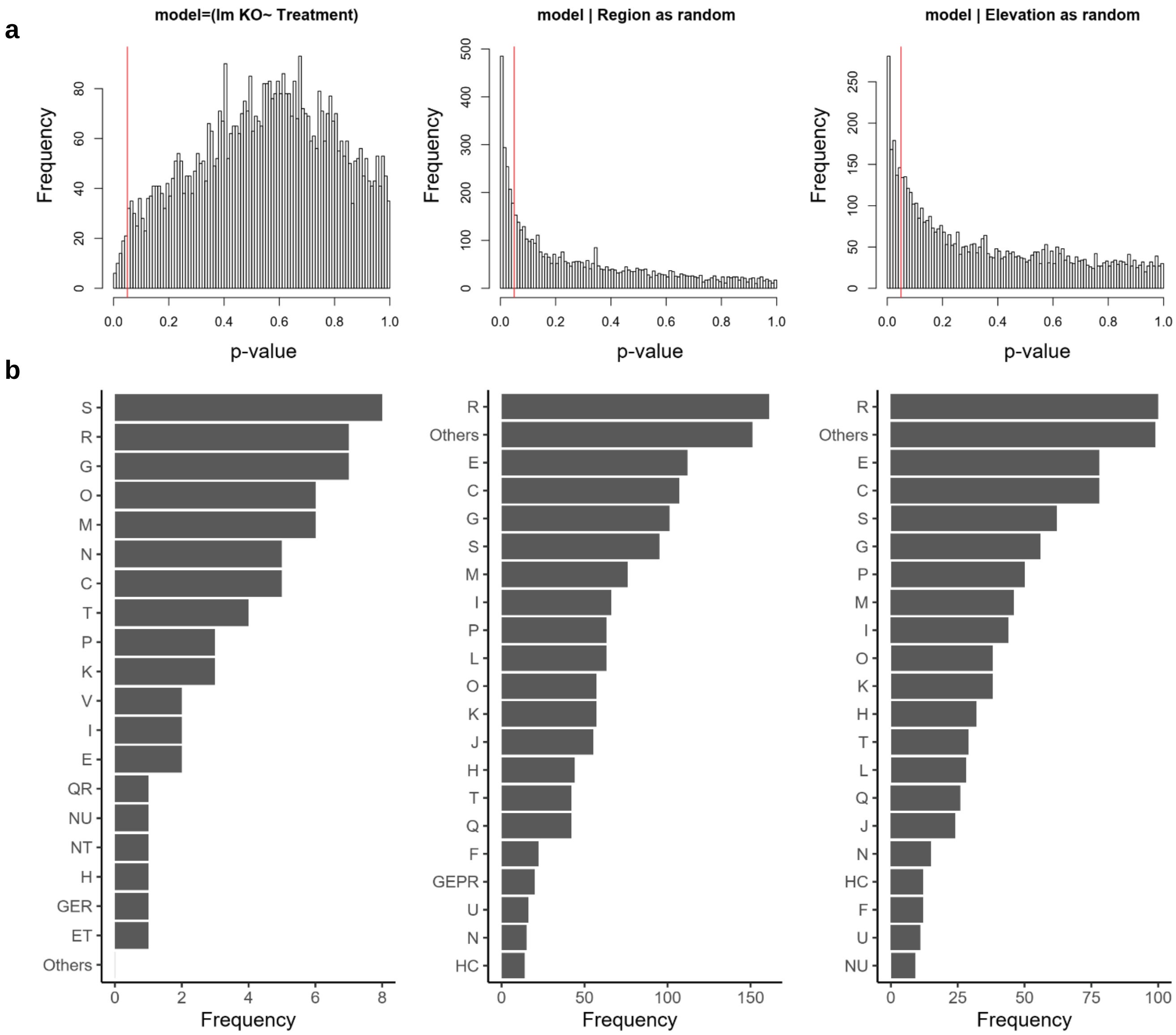


Figure 1. PCA of the climatic variables from the CHELSA database to represent the “climate space”. **a.** Polygons group samples by “region” Alps (CH, FR and DE), Scandinavia (NO) and USA (US). **b.** Distance between the HH (squares) and LL (triangles) sites per locality reflects the similarity of their climatic patterns.

Once the variance generated by the localities and their elevation was accounted for using the mixed-effect model approach, 1418 and 911 KOs (out of 5464 KOs evaluated) showed a better model fit (Figure 2a), indicating that the abundance of those KOs, in the different samples, was indeed affected by the transplantation treatments. Using a fixed effect model (i.e. with the confounding variation of “locality” and “elevation”) on the gene abundances, revealed 70 KOs (out of 5464 KOs evaluated) that were significantly affected by the treatments ($P < 0.05$). The KOs significantly affected by the treatments were assigned to COG categories (Figure 2b).



A	RNA processing and modification
B	Chromatin Structure and dynamics
C	Energy production and conversion
D	Cell cycle control and mitosis
E	Amino acid metabolism and transport
F	Nucleotide metabolism and transport
G	Carbohydrate metabolism and transport
H	Coenzyme metabolism
I	Lipid metabolism
J	Translation
K	Transcription
L	Replication and repair
M	Cell wall/membrane/envelop biogenesis
N	Cell motility
O	Post-translational modification
P	Inorganic ion transport and metabolism
Q	Secondary Structure
T	Signal Transduction
U	Intracellular trafficking and secretion
Y	Nuclear structure
Z	Cytoskeleton
R	General Functional Prediction only
S	Function Unknown

Figure 2. Once the variation generated by the localities is accounted for, it is possible to observe an effect of the transplantation treatments. **a)** Histograms of the Likelihood Ratio Test applied to the comparisons of the Linear Mixed-effects Model (LMM) vs the null model the model with locality and elevation treated as random effects. The red line marks the *P* < 0.05 threshold. **b)** Bar plots of the 20 most frequent COG categories for the genes that showed a model improvement (LMM vs the null model) as defined by the AIC score.

A large proportion of the KOs were assigned to the COG categories “R: General Functional Prediction only” and “S: Function Unknown”. Around 11% were grouped as “others”, comprising categories with low representation (each one below 4%). After “R” and “S”, the most abundant categories were “E: Amino Acid metabolism and transport”, “C: Energy production and conversion” and “G: Carbohydrate metabolism and transport”. All in all, these 5 COGs categories accounted for almost 50% of the KOs classified.

The differential gene abundance comparisons (using the DESeq2 approach) between treatments, when grouping all the localities irrespective of locality, showed no differentially abundant genes for the HL vs HH comparison (i.e. comparing changes in gene abundance in the soil of the transplanted turfs with the original high elevation site), 3234 for HH vs LL (i.e. comparing the high and low elevation sites) and 1143 for HL vs LL (i.e. comparing the transplanted turfs with the receiving low elevation site) of a catalogue of 464751 entries (Figure 3a-c).

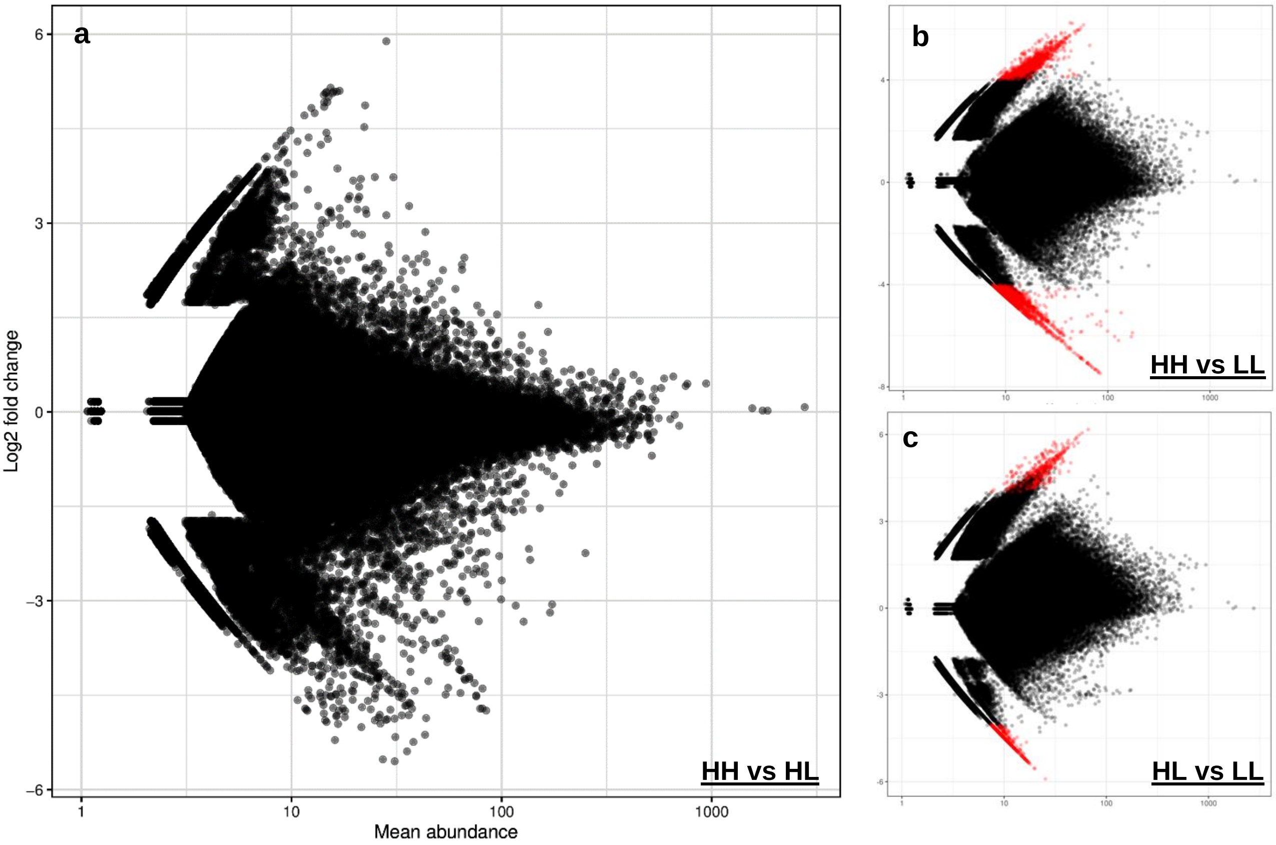


Figure 3. Log2 fold change and mean abundance of the gene count matrix. Each panel shows a different comparison between treatments (bottom right label). Red dots show genes with a significant (P < 0.01) log$_2$ fold change greater than 4. The axis labels in **a)** apply to all the graphs.

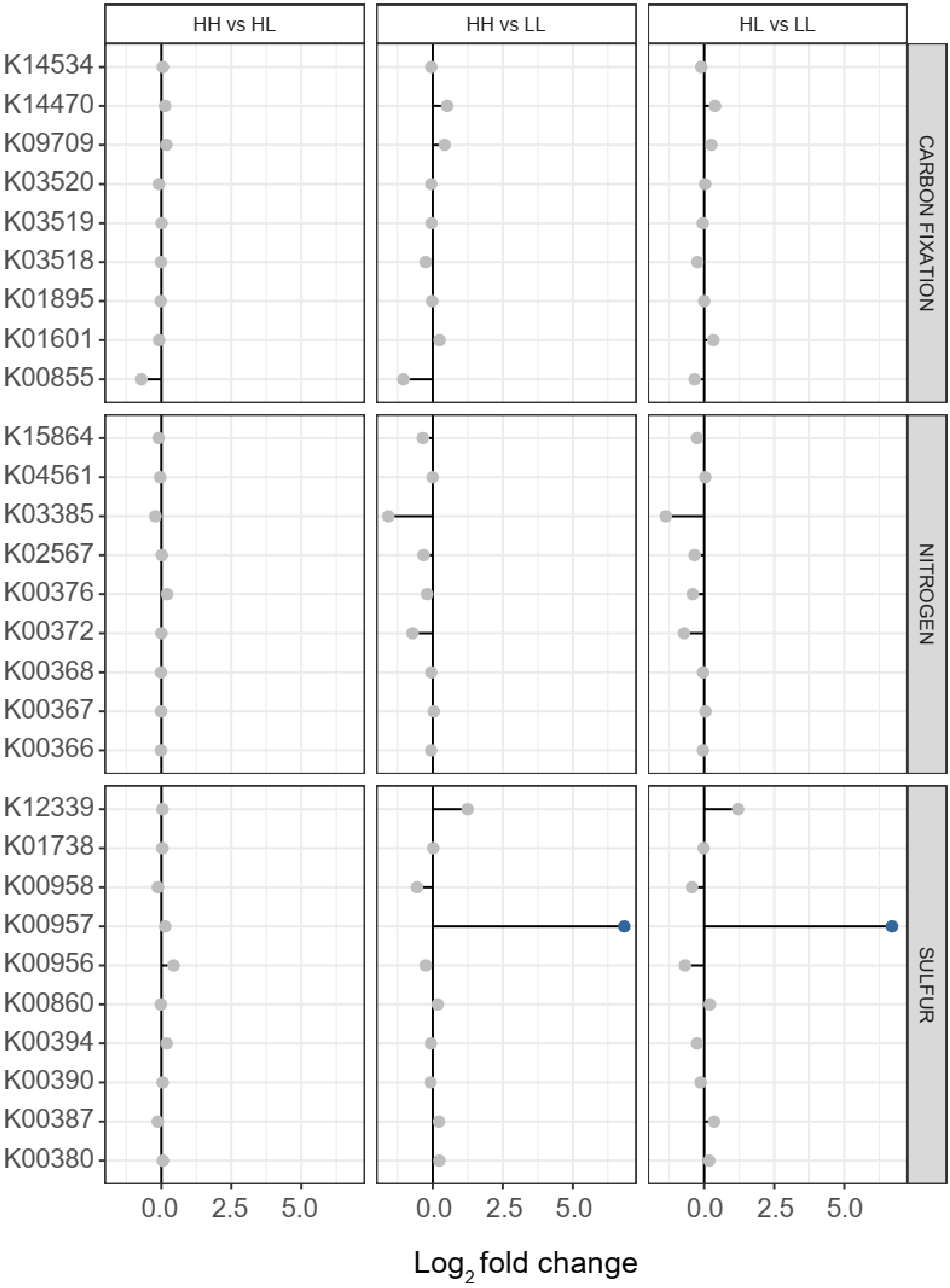


Figure 4. Differences in KO abundance of metabolic marker genes across localities. The data points show the log2 fold change differences computed using DESeq2. Differences were considered significant if *P* ≤ 0.01 after correction (blue dots).

When repeating the DESeq2 approach, but using the KEGG Orthology annotation (which offers information about metabolic pathways), the contrast test showed 60 differentially abundant KOs in the HL vs HH comparison, 104 in the HH vs LL comparison and 82 in the HL vs LL comparison, from a catalogue of 5464 entries. This analysis showed the same trend as that observed in the analysis using the whole gene catalogue where HH and HL sites differ the least and HH and LL sites the most (Figure S1).

Differences were observed in $\log_2$ fold changes of KO abundances of a series of marker genes that were involved in specific metabolic processes in each of the three comparisons (HH vs LL, HL vs LL and HH vs HL; Figure 4). Only K00957 (cysD; sulfate adenylyltransferase subunit 2); a marker for sulfate reduction, was found to be differentially abundant when comparing HH vs LL and HL vs LL treatments.

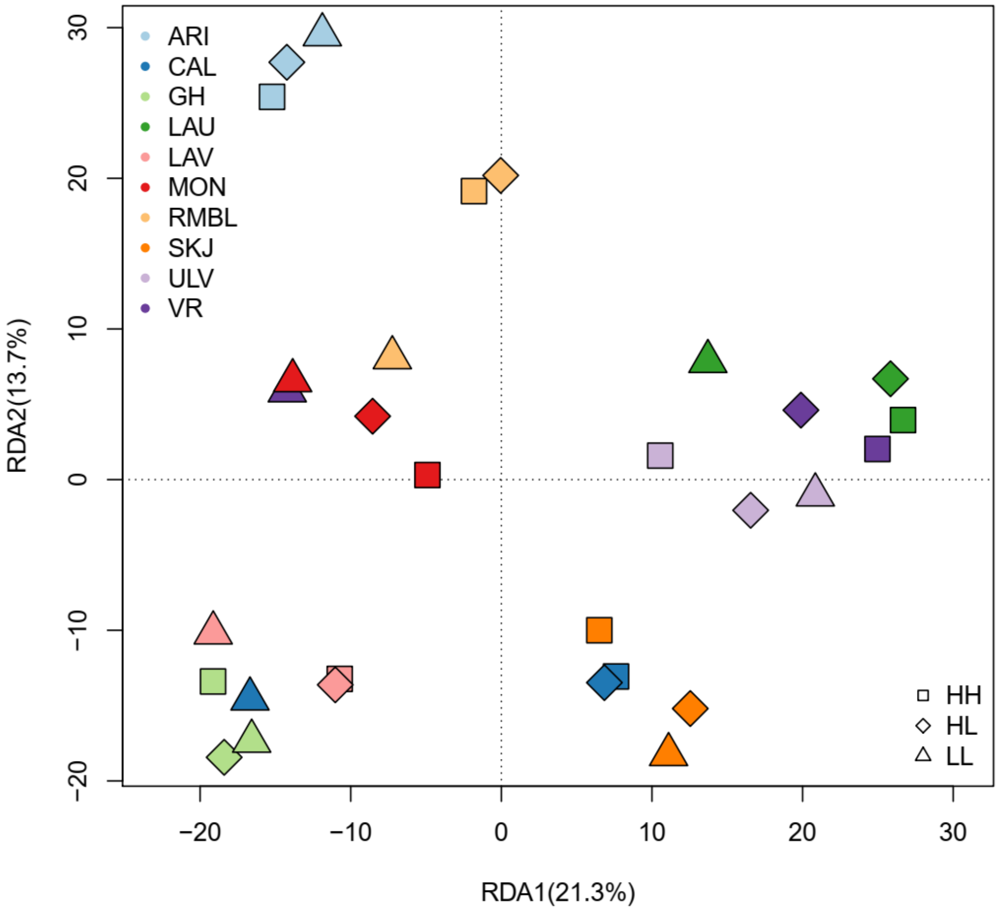


Figure 5. Redundancy Analysis (RDA) of the beta diversity (Bray-Curtis) of the gene catalog using the CHELSA climate data as explanatory variables. Loadings were omitted for clarity. Localities are represented by the different colors and treatments within one locality by shape.

The ordination of beta diversity dissimilarity of the gene composition matrices using the climatic data as explanatory variables (RDA), showed a clustering of samples per locality, indicating higher gene composition similarity within a locality than between localities (Figure 5). This was the case in most of the localities, with the exception of the Low-to-Low (LL) sample from VR (France) that grouped with MON (US) and the LL site from CAL (Switzerland) that grouped with GH (Germany). Additionally, the distance in the plot of High-to-Low (HL) samples to either High-to-High (HH) or Low-to-Low (LL) within a locality depicts a higher similarity of gene composition towards one or another. RDA1 and RDA2 explained 35% of the variance. An ANOVA-like permutation test showed a significant effect ($P < 0.001$) of the climate variables used as constrains in this ordination method. After forward selection the VIFs of the retained variables were less than 10 [VIF above 10 indicates strong collinearity (Blanchet *et al.*, 2008)] and the ANOVA-like permutation test showed a significant effect ($P < 0.001$) of the retained variables on the ordination (Figure S2).

There was no obvious pattern of HL samples being consistently closer to either HH or LL in terms of gene composition (measured as beta diversity). The calculated dissimilarity ratio calculated with five different indices of beta diversity showed a relatively consistent trend in which the HL treatment in SKJ, ULV (Norway) and GH (Germany) were more similar in terms of gene composition to the LL treatment of their respective locality (Figure 6a). The HL treatment in MON (US) was along the line defining the threshold at zero. Subsequently, when elevation of the low site per locality was used to color the ratio of dissimilarity it was revealed that the gene composition of HL sites tended to resemble that of LL sites when the elevation of the locality is at a relative low altitude (< 900 masl). Likewise, it showed that the HL sites tend to be more similar to HH sites when the whole locality (high and low site per locality) lies at a “high” elevation (>1400 masl) (Figure 6b).

The dissimilarity ratio calculated with the Bray-Curtis index tended to correlate to mean annual air temperature, mean daily air temperature of the driest quarter and precipitation seasonality. Elevation mean per locality also showed a relatively high correlation coefficient. However, none of the those correlations were statistically significant (Figure S3).

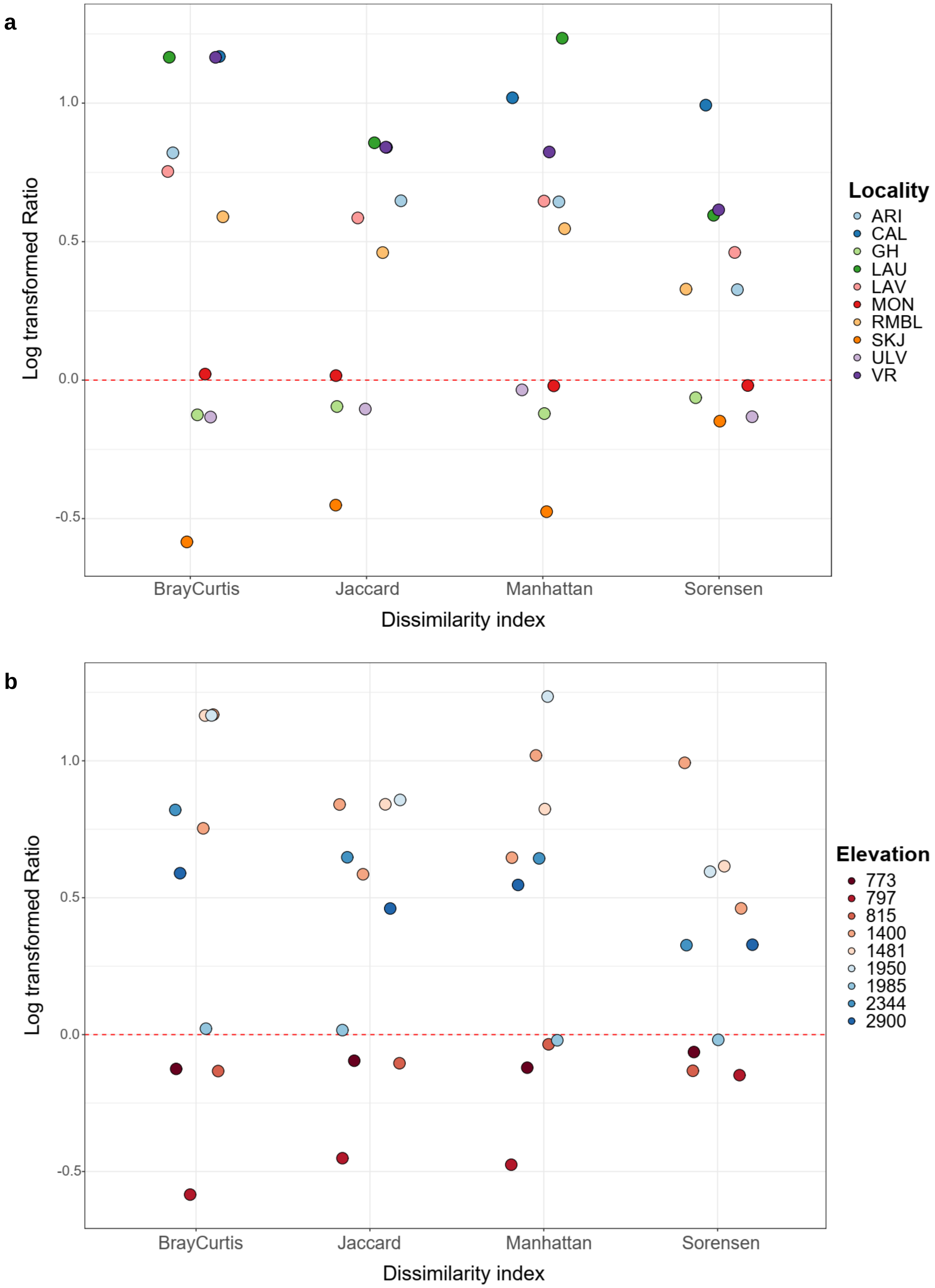


Figure 6. A ratio of dissimilarity indices between treatments within localities was calculated (see methods). The dashed red line indicates the threshold in which the HL sample resembles more the HH treatment (above 0) or LL treatment (below 0) in terms of gene composition. Values were log transformed for visualization. **a)** values were colored by locality. **b)** values were colored by the elevation (in meters above sea level) of the low site of the locality.

Discussion

The soil microbiome is extremely diverse and an integral part of every ecosystem. The microbial community composition, as well as its functionality, have profound impacts on processes that in turn have a direct effect on aboveground functionality. At the same time, changes in plant communities' composition and functioning and alterations on soil chemistry derived from transplantation are likely to have an effect on microbial communities. The collaborators on this project collected information related to nutrient cycling and plant communities performance which will be later integrated with the results presented here.

The differential gene abundance analyses across the localities suggested that the transplantation had little effect over the gene composition of the microbial communities. However, this trend was not consistent for all the localities. Values of gene beta diversity of Granau-Hochaml (GH) in Germany and Skjaelligehaugen (SKJ) and Ulvehaugen (ULV) in Norway suggested that the treatments had an effect, making the transplanted plot more similar in its gene composition to the destination community at low elevation. Looking for a factor linked to this finding, the elevation of the destination plot arose as a component potentially driving the gene composition of the observed communities which suggest, in turn, that the main constraining factors affecting the metagenome likely remain linked to the climatic variables.

The PCA of the climate space showed that there is a low degree of overlapping between the climatic variables of the regions, there was also, however, a large climatic variation within regions. When modeling the gene composition data with the climatic variables it was clear that gene composition grouped by region. This, in turn, indicates that the gene composition was explained, to a certain extent, by the environmental variables. As there was a large variation in climatic conditions across localities, the use of a LMM, introducing "localities" as random effect, allowed us to estimate variation among localities rather than the specific effects of each locality on, in this case, gene abundance (Bolker *et al.*, 2009). The results suggested that once the variation among localities is accounted for, it is possible to observe an effect of the treatments.

The elevation of the destination plot partially explained the observed patterns, independently localities, having a wide range of climatic conditions (Figure 1, Tables S1 and S2). This indicates that the changes generated by the transplantation itself on the gene composition are not as strong as the constraints imposed by environmental variables 'summarized' by the elevation (notably temperature and precipitation). This agrees with the findings of Shen *et al.* (2016) who showed changes in microbial gene composition along an elevational gradient from 530 to 2200 masl using GeoChip microarrays. PCoA revealed a clear distinction in gene composition between samples coming from 1600 masl and below and samples from 1900 masl and above. We observed a similar trend where gene composition of transplanted plots (HL) changed to resemble the low elevation plots (LL) in sites where the elevation was below 1200 masl.

We found that few genes that were affected by the transplantation treatment were assigned to a given function within the COG categories. Of those that were assigned, a large portion corresponded to housekeeping functions. It is possible to link differentially abundant COGs to a particular condition in microbial populations in intensively studied microbial communities, such as the human gut where a diverse range of reference genomes is available (White *et al.*, 2009; Sczyrba *et al.*, 2017; Knight *et al.*, 2018). The analysis of more diverse environments, like soil, is often limited by the amount reference genomes and gene annotations. This, in turn, might explain why functional traits are often consistent across different samples and environments (Quince *et al.*, 2017) as it was found here at the several of the localities. Another limiting factor while profiling the metabolic potential of a microbial community is the lack of annotations for accessory genes in most microbial species. Thus, it was expected that mainly highly conserved pathways would to be detected and quantified in the metagenomes of the present study also linked to the relatively shallow sequencing depth obtained. Total abundance of genes involved in N cycling respond to increasing temperature derived from transplantation (Zhao *et al.*, 2014). Here, a selection of N cycle-related genes were unaffected by transplantation (assessed as KO terms in Figure 4). Previous research has shown that broadly distributed functions such as respiration, overall carbon catabolism and biomass production often seem more resistant to taxonomic changes than narrow functions such as the degradation of specific compounds (Louca *et al.*, 2018). Also, soil microbial functional profiles showed greater resilience than taxonomic ones in response to chemical contamination (Jiao *et al.*, 2019). This previously observed resilience pattern in soil functionality, and the fact that the metabolic potential of the communities assessed in this study corresponds primarily to broad functions, may explain the apparent overall lack of an effect of the applied treatments.

While the majority of studies exploring the impact of elevation or warming on microbial communities has focused on taxonomic changes, this study has explicitly determined the influence of transplantation on potential microbial functional through the genes comprising the soil microbial metagenome. Although the KEGG annotation terms observed were broad, they can be associated with diverse environment-specific adaptations. For example, when studying cold adaptation of *Pseudoalteromonas* strains, Mocali *et al.*, (2017) found that features linked to this adaptation involved multiple metabolic pathways including “Glutathione metabolism”, “Arginine and Proline metabolism”, “Fatty acid biosynthesis” and “Biosynthesis of amino acids” among others. Similarly, when studying cold resistance of the food contaminant *Vibrio parahaemolyticus,* Xie *et al.*, (2019) reported that differentially expressed genes were commonly enriched in pathways such as “Carbon metabolism”, “Pyruvate metabolism”, “ATP-binding cassette (ABC) transporter” and “Biosynthesis of amino acids”. In the present case, the pathways ko00480 Glutathione metabolism and ko02010 ABC transporters were found to be more abundant in HH (high elevation sites) when compared to LL (low elevation sites) (Table S5). However, experimental validation is needed to directly link the differentially abundant features (genes or KOs) found in the present study to specific adaptations of the microbial communities following the transplantation treatments. Although a meta-transcriptomic approach is also desirable to determine whether these genes were being actively transcribed, recent studies have shown that gene abundance generally correlates to transcript abundance (Salazar *et al.*, 2019) and to specific ecosystem processes (Zhao *et al.*, 2014).

The inclusion of microbial gene abundances of key metabolic processes into climate models will likely improve accuracy predictions of climate change effects on microbial communities. The large geographical scale used in this study coupled with plant community data and soil biochemistry measurements here provide valuable insights improving confidence in global biogeochemical modelling.

References

Alexander JM, Diez JM, Levine JM. 2015. Novel competitors shape species' responses to climate change. Nature 525: 515–518.

Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, et al. 2018. Structure and function of the global topsoil microbiome. Nature 560: 233–237.

Blanchet FG, Legendre P, Borcard D. 2008. Forward selection of explanatory variables. Ecology 89: 2623–2632.

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution 24: 127–135.

Borcard D, Gillet F, Legendre P. 2011. Numerical Ecology with R. ISBN 978-1-4419-7975-9

Bryant JA, Lamanna C, Morlon H, Kerkhoff AJ, Enquist BJ, Green JL. 2009. Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. In the Light of Evolution 2: 127–148.
Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. Nature Methods 12: 59–60.

Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, et al. 2019. Scientists' warning to humanity: microorganisms and climate change. Nature Reviews Microbiology.

Crowther TW, van den Hoogen J, Wan J, Mayes MA, Keiser AD, Mo L, Averill C, Maynard DS. 2019. The global soil community and its influence on biogeochemistry. Science 365.

DeAngelis KM, Pold G, Topçuoglu BD, van Diepen LTA, Varney RM, Blanchard JL, Melillo J, Frey SD. 2015. Long-term forest soil warming alters microbial communities in temperate forest soils. Frontiers in Microbiology 6: 1–13.

Donhauser J, Frey B. 2018. Alpine soil microbial ecology in a changing world. FEMS Microbiology Ecology 94: 1–31.

Fierer N, McCain CM, Meir P, Zimmermann M, Rapp JM, Silman MR, Knight R. 2011. Microbes do not follow the elevational diversity patterns of plants and animals. Ecology 92: 797–804.

Finlay BJ. 2002. Global dispersal of free-living microbial eukaryote species. Science 296: 1061–1063.

Frindte K, Pape R, Werner K, Löffler J, Knief C. 2019. Temperature and soil moisture control microbial community composition in an arctic–alpine ecosystem along elevational and micro-topographic gradients. ISME Journal: 2031–2043.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics 28: 3150–3152.

Gao Y, Wang S, Xu D, Yu H, Wu L, Lin Q, Hu Y, Li X, He Z, Deng Y, et al. 2014. GeoChip as a metagenomics tool to analyze the microbial gene diversity along an elevation gradient. Genomics Data 2: 132–134.

Hayden HL, Mele PM, Bougoure DS, Allan CY, Norng S, Piceno YM, Brodie EL, Desantis TZ, Andersen GL, Williams AL, et al. 2012. Changes in the microbial community structure of bacteria, archaea and fungi in response to elevated CO2 and warming in an Australian native grassland soil. Environmental Microbiology 14: 3081–3096.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. 2016. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research 44: D286–D293.

Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11.

Jiao S, Chen W, Wei G. 2019. Resilience and assemblage of soil microbiome in response to chemical contamination combined with plant growth (I Cann, Ed.). Applied and Environmental Microbiology 85: 1–16.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44: D457–D462.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. Scientific Data 4: 1–20.

Klironomos JN. 2002. Feedback with soil biota contributes to plant rarity and invasiveness in communities. Nature 417: 67–70.

Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D, et al. 2018. Best practices for analysing microbiomes. Nature Reviews Microbiology 16: 1–13.

Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31: 1674–1676.

Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF, Doebeli M. 2017. High taxonomic variability despite stable functional structure across microbial communities. Nature Ecology & Evolution 1: 1–12.

Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, Ackermann M, Hahn AS, Srivastava DS, Crowe SA, et al. 2018. Function and functional redundancy in microbial systems. Nature Ecology & Evolution 2: 936–943.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 1–21.

Luo X, Fu X, Yang Y, Cai P, Peng S, Chen W, Huang Q. 2016. Microbial communities play important roles in modulating paddy soil fertility. Scientific Reports 6.

Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: Evaluation of metagenome assemblies. Bioinformatics 32: 1088–1090.

Mocali S, Chiellini C, Fabiani A, Decuzzi S, Pascale D, Parrilli E, Tutino ML, Perrin E, Bosi E, Fondi M, et al. 2017. Ecology of cold environments: New insights of bacterial metabolic adaptation through an integrated genomic-phenomic approach. Scientific Reports 7: 1–13.

Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P, et al. 2019. Vegan : Community Ecology Package. R package version 2.5-6.

Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 35: 501–504.

van der Putten WH, Bradford MA, Pernilla Brinkman E, van de Voorde TFJ, Veen GF. 2016. Where, when and how plant–soil feedback matters in a changing world. Functional Ecology 30: 1109–1121.

Qi Q, Zhao M, Wang S, Ma X, Wang Y, Gao Y, Lin Q, Li X, Gu B, Li G, et al. 2017. The biogeographic pattern of microbial functional genes along an altitudinal gradient of the Tibetan pasture. Frontiers in Microbiology 8: 1–12.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology 35: 833–844.

R Core Team. 2019. R: A Language and Environment for Statistical Computing.

Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh H, Cuenca M, Field CM, Coelho LP, Cruaud C, Engelen S, et al. 2019. Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. Cell 179: 1068-1083.e21.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863–864.

Schmieder R, Lim YW, Rohwer F, Edwards R. 2010. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. BMC bioinformatics 11: 341.

Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, et al. 2017. Critical Assessment of Metagenome Interpretation - A benchmark of metagenomics software. Nature Methods 14: 1063–1071.

Sheik CS, Beasley WH, Elshahed MS, Zhou X, Luo Y, Krumholz LR. 2011. Effect of warming and drought on grassland microbial communities. ISME Journal 5: 1692–1700.

Shen C, Shi Y, Ni Y, Deng Y, Van Nostrand JD, He Z, Zhou J, Chu H. 2016. Dramatic increases of soil microbial functional gene diversity at the treeline ecotone of changbai mountain. Frontiers in Microbiology 7: 1–12.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, et al. 2015. Structure and function of the global ocean microbiome. Science 348: 1261359.

Tatusov RL. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research 28: 33–36.

de Vries FT, Griffiths RI. 2018. Impacts of Climate Change on Soil Microbial Communities and Their Functioning. In: 111–129.

Wagg C, Bender SF, Widmer F, van der Heijden MGA. 2014. Soil biodiversity and soil community composition determine ecosystem multifunctionality. Proceedings of the National Academy of Sciences of the United States of America 111: 5266–70.

Wagg C, Jansa J, Stadler M, Schmid B, Van Der Heijden MGAA. 2011. Mycorrhizal fungal identity and diversity relaxes plant-plant competition. Ecology 92: 1303–1313.

Weber CF, Zak DR, Hungate BA, Jackson RB, Vilgalys R, Evans RD, Schadt CW, Megonigal JP, Kuske CR. 2011. Responses of soil cellulolytic fungal communities to elevated atmospheric CO 2 are complex and variable across five ecosystems. Environmental Microbiology 13: 2778–2793.

White JR, Nagarajan N, Pop M. 2009. Statistical Methods for Detecting Differentially Abundant Features in Clinical Metagenomic Samples. PLoS Computational Biology 5.

Xie T, Pang R, Wu Q, Zhang J, Lei T, Li Y, Wang J, Ding Y, Chen M, Bai J. 2019. Cold tolerance regulated by the pyruvate metabolism in vibrio parahaemolyticus. Frontiers in Microbiology 10: 1–11.

Yang Y, Gao Y, Wang S, Xu D, Yu H, Wu L, Lin Q, Hu Y, Li X, He Z, et al. 2014. The microbial gene diversity along an elevation gradient of the Tibetan grassland. ISME Journal 8: 430–440.

Yang Y, Halbritter AH, Klanderud K, Telford RJ, Wang G, Vandvik V. 2018. Transplants, open top chambers (OTCS) and gradient studies ask different questions in climate change effects studies. Frontiers in Plant Science 871: 1–9.

Yue H, Wang M, Wang S, Gilbert JA, Sun X, Wu L, Lin Q, Hu Y, Li X, He Z, et al. 2015. The microbe-mediated mechanisms affecting topsoil carbon stock in Tibetan grasslands. ISME Journal 9: 2012–2020.

Zhang Y, Lu Z, Liu S, Yang Y, He Z, Ren Z, Zhou J, Li D. 2013. Geochip-based analysis of microbial communities in alpine meadow soils in the Qinghai-Tibetan plateau. BMC Microbiology 13.

Zhao M, Xue K, Wang F, Liu S, Bai S, Sun B, Zhou J, Yang Y. 2014. Microbial mediation of biogeochemical cycles revealed by simulation of global changes with soil transplant and cropping. ISME Journal 8: 2045–2055.

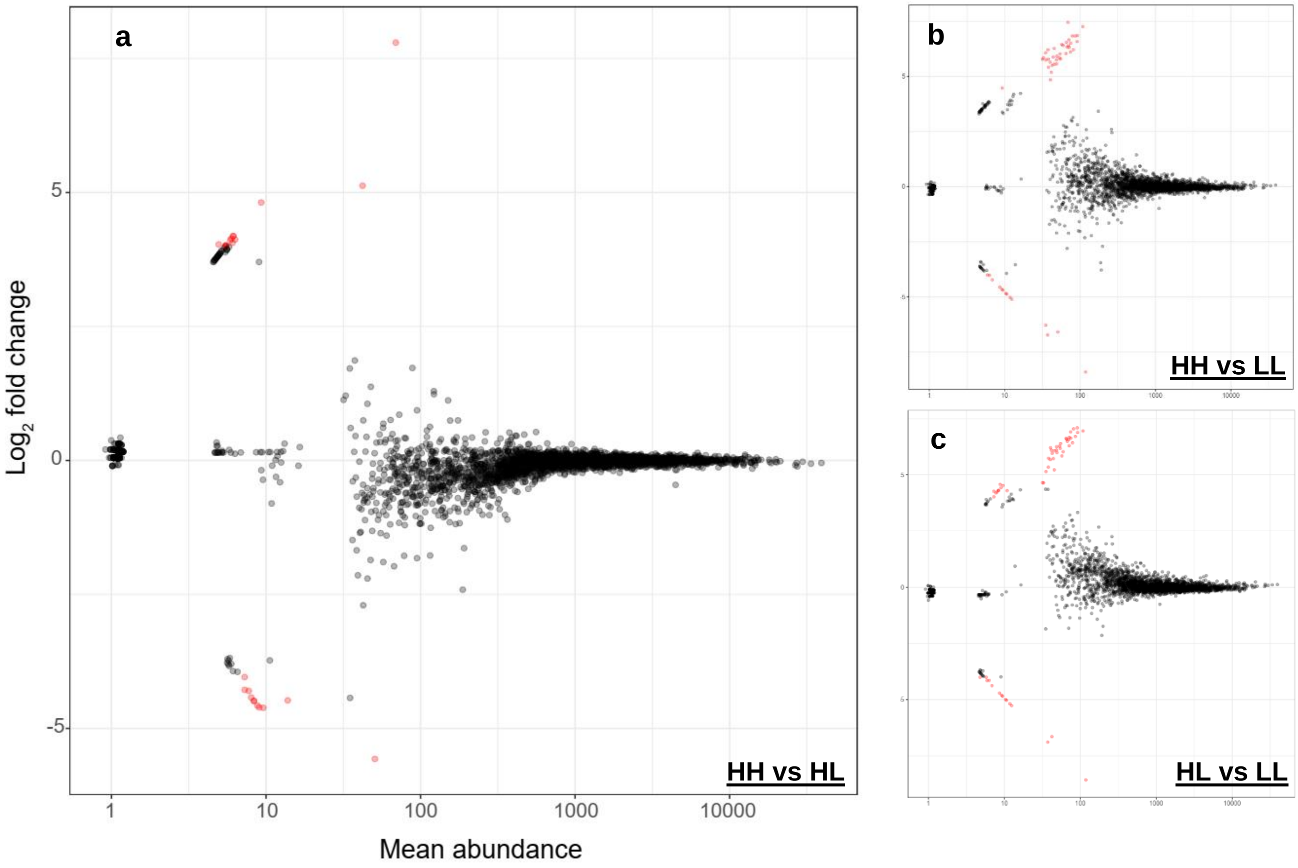


Figure S1. Log$_2$ fold change and mean abundance of the KEGG orthology (KO) count matrix. Each panel shows a different comparison between treatments (bottom right label). Red dots show genes with a significant (*P* < 0.01) log$_2$ fold change greater than 4. The axis labels in **a)** apply to all the graphs.

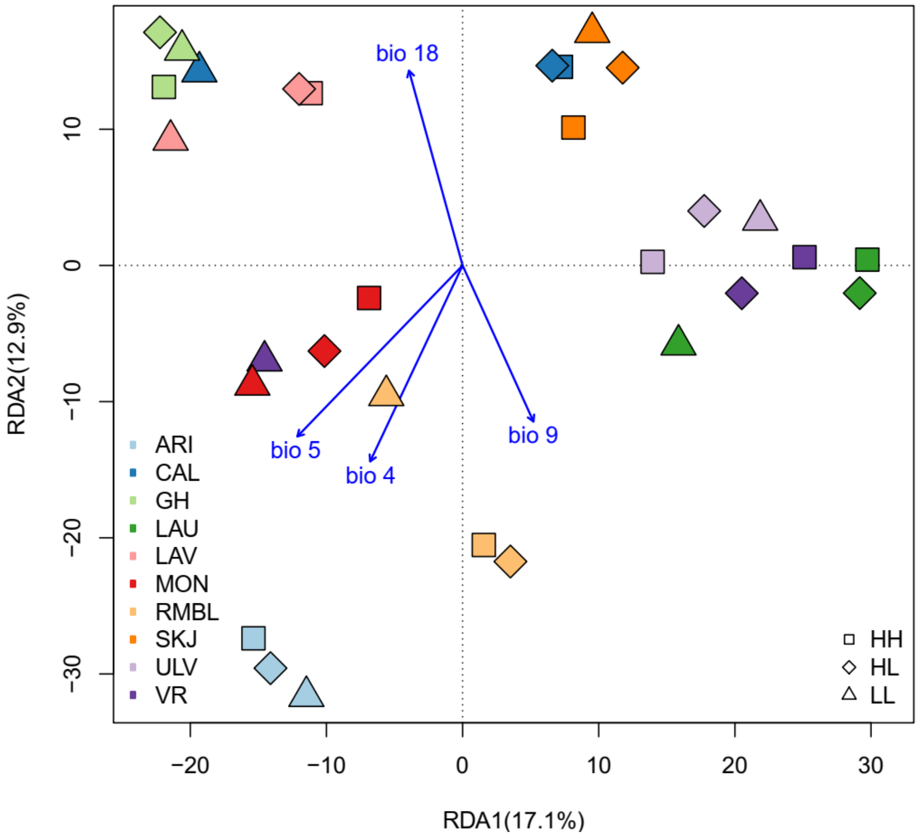


Figure S2. RDA of the gene catalog and the CHELSA climate data as explanatory variables after forward selection. The loadings presented are those of the retained variables. Regions are represented by the different colors and treatments within one region by shape.

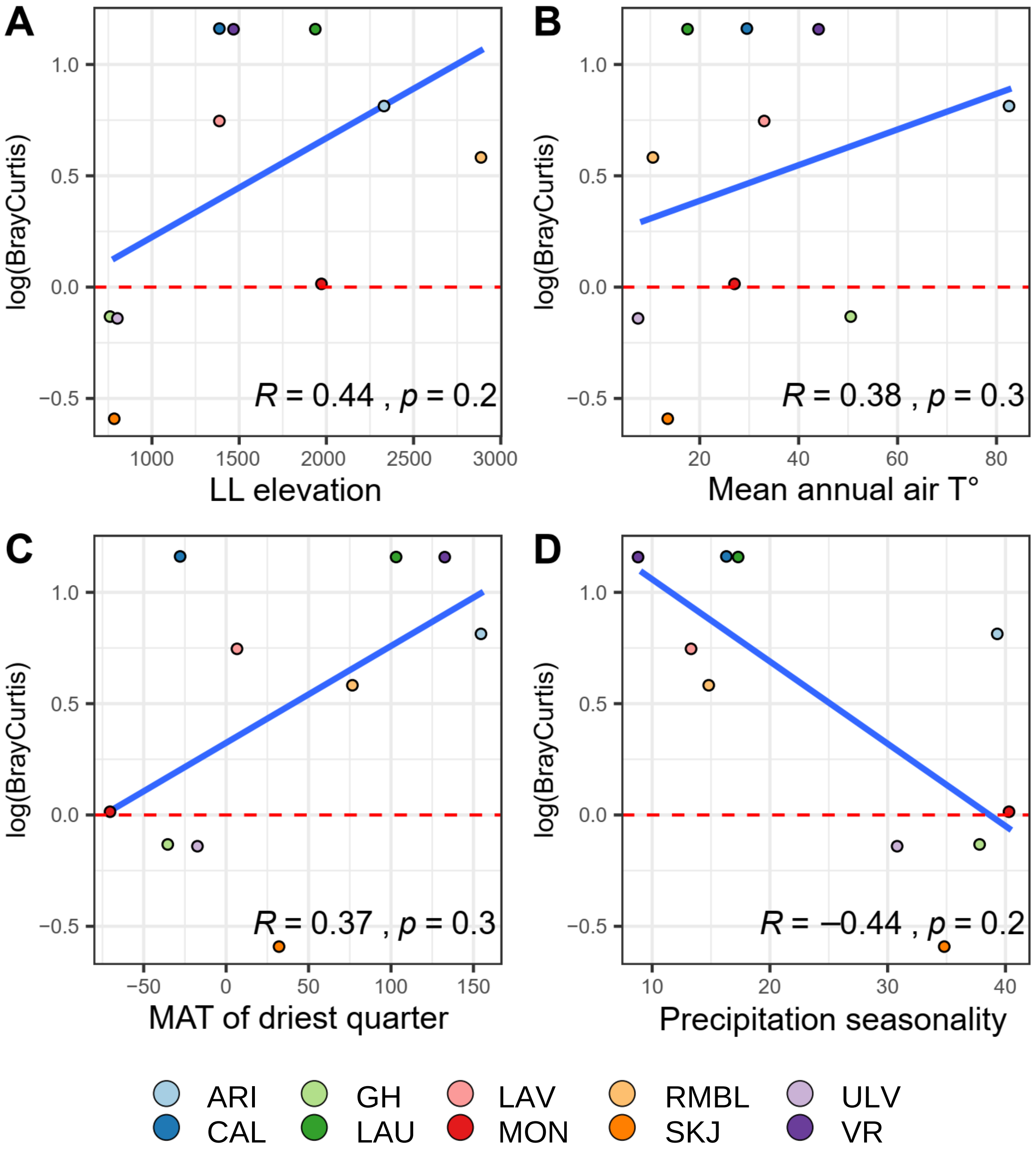


Figure S3. Bray Curtis based ‘closeness’ index and its relation with CHELSA climatic variables. Spearman correlation was calculated and the plots show the variables for which ρ (denoted as R) was higher than 0.35. The red dashed line shows the threshold in which a sample is closer in its gene composition to either the high site (above 0) or the low site (below 0). MAT: Mean air temperature

Supplementary Table 1: Generalities of the different regions sampled. Year range correspond to the time interval between the establishment of the experimental plot and the sampling time.

Locality	Site	Region	Country	Lat	Lon	Habitat	Plot size	Replicates	Elevation	Year range
ARI	high	USA	US	35.35	-111.73	Subalpine meadow	0.3 x 0.3 m	8	2620	16
	low			35.42	-111.67	Forest meadow			2344	16
CAL	high	Alps	CH	46.89	9.49	Calcareous grassland	2 x 2 m	8	2000	6
	low			46.87	9.49	Subalpine pasture			1400	6
GH	high	Alps	DE	47.44	11.06	Subalpine meadow	0.5 x 0.5 m	8	1714	4
	low			47.48	11.01	Meadow			773	4
LAU	high	Alps	FR	45.05	6.40	Alpine grassland	2 x 2 m	8	2450	1
	low			45.04	6.42	Subalpine grassland			1950	1
LAV	high	Alps	CH	46.20	7.06	Alpine grassland	1 x 1 m	8	2200	2
	low			46.22	7.04	Forest pasture			1400	2
MON	high	USA	US	45.31	-111.50	Alpine grassland	0.5 x 0.5 m	8	2185	2
	low			45.31	-111.50	Grassland			1985	2
RMBL	high	USA	US	38.97	-107.05	Subalpine meadow	0.5 x 0.5 m	8	3300	1
	low			38.93	-107.01	Meadow			2900	1
SKJ	high	Scandinavia	NO	60.93	6.42	Calcareous grassland	0.25 x 0.25 m	5	1088	8
	low			60.54	6.51	Calcareous grassland			797	8
ULV	high	Scandinavia	NO	61.02	8.12	Calcareous grassland	0.25 x 0.25 m	5	1208	8
	low			60.82	8.70	Calcareous grassland			815	8
VR	high	Alps	FR	45.10	6.06	Alpine grassland	0.5 x 0.5 m	8	2072	4
	low			45.09	6.04	Grassland			1481	4

Supplementary Table 2: Climatic data extracted from the CHELSA database based on GPS coordinates. Temperature related variables are expressed in C° times 10 and precipitation in kg m^{-2}. **bio1:** mean annual air temperature; **bio2:** mean diurnal air temperature range; **bio3:** isothermality; **bio4:** temperature seasonality; **bio5:** mean daily maximum air temperature of the warmest month; **bio6:** mean daily minimum air temperature of the coldest month; **bio7:** annual range of air temperature; **bio8:** mean daily mean air temperatures of the wettest quarter; **bio9:** mean daily mean air temperatures of the driest quarter **bio10:** mean daily mean air temperatures of the warmest quarter; **bio11:** mean daily mean air temperatures of the coldest quarter; **bio12:** annual precipitation amount; **bio13:** precipitation amount of the wettest month; **bio14:** precipitation amount of the driest month; **bio15:** precipitation seasonality; **bio16:** mean monthly precipitation amount of the wettest quarter; **bio17:** mean monthly precipitation amount of the driest quarter; **bio18:** mean monthly precipitation amount of the warmest quarter; **bio19:** mean monthly precipitation amount of the warmest quarter.

Locality	Site	bio 1	bio 2	bio 3	bio 4	bio 5	bio 6	bio 7	bio 8	bio 9	bio 10	bio 11	bio 12	bio 13	bio 14	bio 15	bio 16	bio 17	bio 18	bio 19
ARI	high	73	124	358	7831	256	-89	345	179	145	179	-32	589	86	14	40	251	50	251	154
	low	93	123	352	8029	279	-72	351	202	167	202	-15	350	49	9	39	143	30	99	107
CAL	high	16	79	300	6383	155	-108	263	106	-40	106	-68	1151	124	73	17	361	220	361	241
	low	44	79	296	6490	184	-82	266	135	-13	135	-42	1010	103	65	16	301	197	301	219
GH	high	23	83	307	6513	164	-108	272	108	-61	113	-64	1377	209	67	46	619	204	587	215
	low	79	83	301	6705	222	-54	276	167	-7	171	-10	963	121	56	30	359	172	355	186
LAU	high	3	83	304	6438	150	-122	272	-23	89	95	-80	912	98	50	18	292	168	172	251
	low	33	83	301	6547	181	-94	275	6	120	126	-52	799	84	47	17	251	156	159	230
LAV	high	7	81	303	6388	151	-116	267	98	-19	98	-76	1448	151	106	12	441	319	441	351
	low	60	81	298	6577	206	-66	272	153	35	153	-26	1236	134	89	15	394	269	394	290
MON	high	24	110	309	8238	227	-129	355	89	-72	147	-84	415	64	18	40	190	59	100	69
	low	31	110	307	8297	235	-122	357	96	-66	155	-78	394	61	17	41	182	55	94	63
RMBL	high	0	121	344	7677	189	-163	352	-59	66	110	-101	747	76	38	15	221	131	174	207
	low	22	121	340	7809	213	-144	357	-37	90	133	-81	491	49	25	15	142	85	123	138
SKJ	high	4	55	236	6263	133	-99	232	-59	24	92	-74	1727	217	69	37	634	220	330	593
	low	24	55	235	6348	155	-81	236	-40	43	113	-55	2083	253	84	33	727	259	434	574
ULV	high	-3	57	222	7005	141	-117	258	95	-26	96	-92	538	63	24	28	185	82	163	127
	low	19	59	225	7073	166	-95	261	118	-6	119	-70	690	89	30	34	253	96	250	141
VR	high	34	82	302	6444	181	-90	271	8	126	126	-49	1369	132	93	9	389	301	301	357
	low	55	82	300	6516	203	-70	273	28	142	148	-29	1209	117	85	9	350	273	273	307

Supplementary Table 3: Quality control and annotation metrics of the raw data and the assemblies, respectively. Blue highlights the two samples from France with an above-average number of contigs obtained after the assembly.

		Quality control			Assembly and annotation		
Sample		**Input reads**	**After QC**	**%Good**	**Total contigs**	**Predicted ORF**	**Annotated NR**
ARI	HH	40264876	36999786	91.9%	318851	603197	519915
	HL	37304808	34177023	91.6%			
	LL	33942554	31276449	92.1%			
CAL	HH	26301535	23892448	90.8%	382282	743911	643322
	HL	32425599	29883873	92.2%			
	LL	26765139	24767291	92.5%			
LAV	HH	32646342	30335746	92.9%	314551	589783	518796
	HL	29801175	27509761	92.3%			
	LL	28706813	26719839	93.1%			
RMBL	HH	35639831	32895605	92.3%	323778	620246	540107
	HL	38060119	34945001	91.8%			
	LL	32300752	29537676	91.4%			
GH	HH	24486037	22447599	91.7%	280605	522960	445897
	HL	29567742	27781428	94.0%			
	LL	22166868	20731204	93.5%			
VR	HH	46482708	44131794	94.9%	645588	1325601	1116308
	HL	41849017	38273982	91.5%			
	LL	38028592	35483236	93.3%			
MON	HH	45249004	42606931	94.2%	292426	529936	468603
	HL	34342814	31228450	90.9%			
	LL	35856616	32793806	91.5%			
LAU	HH	35464930	33251712	93.8%	650550	1421510	1211362
	HL	30931612	28704441	92.8%			
	LL	43902544	40749159	92.8%			
SKJ	HH	27217188	25394734	93.3%	307690	564012	482636
	HL	35904614	33257318	92.6%			
	LL	27536338	25940272	94.2%			
ULV	HH	26872576	25152804	93.6%	386880	742562	642108
	HL	31711546	29439649	92.8%			
	LL	29093083	26815461	92.2%			

Supplementary Table 4: Summary of metrics of the assemblies as reported by metaQUAST.

	Locality									
Metric	**ARI**	**CAL**	**LAV**	**RMBL**	**GH**	**VR**	**MON**	**LAU**	**SKJ**	**ULV**
Input reads	102,453,258	78,543,612	84,565,346	97,378,282	70,960,231	117,889,012	106,629,187	102,705,312	84,592,324	81,407,914
Total contigs	318,851	382,282	314,551	323,778	280,605	645,588	292,426	650,550	307,690	386,880
contigs ≥ 750 bp	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
contigs ≥ 1000 bp	46.02%	53.98%	48.61%	48.87%	48.05%	55.00%	45.59%	58.68%	48.98%	52.52%
contigs ≥ 1500 bp	15.38%	21.37%	16.96%	17.77%	16.44%	23.62%	14.79%	27.42%	17.16%	19.60%
contigs ≥ 3000 bp	2.44%	3.45%	2.41%	2.86%	2.50%	5.10%	1.98%	6.98%	2.25%	2.68%
contigs ≥ 5000 bp	0.71%	0.75%	0.49%	0.72%	0.67%	1.52%	0.46%	2.45%	0.45%	0.50%
contigs ≥ 10000 bp	0.14%	0.07%	0.04%	0.13%	0.12%	0.23%	0.06%	0.53%	0.05%	0.08%
contigs ≥ 25000 bp	0.01%	0.00%	0.00%	0.01%	0.01%	0.01%	0.00%	0.05%	0.00%	0.01%
Largest contig bp	40634	71369	64258	87543	63582	119577	29801	128328	62575	66902
Total length bp	382,049,989	493,899,280	379,103,548	400,837,387	340,674,456	891,913,110	341,853,485	991,650,128	370,681,863	484,279,066
N50	1132	1271	1157	1184	1159	1370	1111	1562	1161	1224
N75	889	943	901	907	900	966	884	1025	903	928
L50	105029	120408	104875	103731	92085	184452	99830	164847	103062	126033
L50	201250	234472	198604	201546	176348	381270	186948	364487	194480	240831

Supplementary Table 5. Top 3 differentially abundant KO per comparison and correspondent KEGG pathway assignation. Positive $\log_2$ fold change indicates higher abundance of the KO in the first term compared to the other (e.g. K01678 was more abundant in HH compared to HL).

Comparison	KO term	Mean abundance	Log_2 fold change	Adjusted p-value	KEGG pathway
HH vs HL	K01678	42.17	5.13	4.00E-03	ko00020 Citrate cycle (TCA cycle) ko00620 Pyruvate metabolism ko00720 Carbon fixation pathways in prokaryotes ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites ko01120 Microbial metabolism in diverse environments ko01130 Biosynthesis of antibiotics ko01200 Carbon metabolism
	K15314	13.81	-4.48	4.84E-03	ko01059 Biosynthesis of enediyne antibiotics ko01100 Metabolic pathways ko01130 Biosynthesis of antibiotics
	K00895	8.35	-4.49	9.15E-04	ko00010 Glycolysis / Gluconeogenesis ko00030 Pentose phosphate pathway ko00051 Fructose and mannose metabolism ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites ko01120 Microbial metabolism in diverse environments ko01130 Biosynthesis of antibiotics
HL vs LL	K00957	86.00	6.70	2.20E-05	ko00230 Purine metabolism ko00261 Monobactam biosynthesis ko00450 Seleno compound metabolism ko00920 Sulfur metabolism ko01100 Metabolic pathways ko01120 Microbial metabolism in diverse environments ko01130 Biosynthesis of antibiotics
	K21166	81.38	7.04	4.88E-05	ko01059 Biosynthesis of enediyne antibiotics ko01100 Metabolic pathways ko01130 Biosynthesis of antibiotics
	K13593	74.96	6.63	3.20E-05	ko04112 Cell cycle - Caulobacter
HH vs LL	K07232	71.56	6.32	1.65E-04	ko00480 Glutathione metabolism ko01100 Metabolic pathways
	K03430	65.88	6.55	2.22E-04	ko00440 Phosphonate and phosphinate metabolism ko01100 Metabolic pathways ko01120 Microbial metabolism in diverse environments
	K15598	64.91	6.04	3.51E-04	ko02010 ABC transporters

Chapter 2: Inoculation of cassava with arbuscular mycorrhizal fungi alters gene diversity and gene composition of the soil bacterial metagenome without altering its taxonomic structure and composition

Cristian Rincón[1,3], Diego Peña-Quemba[2,3], Alia Rodríguez[2] and Ian R. Sanders[1*].

[1]Department of Ecology and Evolution; University of Lausanne; Lausanne, Switzerland. [2]Department of Biology; Universidad Nacional Colombia; Bogotá, Colombia. [3] These authors contributed equally.

Abstract

Soil microbial communities are among the most complex and diverse environments on the planet. They play key roles in biogeochemical cycles which are essential to terrestrial ecosystems. Seeking benefits from microbial communities to improve crop productivity, has increased over the last years, via the use of microbial inoculants, as such as arbuscular mycorrhizal fungi (AMF). However, these inoculations have been done without full understanding of the ecological impacts of such practices. In a field experiment using the globally important crop cassava, a shotgun metagenomic approach (which offers information both about taxonomic and genomic composition of the microbial community) was used to study the impact of inoculation with two isolates of *Rhizophagus irregularis.* We characterized the effects of inoculation on alpha and beta diversity of taxa and genes in soil microbial communities. It was found that the gene composition of the soil microbial metagenome was altered by AMF inoculation even though changes in the taxonomic composition were undetectable. As very little information currently exists on how inoculation of crops by AMF, or other beneficial microbes, influences the pre-existing soil bacterial microbiome, this study offers important insights into this topic in agriculture. The possibility of modifying the soil metagenome to obtain positive responses (e.g. plant yield) opens the door to research that will allow the harnessing of ecological services provided by soil microbial communities towards a more sustainable agriculture.

Introduction

Soil microbial communities are considered the most diverse and complex ecosystems on the planet (Bardgett & Van Der Putten, 2014) and play a key role in ecological processes. These include biogeochemical cycling of nutrients, which in turn is fundamental to the functioning of terrestrial ecosystems. In the past few years, advances in sequencing technologies and bioinformatic tools have allowed a better understanding of the structure and diversity of soil microbial communities and improved the understanding of their biogeography and ecology (Chu *et al.*, 2020). However, this knowledge has mostly focused on the microbiome of temperate soils and often using a descriptive approach. There are fewer investigations adopting an experimental framework where responses of the soil microbiome to experimental manipulations of the environment have been studied. The lack of knowledge of tropical soil microbiomes hinders our ability to harness their ecological services towards more sustainable agriculture. This gains particular relevance in the tropics as their inherently low fertility soils are experiencing increasing anthropogenic pressure and the effects of climate change (Jiao *et al.*, 2019).

In an attempt to develop more sustainable agriculture in the tropics the use of microbial inoculants, such as arbuscular mycorrhizal fungi (AMF), has increased. Results of field experiments with the security crop cassava (a clonally reproduced plant producing a starchy tuberous root) have shown higher plant yield as an effect of AMF inoculation (Ceballos *et al.*, 2013). More recent field experiments conducted in Colombia, Kenya and Tanzania have shown that inoculation with genetically different AMF isolates of the fungus *Rhizophagus irregularis* generated differential growth responses in cassava (Ceballos *et al.*, 2019). The large differences in cassava productivity led to a higher allocation of carbohydrates to the roots, and this could have direct consequences on the surrounding microbial community. However, the introduction of AMF inoculants to cassava, and to crops in general, has been done with little knowledge about the potential impact of such practices on the soil microbiome, the structure of the microbial community, its diversity or potential functional changes to the soil metagenome.

Several studies have addressed the effects of AMF inoculation on pre-existing local AMF communities and have reported contrasting results. Adding an AMF inoculant in a greenhouse experiment was shown to have no impact on the structure of the resident AMF community T-RFLP fingerprint (Antunes *et al.*, 2009). Other studies found that inoculation reduced the abundance and diversity of resident AMF communities (Koch *et al.*, 2011; Symanczik *et al.*, 2015; Thioye *et al.*, 2019). There have been fewer studies on the effect of inoculation with AMF on other non-AMF components of the soil microbial community. In an amplicon sequencing study it was recently reported that AMF inoculation significantly influenced rhizospheric community structure, particularly bacterial taxa (Akyol *et al.*, 2019). Notably, amplicon sequencing of a marker gene (i.e. rRNA genes) with the currently used sequencing technologies typically has a resolution that is limited to genus level at best and functional information is, thus, based on knowledge of metabolic capabilities of given families and genera of bacteria (Knight *et al.*, 2018). An alternative, and potentially more informative approach, is to analyze the gene composition of the metagenome. This can be used to provide taxonomic information about the microbial community but can also be used to reveal the relative abundance of the genes making up the metagenome, thus, giving insights into its specific metabolic capabilities. Additionally, these data can be interpreted in a qualitative way by reconstructing putative metabolic pathways. Such an approach is particularly interesting for studies of soil bacterial communities because some genes move horizontally on plasmids among bacterial taxa. Because of this, a change in the environment might not necessarily elicit observable changes in the taxonomic composition of the bacterial community but could alter the diversity and abundance of given groups of genes in the microbial metagenome. This, in turn could alter the metabolic and degradation capabilities of the soil microbiome. To date, such an approach has not been used to study the response of the microbial metagenome to inoculation of a crop with AMF and with different AMF genotypes.

In this study, we analyzed the bacterial metagenome in soil taken from around the roots of cassava that was inoculated in the field in Colombia with two genetically different *R. irregularis* isolates or non-inoculated. In this way, we were able to experimentally assess the effect of inoculation with genetically different AMF isolates on the soil bacterial metagenome. The sequences were classified and annotated to create both a taxonomic profile and a catalogue of the genes present in the community. We tested whether the inoculation altered the alpha and beta diversity profiles of the resident microbial taxa. Also, by comparing gene abundance between treatments we identified the potential metabolic pathways affected by inoculation. Soil variable data as respiration and aggregation was used to discern potential factors structuring the soil microbiome or possible effects of microbial community changes on soil properties. This is, to the best of our knowledge, the first work addressing the effect of AMF inoculation on the metagenomic profile of soil microbiomes in field conditions.

The goals of our study were to characterize the potential changes on the soil metagenome due to inoculation with AMF and two AMF genotypes. Studying the potential effects generated by AMF inoculation would provide much-needed knowledge to contribute to the understanding on how to harness ecological services from AMF and soil microbiomes to contribute to plant productivity and agroecosystems sustainability.

Materials and methods

Experimental design

The experiment was set up in the eastern plains of Colombia (Tauramena, Casanare, 4°57'32.14"N, 72°34'23.28"W, 220 masl). Cassava (*Manihot esculenta* Crantz) variety CM 4574 (from CIAT, Cali, Colombia) planted as stem cuttings at a density of 10,000 plants ha^{-1} was inoculated with two genetically different *R. irregularis* strains originally isolated from a field in Tänikon, Switzerland (Koch *et al.*, 2004). The selection of the fungal isolates (called C2 and C3) was made having into account their contrasting effects on cassava yield in a previous field experiment (Ceballos *et al.*, 2019). The fungal inoculum was upscaled in an *in vitro* culture system by Symbiom (Lanskroun, Czech Republic) and mixed with a sterile carrier (calcified diatomite). Both the *in vitro* system and the sterile carrier ensured that only AMF are being inoculated and that no unknown organism was present in the inoculum. Plants were inoculated at planting with one of two fungi with 1g of the inoculum, containing 1,000 *R. irregularis* spores or mock inoculated using 1g of the carrier as a control (hereon referred to as the carrier treatment).The experiment was set up as a randomized block design with 6 blocks. Within each block each inoculated plant was surrounded by 8 uninoculated plants to avoid cross-contamination. Plants were fertilized in total with 100 kg ha^{-1} urea, 100 kg ha^{-1} di-ammonium phosphate, 106 kg ha^{-1} potassium chloride. Fifty percent of this was applied at 43 days after planting date and the other 50% at 61 days after planting date. There was no irrigation and conventional crop management for the region was applied depending on pests, diseases and weed incidence.

Variables measured during the experiment

To determine CO_2 efflux, 20 cm diameter, 11 cm height PVC rings were pressed into the soil 10 cm from the stem of the inoculated plant in each one of the blocks. The rings remained in the same position during the entire study period. Measurements of CO_2 efflux were made with a portable flow measurement chamber equipped with an infrared gas analyzer (IRGA) Licor LI-8100A with a volume of 6 186 cm^3 (Zhao *et al.*, 2018). Four consecutive measurements were made per plant and then averaged. Additionally, at sampling depth three measurements of soil humidity content were taken then averaged with the portable measurement system (ΔT®) and three soil temperature measurements were taken then averaged with a 12 cm soil thermometer (Spectrum®).

Microbial soil respiration was determined using the alkali absorption method (Alef, 1995) in which 25 grams of 2 mm sieved soil were incubated at 25°C for 24h in the dark in tightly sealed containers. Each container had a beaker with 15 ml of a 0.1 M NaOH solution and another one with 15 ml of distilled water to keep moisture in the container. After the incubation time, the beaker with NaOH was removed and 3 ml of a 0.5 M $BaCl_2$ solution were added to precipitate the stored C. The resulting solution was titrated with 0.1 M HCl to give the rate of soil respiration in mg of CO_2 per dry weight in the incubation time (CO_2/g/h).

Total soil water-stable aggregates were determined at two depths (10 cm and 30 cm) using a wet sieving method (Kemper & Rosenau, 1986). Briefly, 100 g of fresh soil were placed on the top of a stack of five sieves with the following screen sizes 6.3, 4, 2, 1, 0.5 mm to obtain the aggregate classes. The sieve stack was shaken up and down for 30 min with the column submerged in water throughout the whole process. The aggregates remaining on each sieve were collected and weighed after being oven-dried. In conjunction with the of aggregation measurements, soil moisture content was determined by drying 50 g of soil overnight at 105°C and calculating the percentage of the soil sample that was water.

Plants were directly weighed in the field separating the above-ground and below-ground biomass. The reported values correspond to fresh weight as it has been shown that fresh and dry weight are highly correlated (Ceballos *et al.*, 2019).

Soil sampling for metagenome sequencing

Soil was sampled from C2 and C3 treatments as well as the mock-inoculated treatment (as a control) in six replicates in different blocks, thus giving 18 samples in total. The sampling was performed by taking soil at three points near the same plant at a depth of 10 cm and pooling it to have a representative sample. The samples were immediately frozen in liquid nitrogen and transferred to -80°C until extraction.

DNA extraction, library preparation and sequencing

Total DNA was extracted from 2 g of each of the 18 pooled soil samples using a PowerSoil total DNA isolation kit (MoBio Laboratories, Inc.) according to the instructions of the manufacturer. Subsequently, libraries were prepared for sequencing with a TruSeq DNA PCR-

Free kit (Illumina) and validated and quantified with KAPA ROX Low qPCR Master Mix (Roche). A Fragment Analyzer was used to assess the quality of the libraries. Libraries were sequenced with Illumina HiSeq 4000 150 base pair (bp) paired-end sequencing.

Data processing

Raw data of the 18 libraries (C2-1 to C2-6, C3-1 to C3-6 and carrier-1 to carrier-6) were initially checked with FastQC v0.11.4 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After library quality check completion adaptor sequences were trimmed using TagCleaner (Schmieder *et al.*, 2010). Reads were then quality-filtered (`min_qual_mean 20`) and trimmed using Prinseq-lite version 0.20.4 (Schmieder & Edwards, 2011). Low quality 3'-ends were trimmed and reads containing uncalled bases (N) removed. Only reads at least 120 bp long were kept for further analyses.

Quality filtered reads were taxonomically assigned using kraken 2 (Salzberg & Wood, 2014) with a reduced version (24GB) of the preformatted bacterial and fungal database available with the sofware. Taxonomic assignation counts were then summarized at the class, order and family level using the R package Pavian (Breitwieser & Salzberg, 2016). This constituted the taxa classification matrix used for the diversity analyses.

To construct the gene abundance matrices, co-assemblies of the metagenomes were performed by treatment (3 co-assemblies each with 6 replicates) using Megahit v1.1.4 (Li *et al.*, 2015) (`--k-list 33,47,63,77,93,117, --min-contig-len 750`). Metrics of the assemblies were assessed with metaQUAST v5.0.2 (Mikheenko *et al.*, 2016). Contigs from the three treatments were then concatenated into a single file and processed with cd-hit v4.6.8 (Fu *et al.*, 2012) to reduce redundancy by clustering reads using a similarity threshold of 95% over at least 90% of the alignment. The resulting file was annotated in parallel by performing open reading frame (ORF) predictions with prodigal v2.6.3 (Hyatt *et al.*, 2010) (using the `-p meta` option). Subsequently, blast-like annotation was made against the NCBI RefSeq non-redundant (NR) sequence database of proteins (Pruitt *et al.*, 2007) and eggNOG v4.5 (Huerta-

Cepas *et al.*, 2016) using diamond v0.9.18 (Buchfink *et al.*, 2014) with an e-value threshold of 0.001.

Quality filtered reads were mapped back to the annotated contigs to obtain a count matrix of the abundance of each contig with BBMap v37.82 (Bushnell B., https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/) using default parameters. These matrices, containing the abundance per sample of each annotated contig were used as input data for the R packages and will be referred to as the gene catalogues.

Statistical analysis

Analyses were conducted in R (R Core Team, 2019). Phyloseq (McMurdie & Holmes, 2013) was used to estimate richness and alpha diversity indices and the non-parametric Kruskal-Wallis test was applied to identify differences between treatments. To visualize the level of dissimilarity (beta diversity) between the samples non-metric dimensional scaling plots (NMDS) of the Bray-Curtis metric were build using Vegan v2.5-5 (Oksanen, 2013). Significant differences in beta diversity among treatments were tested with ANOSIM in Vegan. An extension of Vegan's `bioenv` function, implementing a "BVSTEP" routine (http://menugget.blogspot.com/2011/06/clarke-and-ainsworths-bioenv-and-bvstep.html) was used to overcome the inflexibility of the original function which uses only a similarity matrix based on normalized euclidean distance. This routine was used to search for a combination of environmental variables with the highest correlation to dissimilarities of the community. The result of the routine was then plotted on an NMDS plot. The variables shown in Table S1 were not measured in the carrier treatment, therefore, the relationship between any of these variables and the bacterial metagenome were only analyzed in treatments C2 and C3.

DESeq2 (Love *et al.*, 2014) was used to identify differentially abundant genes between the samples and treatments. The gene catalogs were filtered to only keep entries with more than 10 reads mapped in each of at least 3 of the replicates. Genes were considered to have significantly different abundance if the absolute value of the $\log_2$ fold change was higher than

4 and a FDR Benjamin-Hochberg adjusted p-value less than 0.001. Finally, to contextualize gene abundance patterns with known biochemical processes pathway enrichment analysis (PEA) was performed using GAGE (Luo *et al.*, 2009) and visualization with Pathview (Luo & Brouwer, 2013). KEGG orthologs (KO) (Kanehisa *et al.*, 2016) obtained from the eggNOG annotation were compiled into a KO catalogue. This was catalogue was subsequently filtered to only keep entries with more than 10 reads mapped in each of at least 5 of the replicates and subsequently used to perform the PEA. Significant pathways were considered differentially abundant at a cutoff q-value of 0.1.

Results

Soil variables

A Wilcoxon rank sum test was applied to the measured soil variables of the AMF inoculated samples and no statistically significant differences were found between treatments (Table S1). Measurements taken during soil sampling (temperature and humidity) and pH showed minimal variation of sampling conditions. Moisture at different depths indicated similar environmental conditions needed for accurate respiration measurements comparisons. Plant biomass, total aggregation at both depths and respiration variables showed no differences between AMF isolates.

Sequencing data characteristics

A total of 409 million reads were obtained with an average of 22 million reads per sample. Approximately 95% of the reads across all the samples were considered of high enough quality after filtering and trimming (Table S2). Only approximately 7 million reads were obtained from the two samples, C3-4 and C2-3. However, given that the quality threshold parameters were passed, these two samples were kept for the analyses despite the relatively low number of reads.

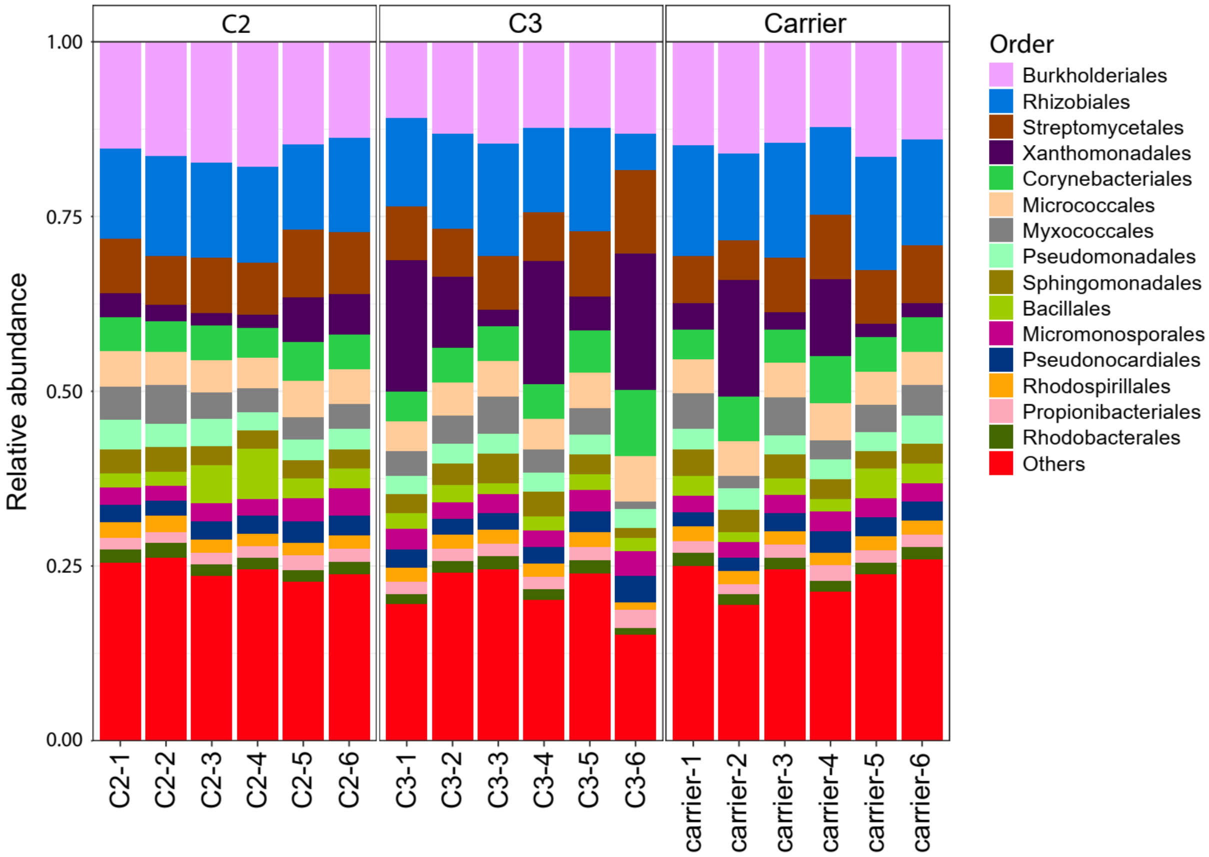


Figure 1. Relative abundance of the 20 most abundant bacterial orders obtained using the kraken classification algorithm. The remaining orders of lower abundance were grouped into the category called “others”.

On average, 23.2% of the reads were successfully assigned to taxa using kraken. Out of the classified reads, on average 97.7% of them were tagged as of bacterial origin and the remaining 2.3% as of fungal origin (Table S2). The most abundant 20 orders, as determined by the kraken classification algorithm, occurred in all samples (Figure 1) and a Kruskal-Wallis rank sum test revealed that no order differed significantly in abundance among treatments following a family-wise error rate correction at 0.05 (data not shown).

After co-assembly, 1 608 366, 1 039 875 and 478 561 contigs were obtained from the reads of treatments C2, C3 and carrier, respectively. Metrics calculated by MetaQUAST showed a similar distribution of contig length in C2 and C3 and slightly lower in carrier. The total number of nucleotides integrated in the assembly showed a similar trend with the carrier treatment exhibiting a lower total base length than in treatments C2 and C3 (Table S3).

There were 2 581 287, 1 784 819 and 682 736 predicted ORFs in the C2, C3 and carrier treatments, respectively. Of these, 3 947 533 (78.18%) were annotated using the NR protein database. The annotation to the KEGG orthology showed 436 488 ORFs annotated in the C2 treatment (27.1%), 341 433 in the C3 treatment (32.8%) and 87 160 in carrier treatment (18.2%). Gene catalogs were constructed by compiling all the annotated genes and used in the subsequent analyses performed.

Taxa and gene diversity

No significant differences were found in richness and alpha diversity among inoculation treatments when the indices were based on taxa counts at the species level (Figure 2a). In contrast, observed and Chao1 richness of genes, as well as and Shannon and Fisher alpha diversity indices of gene counts differed significantly among the inoculation treatments (Figure 2b). The Simpson alpha diversity index showed no differences. Richness indices were significantly higher in the C2 treatment than the carrier treatment. Richness and diversity of genes in C3 tended to be higher than the carrier treatment. Richness and diversity of genes between the treatments with the two genetically different *R. irregularis* isolates, C2 presented higher values than C3.

There were no discernable differences in taxon beta diversity among treatments when taxa counts were used to calculate the Bray-Curtis dissimilarity matrix (ANOSIM R: -0.053, 9 999 permutations, *p* = 0.7127) (Figure 3a). Highly significant differences in gene beta diversity among treatments were revealed when Bray-Curtis dissimilarities were calculated using gene counts (ANOSIM R: 0.814, 9 999 permutations, *p*= 1e-04) (Figure 3b). Beta diversity of genes among all treatments were clearly separated and significantly different (Figure 3b).

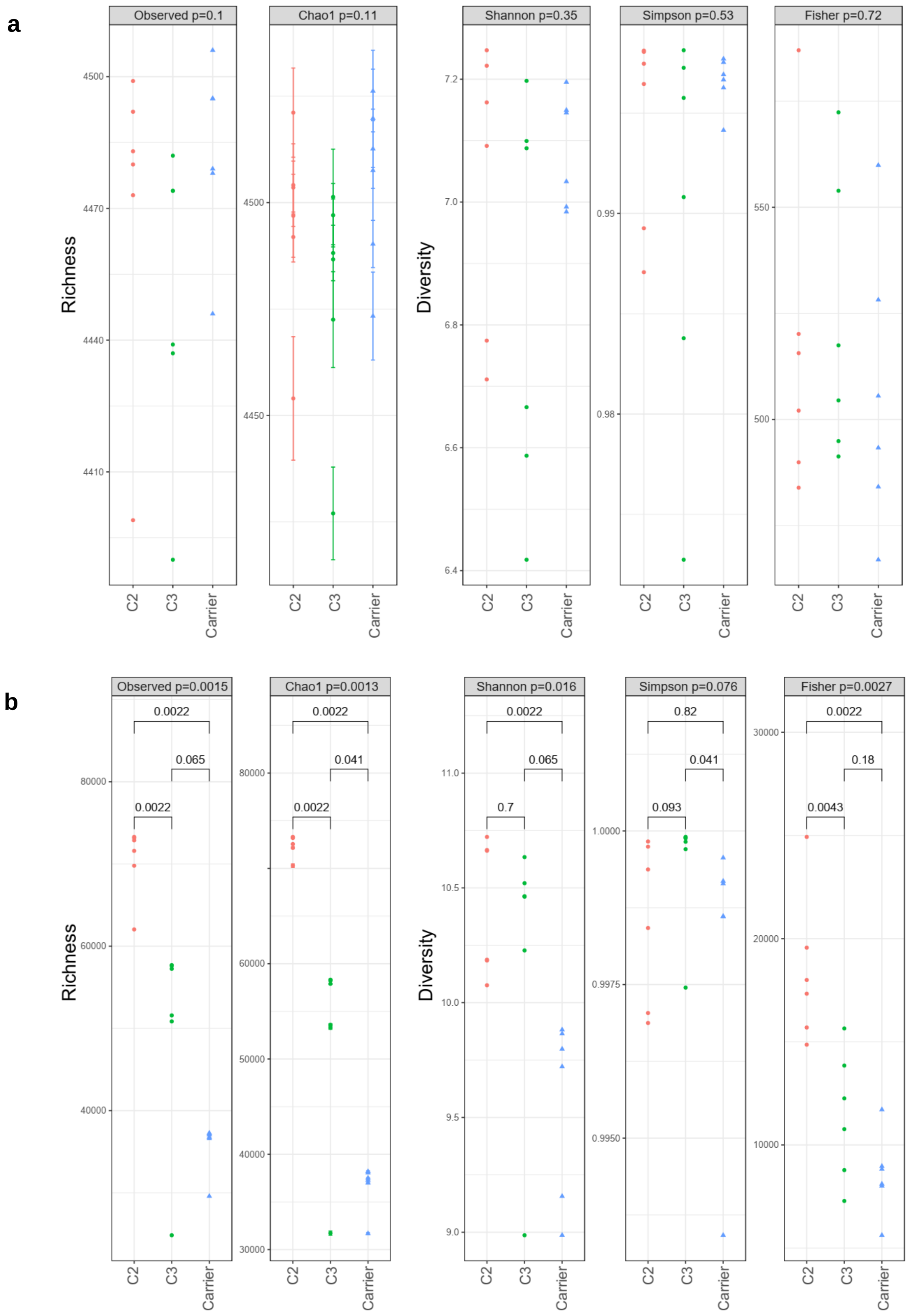


Figure 2. Estimates of richness and alpha diversity using taxon-based metrics at ‘species’ level and (a) and gene count metrics (b). The p values within gray boxes show the non-parametric Kruskal-Wallis test comparing all treatments. The pairwise comparisons between treatments were made with the non-parametric Wilcoxon test. The absence of error bars for the gene count metrics Chao1 index is due to the presence of multiple zero-filled columns.

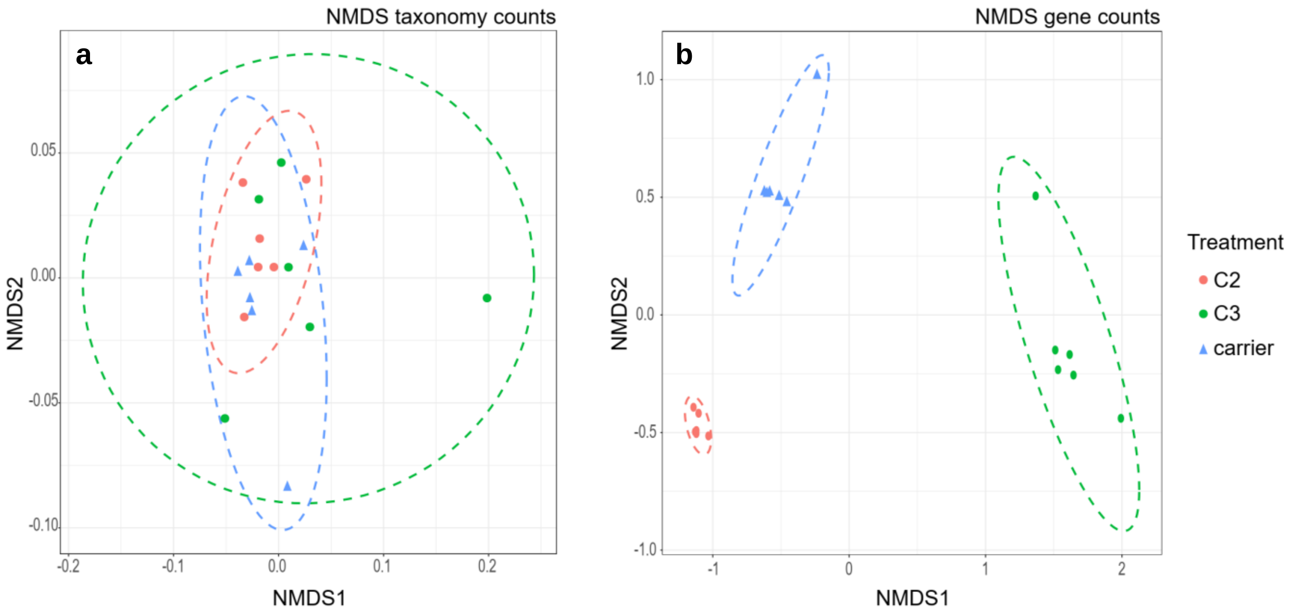


Figure 3. Non-metric multi-dimensional scaling (NMDS) of the Bray-Curtis dissimilarity matrix describing beta diversity. Results from the taxa classification matrix are shown in panel **a**, from the gene count matrix in panel **b**. Each point represents a replicate (n=18). Ellipses represent 95% confidence intervals around centroids and the colors denote the 3 different inoculation treatments.

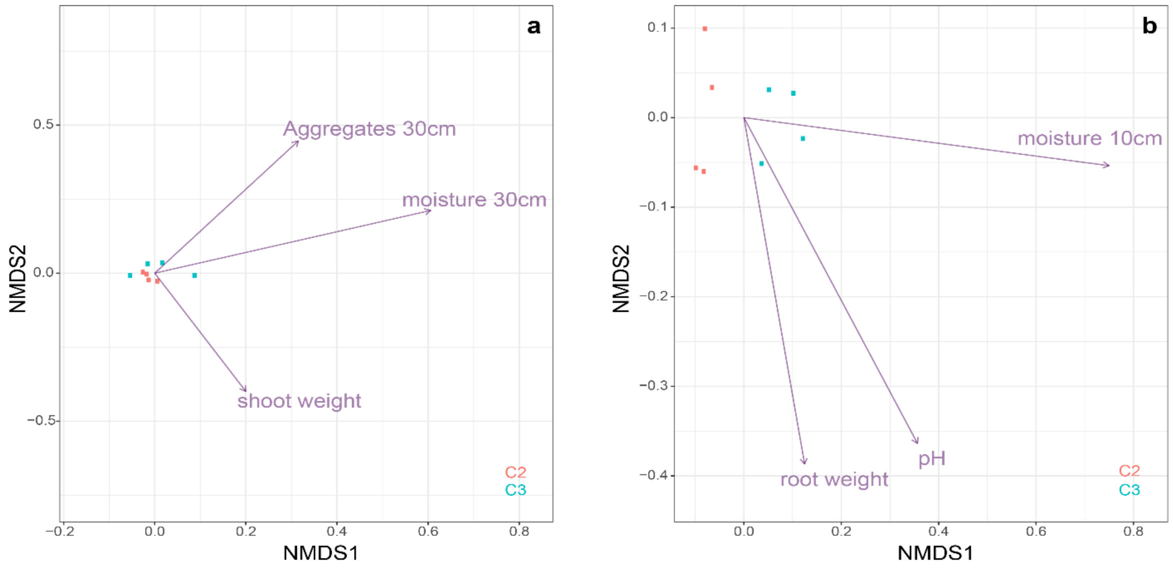


Figure 4. Non-metric multi-dimensional scaling (NMDS) of the Bray-Curtis distance matrix depicting the combination of environmental variables with highest rank correlation to the dissimilarities among samples. Results from the taxa classification matrix are shown in (a) and from the gene count matrix in (b). Aggregates 30cm: total soil aggregation at 30 cm depth; moisture 30cm: moisture percentage at 30 cm depth; shoot weight: cassava shoot fresh weight; moisture 10cm: moisture percentage at 10 cm depth; root weight: cassava root fresh weight.

BVSTEP analyses showed that beta diversity estimated using taxa counts was correlated with above ground plant biomass (“shoot weight”) the total soil aggregation at 30 cm depth (“Aggregates 30cm”) and moisture at 30 cm depth (“moisture 30cm”) (Figure 4a). On the other hand, the beta diversity matrix calculated using the gene counts was correlated with below ground plant biomass (root weight) pH and moisture at 10 cm depth (moisture 10cm) (Figure 4b).

Differential gene abundance and metabolic pathways

A contrast test showed that 55 118 genes of the bacterial metagenome were differentially abundant between the C2 and C3 treatments, 15 826 genes were differentially abundant between the C2 and carrier treatments, 35 224 genes were differentially abundant between the C3 and carrier treatments and 6 511 genes differed in abundance when comparing the two AMF inoculated treatments versus the carrier treatment (Figure S1). All these comparisons were made using an annotated gene catalog of 142 065 entries.

The PEA showed that several metabolic pathways were differentially represented among treatments. Twelve significantly gene sets exhibited higher abundance and 63 gene sets exhibited significantly lower abundance in the AMF inoculated treatments compared to the carrier treatments (Figure 5). Two pathways were significantly higher and 5 pathways were significantly lower in abundance in the C3 treatment compared to the C2 treatment. Thirty-six pathways were in lower abundance and 6 were higher in abundance in C3 compared to the carrier treatment. Finally, there were 31 pathways revealing lower abundance and 11 pathways revealing higher abundance in the C2 treatment compared to the carrier treatment. The greater abundance pathways for the inoculated treatments compared to the carrier were related to the degradation of aromatic compounds. The lower abundance pathways for the inoculated treatments compared to the carrier were related to more general metabolic processes as amino acids or protein synthesis. Table S4 shows the top 5 entries generated in the PEA for the above-mentioned pair-wise comparisons.

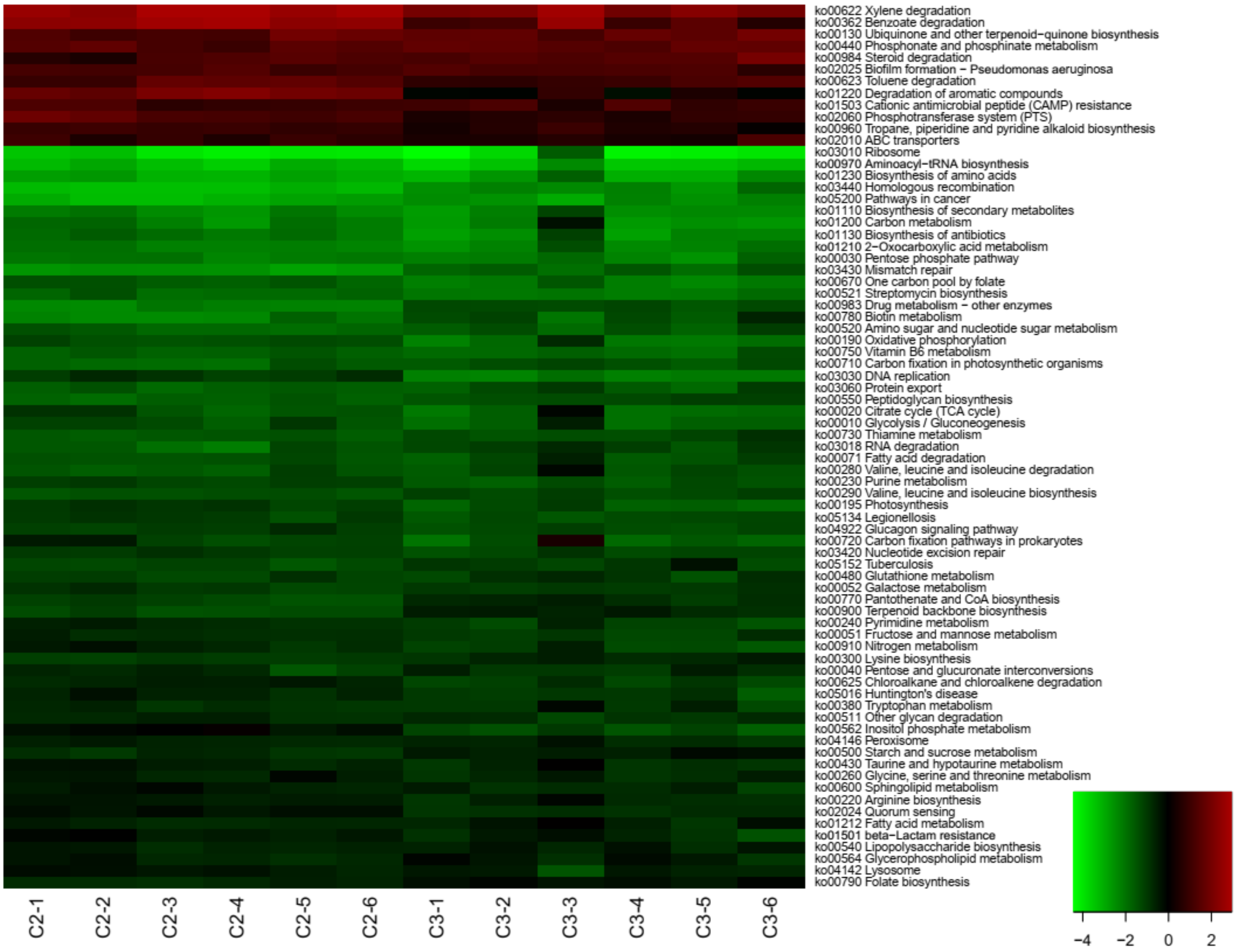


Figure 5. Heatmap plotting Generally Applicable Gene-set Enrichment (GAGE) for pathway analysis. The analysis shows the AMF inoculation treatments compared to the carrier. Significant KEGG pathways up (red) or down (green) regulated at a q value of 0.1 are shown There are 12 significantly up-regulated gene sets and green 63 significantly down-regulated gene sets. Color shade indicates the $\log_2$ fold change according to the color key at the bottom right corner of the figure.

Discussion

Very little information currently exists on how inoculation of crops by AMF, or other beneficial microbes, influences the pre-existing soil bacterial microbiome. In this study, we demonstrated that inoculation of cassava with the AMF *Rhizophagus irregularis* mostly increased gene richness and gene alpha diversity of the microbial metagenome compared to the mock-inoculated treatment while there was not an observable shift in the taxonomic richness and alpha diversity of the microbiome. In addition, genetically different *R. irregularis* isolates also induced significant changes in gene richness, gene alpha and beta diversity of the microbial metagenome even though these two fungi did not alter the observable taxonomic richness, alpha and beta diversity of the microbiome. Biologically, this means that the gene composition of the soil microbial metagenome was altered by inoculation with genetically different AM fungi even though taxonomic differences in the composition of the microbial community were undetectable. All the effects of AMF inoculation on the gene diversity and composition of the microbial metagenome occurred in the absence of an effect of inoculation on plant growth or any of the other variables measured in the soil.

There are remarkably few experimental studies looking at the effects of AMF inoculation on the soil microbiome. Recently, the results of a study using 16S and 18S rRNA amplicon data revealed that AM fungal inoculation significantly influenced the root microbial community structure, changing the abundance of indigenous AMF and other soil taxa, including bacteria (Akyol *et al.*, 2019). In the present study, we mostly captured the bacterial metagenome and, to date, there are no other such published studies on effects of AMF inoculation on the bacterial soil metagenome with which we can directly compare. It is worth noting that in this study the inocula were prepared in a sterile *in-vitro* culture system and put into a sterilized carrier (calcified diatomite), that is free of unknown microorganisms. This means that the addition of AMF inoculum, or the carrier, has not resulted in adding any other microorganisms and thus, the observed effects were AMF mediated. This is important because many AMF inocula are produced in non-sterile conditions and contain other unknown microorganisms which would make it difficult to be sure the effects are due to AMF and not due to other microbes.

Due to the results of previous field experiments, conducted in Colombia, where cassava was inoculated with several AMF isolates (Ceballos *et al.*, 2019), we expected that the two genetically different *R. irregularis* isolates would induce differential growth responses of cassava roots. Because of this, we expected that the alteration in cassava root growth, induced by different AMF isolates, would alter the soil environment experienced by soil bacteria and that this could then potentially alter the soil bacterial community. It is, therefore, particularly surprising that even though we did not observe differential effects of AMF isolates on cassava root biomass, or any of the other variable measured in the soil, there was a clear effect of the fungi on the gene diversity and gene composition of the microbial metagenome. This could potentially be due to direct effects of the introduced fungus on the bacterial community. For example, it is already known that these two isolates exhibit differences in their quantitative growth traits (Koch *et al.*, 2004), such as extraradical hyphal growth and sporulation that could influence the bacteria living around them. However, the AMF isolate effects on the microbial metagenome could also still be due to other indirect effects of inoculation that we did not measure, such as effects on the pre-existing local AMF community, or some effects on the plants such as alteration of root exudation etc. that could have then indirectly affected the soil bacterial community.

Experimental studies on factors that affect the soil bacterial metagenomes are still in their infancy. One of the most exciting results of this study is that changes in gene richness, diversity and composition of the microbial metagenome were observed without associated changes in the taxonomic diversity and composition of the microbial community. For most ecologists working with plants and animals, rather than bacteria, such an observation appears unintuitive or impossible. However, we have two explanations for this result. The first is that this is a real biological effect that could be due to horizontal gene transfer among bacterial taxa in response to a change of environment (in this case inoculation with AMF). Approximately, 55000 genes of the bacterial metagenome were differentially abundant in the C2 versus C3 inoculation treatments. Given the total number of genes recorded in the soil bacterial metagenome in this study, the number of genes that were differentially abundant among treatments represents a small enough proportion of the total gene set to possibly represent genes that were mobile among taxa within the bacterial community. The second possibility is that, although generally recognized as a more robust method (Manichanh *et al.*, 2008; Shakya *et al.*, 2013; Campanaro *et al.*, 2018), the resolution of the taxonomic data obtained using the metagenomic approach might still not suffice to observe changes in taxonomic diversity of highly diverse and complex bacterial communities as those from soil. It is difficult to compare our data with those of studies on soil bacterial diversity that have employed a metabarcoding approach as the two methods suffer from different limitations and also by lack of knowledge of the true ecological diversity. However, we consider this as a potentially very important result in ecology as almost all ecological studies on relationships between soil microbial diversity, plant and animal diversity and ecosystem processes focus on taxonomic bacterial diversity. Such taxonomic measurements of diversity may well mask more meaningful functional differences in the microbial metagenomic metabolic capabilities. In plants and animals where extensive horizontal gene transfer among taxa is much less prevalent or likely, measurements of taxonomic diversity could be more closely coupled with functional diversity. Louca *et al*. (2018) reported how several studies showed that certain metabolic functions are strongly coupled to particular environmental factors rather than to the taxonomic composition of the sample at a given time.

A relatively large number of genes were found to be differentially abundant when comparing the C2 and C3 treatments (38.8%) but this only translated into 7 enriched pathways. On the other hand, only 4.5% of the genes found to be differentially abundant (when comparing the AMF treatments together against the carrier) resulted in 75 differentially enriched pathways. Pathway enrichment analysis provided a more informative view (compared to individual gene based interpretation) of the potential functional changes that occurred in the bacterial metagenome as a result of AMF inoculation. It should be kept in mind that the plants and the soil in the carrier treatment still contained an AMF community that existed before inoculation so this difference is not due to the absence of AM fungi in the carrier treatment.

When comparing the AMF treatments together against the carrier, 6 out of the 12 pathways found to be more abundant (namely, ko00622 Xylene degradation, ko00362 Benzoate degradation, ko00984 Steroid degradation, ko00623 Toluene degradation, ko01220 Degradation of aromatic compounds and ko02010 ABC transporters) could be linked to degradation and metabolism of aromatic compounds. Many microbes present in the rhizosphere are able to degrade aromatic compounds and several characteristics of this environment favor this process. Firstly, plant secondary metabolites that are part of the exudates are often structurally very similar to organic contaminants (Correa-García *et al.*, 2018). Additionally, the higher allocation of plant-derived carbon through root exudates promotes more active and more abundant microbial communities. Also, microbial communities are known for horizontal gene transfer (HGT), and it has been shown that plasmids acquired via HGT help microorganisms adapt to contamination stress and degrade organic compounds (Sentchilo *et al.*, 2013). In a microcosm experiment measuring biodegradation of aromatic compounds and functional gene composition, functionally diverse soil microbiomes had higher degrading capabilities than specialized assemblages (Bell *et al.*, 2016). In our study, we found an overall higher gene richness and diversity in the AMF inoculated plots and this, added to the higher carbon allocation typically observed in the rhizosphere may explain the relationship between the AMF inoculation and the upregulation of aromatic compound metabolic pathways. A metabolic model of the soil metagenome proposes the use of aromatic compounds derivates to be used in the TCA cycle for energy metabolism, thus, utilizing those as sources of carbon and energy (Bao *et al.*, 2017). Furthermore, a metaproteomic experiment showed that in phosphorus-rich soils, the microbial communities had higher gene abundances for degradation of aromatic compounds as the community was driven to switch from the acquisition of phosphorus towards the acquisition of carbon, nitrogen and sulfur from more recalcitrant substrates as aromatic compounds (Yao *et al.*, 2018). Altogether, this hints that a potential AMF-inoculation-derived higher availability of phosphorus stimulates the microbial community towards different sources of nutrients, possibly avoiding competition with communities in low-phosphorus niches.

It has been reported that abundance of the inoculated AM fungus was the most significant factor that determined plant yield responses to the inoculation. This suggests that dominance of the inoculated fungus is a necessary condition for positive yield responses (Niwa *et al.*, 2018). Although this was not measured in this study, non-inoculated plants may have had a reduced abundance of hyphal networks. It has been reported that mycorrhizal hyphae act as an ecological niche for highly specialized microbial communities (Jansa *et al.*, 2013) and that

AMF can recruit specific bacterial taxa (Agnolucci *et al.*, 2015; Iffis *et al.*, 2017; Akyol *et al.*, 2019). This may partially explain the observed changes in metabolism of amino acids/vitamins/nucleic acids, antibiotics biosynthesis and quorum sensing that can be interpreted as a modification of the bacterial environment due to a higher abundance of the inoculated AM fungus.

We showed that aggregation was one of the several factors correlated with the microbial community structure. It has been reported that exposure to environmental stresses increase the production of diverse bacterial exopolysaccharides (EPSs) (Sandhya & Ali, 2015; Weathers *et al.*, 2015). Among the enriched pathways when comparing AMF and the non-inoculated plots we found "ko02025 Biofilm formation". Biofilm formation is mediated by EPSs and as their biosynthesis is energetically expensive, this suggest that the community shifts its metagenome, and possibly its functional profile, when carbon availability is not a limitation. Microbial EPSs enhance soil aggregation (Deka *et al.*, 2019) and this defines the physical and mechanical properties of soil, such as water retention, water movement, aeration, and temperature, which in turn affect physical, chemical, and biological processes (Costa *et al.*, 2018). This indicates a positive feedback loop in which improved soil quality may generate a augmentation of plant productivity, which in turn increases the amount of available nutrients in the rhizosphere, driving then the microbial community towards further enhancement of the soil quality via, for example, EPSs-mediated soil aggregation.

The observed differences in alpha and beta diversity of the gene content of the metagenome as well as in the enriched metabolic pathways showed that the use of a small amount of AMF inoculum directly (through recruitment of microbial taxa associated to AMF hyphae) or indirectly (through its effects on the rhizosphere) enhances the versatility of the microbial community. As AMF inoculation can bring economic benefits to the farmer (Ceballos *et al.*, 2013) these inoculants are now being used in many parts of the world without knowledge about the potential ecological impacts of such practices. These impacts include potentially invasive AMF isolates (Schwartz *et al.*, 2006) and the fact that inoculation may directly or indirectly be detrimental to microbial diversity (Koch *et al.*, 2011; Symanczik *et al.*, 2015; Akyol *et al.*, 2019; Thioye *et al.*, 2019). Here we showed that the use of AMF induced changes to the soil metagenome. The possibility of modifying the soil metagenome to obtain positive responses of soil variables and plant yield via AMF inoculation opens the door to research that will allow the harnessing of ecological services provided by soil microbial communities towards a more sustainable agriculture.

References

Agnolucci M, Battini F, Cristani C, Giovannetti M. 2015. Diverse bacterial communities are recruited on spores of different arbuscular mycorrhizal fungal isolates. Biology and Fertility of Soils 51: 379–389.

Akyol TY, Niwa R, Hirakawa H, Maruyama H, Sato T, Suzuki T, Fukunaga A, Sato T, Yoshida S, Tawaraya K, et al. 2019. Impact of introduction of arbuscular mycorrhizal fungi on the root microbial community in agricultural Fields. Microbes and Environments 34: 23–32.

Alef K. 1995. Field methods. In: Methods in Applied Soil Microbiology and Biochemistry. Elsevier, 463–490.

Antunes PM, Koch AM, Dunfield KE, Hart MM, Downing A, Rillig MC, Klironomos JN. 2009. Influence of commercial inoculation with *Glomus intraradices* on the structure and functioning of an AM fungal community from an agricultural site. Plant and Soil 317: 257–266.

Bao YJ, Xu Z, Li Y, Yao Z, Sun J, Song H. 2017. High-throughput metagenomic analysis of petroleum-contaminated soil microbiome reveals the versatility in xenobiotic aromatics metabolism. Journal of Environmental Sciences (China) 56: 25–35.

Bardgett RD, Van Der Putten WH. 2014. Belowground biodiversity and ecosystem functioning. Nature 515: 505–511.

Bell TH, Stefani FOP, Abram K, Champagne J, Yergeau E, Hijri M. 2016. A Diverse Soil Microbiome Degrades More Crude Oil than Specialized. Applied and Environmental Microbiology 82: 5530–5541.

Breitwieser FP, Salzberg SL. 2016. Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. bioRxiv: 2014–2017.

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. Nature Methods 12: 59–60.

Campanaro S, Treu L, Kougias PG, Zhu X, Angelidaki I. 2018. Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. Scientific Reports 8: 1–12.

Ceballos I, Mateus ID, Peña R, Peña-Quemba DC, Masso C, Vanlauwe B, Rodriguez A, Sanders IR. 2019. Using variation in arbuscular mycorrhizal fungi to drive the productivity of the food security crop cassava. : 1–21.

Ceballos I, Ruiz M, Fernández C, Peña R, Rodríguez A, Sanders IR. 2013. The in vitro mass-produced model mycorrhizal fungus, *Rhizophagus irregularis*, significantly increases yields of the globally important food security crop cassava. PLoS ONE 8.

Chu H, Gao G-F, Ma Y, Fan K, Delgado-Baquerizo M. 2020. Soil Microbial Biogeography in a Changing World: Recent Advances and Future Perspectives (A Shade, Ed.). mSystems 5: 1–12.

Correa-García S, Pande P, Séguin A, St-Arnaud M, Yergeau E. 2018. Rhizoremediation of petroleum hydrocarbons: a model system for plant microbiome manipulation. Microbial Biotechnology 11: 819–832.

Costa OYA, Raaijmakers JM, Kuramae EE. 2018. Microbial extracellular polymeric substances: Ecological function and impact on soil aggregation. Frontiers in Microbiology 9: 1–14.

Deka P, Goswami G, Das P, Gautom T, Chowdhury N, Boro RC, Barooah M. 2019. Bacterial exopolysaccharide promotes acid tolerance in *Bacillus amyloliquefaciens* and improves soil aggregation. Molecular Biology Reports 46: 1079–1091.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics 28: 3150–3152.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. 2016. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research 44: D286–D293.

Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11.

Iffis B, St-Arnaud M, Hijri M. 2017. Petroleum contamination and plant identity influence soil and root microbial communities while AMF spores retrieved from the same plants possess markedly different communities. Frontiers in Plant Science 8: 1–16.

Jansa J, Bukovská P, Gryndler M. 2013. Mycorrhizal hyphae as ecological niche for highly specialized hypersymbionts - Or just soil free-riders? Frontiers in Plant Science 4: 1–8.

Jiao S, Chen W, Wei G. 2019. Resilience and assemblage of soil microbiome in response to chemical contamination combined with plant growth (I Cann, Ed.). Applied and Environmental Microbiology 85: 1–16.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44: D457–D462.

Kemper WD, Rosenau RC. 1986. Aggregate Stability and Size Dlstributlon’. Methods of soil analysis: Part 1 Physical and Mineralogical Methods 9: 425–442.

Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D, et al. 2018. Best practices for analysing microbiomes. Nature Reviews Microbiology 16: 1–13.

Koch AM, Antunes PM, Barto EK, Cipollini D, Mummey DL, Klironomos JN. 2011. The effects of arbuscular mycorrhizal (AM) fungal and garlic mustard introductions on native AM fungal diversity. Biological Invasions 13: 1627–1639.

Koch AM, Kuhn G, Fontanillas P, Fumagalli L, Goudet J, Sanders IR. 2004. High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. Proceedings of the National Academy of Sciences 101: 2369–2374.

Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31: 1674–1676.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 1–21.

Luo W, Brouwer C. 2013. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 29: 1830–1831.

Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. 2009. GAGE : generally applicable gene set enrichment for pathway analysis. 17: 1–17.

Manichanh C, Chapple CE, Frangeul L, Gloux K, Guigo R, Dore J. 2008. A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. Nucleic Acids Research 36: 5180–5188.

McMurdie PJ, Holmes S. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE 8.

Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: Evaluation of metagenome assemblies. Bioinformatics 32: 1088–1090.

Niwa R, Koyama T, Sato T, Adachi K, Tawaraya K, Sato S, Hirakawa H, Yoshida S, Ezawa T. 2018. Dissection of niche competition between introduced and indigenous arbuscular mycorrhizal fungi with respect to soybean yield responses. Scientific Reports 8: 2–5.

Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P, et al. 2019. Vegan : Community Ecology Package. R package version 2.5-6.

Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 35: 501–504.

R Core Team. 2019. R: A Language and Environment for Statistical Computing.

Salzberg SL, Wood DE. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology 15.

Sandhya V, Ali SZ. 2015. The production of exopolysaccharide by *Pseudomonas putida* GAP-P45 under various abiotic stress conditions and its role in soil aggregation. Microbiology (Russian Federation) 84: 512–519.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863–864.

Schmieder R, Lim YW, Rohwer F, Edwards R. 2010. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. BMC bioinformatics 11: 341.

Schwartz MW, Hoeksema JD, Gehring CA, Johnson NC, Klironomos JN, Abbott LK, Pringle A. 2006. The promise and the potential consequences of the global transport of mycorrhizal fungal inoculum. Ecology Letters 9: 501–515.

Sentchilo V, Mayer AP, Guy L, Miyazaki R, Tringe SG, Barry K, Malfatti S, Goessmann A, Robinson-Rechavi M, Van Der Meer JR. 2013. Community-wide plasmid gene mobilization and selection. ISME Journal 7: 1173–1186.

Shakya M, Quince C, Campbell JH, Yang ZK, Schadt CW, Podar M. 2013. Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. Environmental Microbiology 15: 1882–1899.

Symanczik S, Courty PE, Boller T, Wiemken A, Al-Yahya'ei MN. 2015. Impact of water regimes on an experimental community of four desert arbuscular mycorrhizal fungal (AMF) species, as affected by the introduction of a non-native AMF species. Mycorrhiza 25: 639–647.

Thioye B, Sanguin H, Kane A, de Faria SM, Fall D, Prin Y, Sanogo D, Ndiaye C, Duponnois R, Sylla SN, et al. 2019. Impact of mycorrhiza-based inoculation strategies on Ziziphus mauritiana Lam. and its native mycorrhizal communities on the route of the Great Green Wall (Senegal). Ecological Engineering 128: 66–76.

Weathers TS, Higgins CP, Sharp JO. 2015. Enhanced biofilm production by a toluene-degrading *Rhodococcus* observed after exposure to perfluoroalkyl acids. Environmental Science and Technology 49: 5458–5466.

Yao Q, Li Z, Song Y, Wright SJ, Guo X, Tringe SG, Tfaily MM, Paša-Tolić L, Hazen TC, Turner BL, et al. 2018. Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil. Nature Ecology and Evolution 2: 499–509.

Zhao S, Chen K, Wu C, Mao Y. 2018. Effects of simulated warming on soil respiration to XiaoPo lake. IOP Conference Series: Earth and Environmental Science 113.

Supplementary Table 1. Mean values of 11 variables measured during the experiment. A Wilcoxon rank sum test was applied and no statistically significant differences were found between C2 and C3.

Sample	Soil temperature (C°)	Soil humidity (%)	pH	Soil moisture (%) 10cm depth	Soil moisture (%) 30cm depth	Shoot weight (g)	Root weight (g)	Microbial respiration (mg CO_2/g/h)	CO_2 efflux (CO_2/m^{-2}/h^{-1})	Aggregates (%) 10cm depth	Aggregates (%) 30cm depth
C2-1	25.1	15.1	5.83	16.66	15.44	1 600	1 600	0.391	3.417	68.29	74.93
C2-2	25.1	14.9	5.14	16.66	18.43	5 940	6 375	0.395	7.048	80.58	72.92
C2-3	25.1	16.7	5.25	16.21	14.53	6 120	2 020	0.976	14.283	73.11	49.80
C2-4	26.1	15.2	5.25	21.11	14.90	5 795	5 465	2.059	3.642	66.41	60.54
C3-1	25.2	18.3	6.09	22.37	16.78	3 620	3 930	0.613	6.429	86.06	88.23
C3-2	25.4	18.9	5.34	22.41	19.02	8 000	8 888	0.201	4.282	71.87	82.95
C3-3	25.4	19.3	5.29	24.64	18.64	1 134	4 655	0.175	4.317	78.90	85.37
C3-4	29.7	13.7	5.65	18.62	16.16	4 600	4 265	0.260	4.436	78.50	67.67
Wilcoxon	0.1804	0.3428	0.1912	0.0571	0.1142	0.6857	0.6857	0.1142	1.000	0.3428	0.1142

Supplementary Table 2. Total read count (number of paired reads) after quality control and summary metrics for the results of the kraken classification. Good (%) refers to the percentage of reads passing the quality filters and which were, thus kept for the analyses.

	Quality control (QC)			Kraken classification		
Sample	**Raw reads**	**After QC**	**Good (%)**	**Classified**	**Bacterial**	**Fungal**
C3-1	22 978 139	22 409 692	97.53%	25.97%	98.37%	1.22%
C3-2	19 143 327	18 656 810	97.46%	23.23%	96.42%	2.52%
C3-3	23 972 462	23 391 092	97.57%	22.37%	98.35%	1.18%
C3-4	6 663 084	6 560 311	98.46%	26.24%	98.09%	1.27%
C3-5	11 882 610	11 392 381	95.87%	21.19%	98.00%	1.24%
C3-6	21 518 885	20 527 883	95.39%	28.79%	99.15%	0.65%
C2-1	21 216 179	20 741 203	97.76%	20.85%	97.57%	1.62%
C2-2	34 337 765	33 524 444	97.63%	20.48%	95.67%	2.28%
C2-3	7 003 900	6 815 313	97.31%	22.71%	97.90%	1.40%
C2-4	36 182 061	35 363 904	97.74%	21.68%	97.54%	1.62%
C2-5	28 399 924	27 689 826	97.50%	19.99%	98.15%	1.40%
C2-6	22 899 433	22 342 430	97.57%	18.69%	98.08%	1.35%
carrier-1	11 308 452	10 841 016	95.87%	21.02%	97.49%	1.69%
carrier-2	32 826 479	30 705 695	93.54%	33.33%	97.66%	1.85%
carrier-3	35 261 123	33 273 514	94.36%	23.17%	97.36%	1.86%
carrier-4	21 793 458	20 074 205	92.11%	25.76%	98.41%	1.12%
carrier-5	31 742 712	29 073 410	91.59%	22.25%	97.18%	1.94%
carrier-6	20 671 151	19 025 884	92.04%	19.45%	97.78%	1.70%

Supplementary Table 3. Assembly metrics as assessed by metaQUAST

	Treatment		
Metric	**C2**	**C3**	**carrier**
Input reads	146 477 120	102 938 169	142 993 724
Total contigs	1 608 366	1 039 875	478 561
contigs ≥ 500 bp	100.00%	100.00%	100.00%
contigs ≥ 750 bp	47.27%	49.10%	31.65%
contigs ≥ 1000 bp	25.24%	28.46%	14.32%
contigs ≥ 1500 bp	10.32%	13.22%	4.82%
contigs ≥ 3000 bp	2.26%	3.37%	0.92%
contigs ≥ 5000 bp	0.78%	1.29%	0.31%
contigs ≥ 10000 bp	0.18%	0.40%	0.06%
contigs ≥ 25000 bp	0.02%	0.07%	0.01%
contigs ≥ 40000 bp	0.01%	0.02%	0.00%
Largest contig (bp)	196 162	188 274	107 848
Total length (bp)	1 568 670 520	1 123 536 338	378 215 956
N50	962	1112	739
N75	680	710	586
L50	441 565	242 289	157 513
L75	933 820	565 999	302 825

Supplementary Table 4. Top 5 upregulated and downregulated (increased and decreased pathway genes abundance) pathways as detected by the GAGE R package and sorted by q-value. Comparison, e.g. “C3 vs carrier” should be read as differentially abundant pathways in C3 considering the level in the carrier treatment as the base level. Pathways in bold were not significant.

Treatment	Change	KEGG orthology and pathway name	q.value
C3 vs carrier	Downregulated	ko03010 Ribosome	5.37E-16
		ko00970 Aminoacyl-tRNA biosynthesis	3.79E-12
		ko01230 Biosynthesis of amino acids	3.55E-09
		ko01130 Biosynthesis of antibiotics	4.16E-07
		ko05200 Pathways in cancer	4.55E-07
	Upregulated	ko00622 Xylene degradation	2.27E-05
		ko00130 Ubiquinone and other terpenoid-quinone biosynthesis	0.002595
		ko00440 Phosphonate and phosphinate metabolism	0.005067
		ko00984 Steroid degradation	0.007027
		ko00362 Benzoate degradation	0.016647
C2 vs carrier	Downregulated	ko03010 Ribosome	2.81E-17
		ko00970 Aminoacyl-tRNA biosynthesis	5.48E-12
		ko03440 Homologous recombination	5.48E-12
		ko01230 Biosynthesis of amino acids	3.41E-11
		ko05200 Pathways in cancer	6.06E-10
	Upregulated	ko00362 Benzoate degradation	3.58E-08
		ko00622 Xylene degradation	3.58E-08
		ko01220 Degradation of aromatic compounds	6.08E-05
		ko00130 Ubiquinone and other terpenoid-quinone biosynthesis	0.008596
		ko00623 Toluene degradation	0.010636
C3 vs C2	Downregulated	ko01220 Degradation of aromatic compounds	1.26E-07
		ko00562 Inositol phosphate metabolism	7.54E-05
		ko00362 Benzoate degradation	0.002528
		ko03030 DNA replication	0.002528
		ko00680 Methane metabolism	0.016245
	Upregulated	ko02040 Flagellar assembly	2.35E-07
		ko00906 Carotenoid biosynthesis	0.074136
		ko00261 Monobactam biosynthesis	***0.269983***
		ko00780 Biotin metabolism	***0.282721***
		ko00984 Steroid degradation	***0.314481***

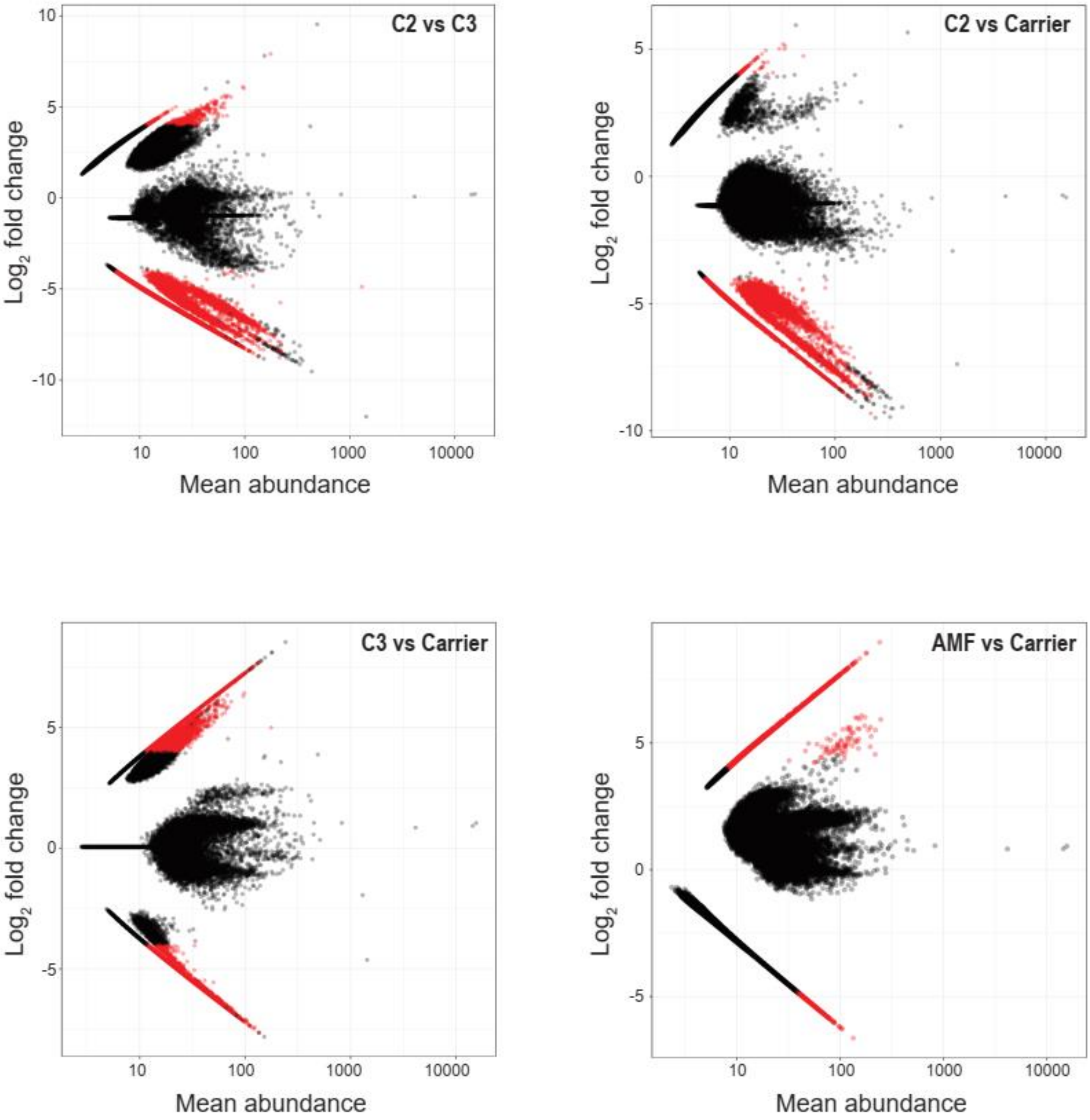


Figure S1. Log$_2$ fold change and mean abundance of the gene counts matrices. Each panel shows a different comparison between two treatments as noted with the top right corner of each panel. Differentially abundant genes (red dots) were denoted as significant when the log$_2$ fold change was over 4 and the adjusted p-value below 0.01.

Chapter 3: Characterizing intraspecific genetic variability of AMF isolates in an attempt to develop strain-specific molecular markers

Cristian Rincón[1],Tania Wyss[1], Frédéric G. Masclaux[1] and Ian R. Sanders[1].

[1]Department of Ecology and Evolution; University of Lausanne; Lausanne, Switzerland.

Abstract

The capacity microbial taxa such as arbuscular mycorrhizal fungi (AMF) to provide ecological services, such as promotion of plant growth and protection against pathogens, has attracted great attention and represent a promising option to help making agriculture more sustainable. It has been shown that inoculation with AMF increases plant yield under field conditions, particularly cassava and cereal crops.

Environmentally safe and economically sustainable use of AMF depends on the understanding of the establishment and persistence of the inoculated AMF isolate. Concerns have been raised about the potential negative effects of AMF inoculations being potentially invasive or detrimental to local AMF diversity. Having a tool to determine the persistence of inocula in the soil and assess its invasiveness will contribute to investigations to better understand and possibly mitigate these processes.

Using ddRAD-seq data small regions with polymorphism were identified. Subsequently, primers were designed to amplify together neighboring small regions into longer sequences containing many more variable sites to first characterize within isolate genetic variation and posteriorly build strain-specific markers.

The selected long regions were amplified and sequenced with ultra-high coverage in isolates used in field experiments. The results (further confirmed by cloning and Sanger sequencing) revealed patterns of within isolate variation inconsistent with current knowledge about the genetics of these fungi. Thus, what began as an approach to develop strain specific markers for tracking AMF isolates, ended up raising additional questions about the genetic organization within these organisms.

Introduction

The capacity of some microbial taxa or microbial communities to provide ecological services, such as promotion of plant growth and protection against pathogens (Berg, 2009), has attracted great attention as some organisms represent a promising option to help making more sustainable agriculture (Backer *et al.*, 2018). One such group of organisms is arbuscular mycorrhizal fungi (AMF). These fungi form one of the commonest plant–microbe mutualisms. The large majority of terrestrial plants, including many important crops, form arbuscular mycorrhizas (van der Heijden *et al.*, 2015). Their main beneficial effect is uptake and transfer of low-mobility minerals (mainly phosphorus) from the soil to plants. It has been shown that inoculation with AMF increases plant yield under field conditions, particularly cassava (Ceballos *et al.*, 2013, 2019) and cereal crops (Zhang *et al.*, 2019).

Environmentally safe and economically sustainable use of AMF depends on the establishment and persistence of the inoculated AMF isolate (Pellegrino *et al.*, 2012). To link plant growth responses to a given inoculum it is necessary to assess its establishment. Also, studying the effects of the inoculated isolate (on, for example, the resident microbial community) requires knowledge of whether and when it establishes and be able to quantify its presence. Being able to evaluate the persistence of an introduced microbial inoculant will also be helpful in order to know if the positive plant growth response following inoculation would then require re-inoculation in future years, and if so, how often. Additionally, as concerns have been raised about the potential negative effects of invasive inoculations worldwide (Schwartz *et al.*, 2006; Hart *et al.*, 2017), having a tool to determine the persistence of inocula in the soil and assess its invasiveness will contribute to investigations to better understand and possibly mitigate these processes. For this, markers need to be developed that can recognize the fungus added as inoculum.

The development of markers for tracking AMF isolates requires considering their life-cycle particularities. Unlike most organisms, there have been no observations of a stage in the AMF cycle in which they develop from a single nucleus (Young, 2015). Also, AMF have no observed sexual cycle, and multiple nuclei share a common cytoplasm. The genetic organization of coexisting nuclei in the Glomeromycota and the within-fungus genetic diversity have been long debated. Recently, the genome organization of AMF was established to be either homokaryotic (coexisting nuclei are genetically similar) (Lin *et al.*, 2014) or dikaryotic (the mycelium harbors two nuclear genotypes) (Ropars *et al.*, 2016).

Recently, data generated in several studies, using different techniques and sequencing platforms, have been used to describe the haploid genome of the model AMF *Rhizophagus irregularis* and its within and among isolate variation. Lin *et al.*, (2014) was the first group to sequence the genome of individual AMF nuclei of the isolate DAOM197198 and reported little variability, thus, suggesting a homokaryotic state. Later, Ropars *et al.* (2016) sequenced the genomes of five *R. irregularis* isolates (A1, A4, A5, B3 and C2) and found heterozygosity in two *R. irregularis* isolates A4 and A5, suggesting the existence of a stable population comprising two dominant divergent haploid nucleus genotypes. However, all studies revealed a level of polymorphism incoherent with a strict homokaryotic or dikaryotic state, although this apparent variation could be due to problems of genome assembly (Masclaux *et al.*, 2019).

Identification of polymorphic sites relies heavily on the quality of the reference genome and depth of coverage and the length of the sequences carrying given polymorphisms that could potentially be used to develop strain specific markers. Thus, considering the importance of tracking inoculated AMF it the field, the goal of this work was to use ultra-high coverage amplicon sequencing to first characterize within isolate genetic variation and posteriorly build strain-specific markers. The approach used in this work comprised of identifying regions that carry polymorphisms on short reads originating from double-digest restriction site-associated sequencing (ddRAD-seq) (Wyss *et al.*, 2016; Savary *et al.*, 2018). The use of ddRAD-seq allows for an inexpensive and fast discovery of thousands of markers in many individuals. Here, we refer to the regions with mapped reads derived from ddRAD-seq as 'RAD loci'. Primers were then designed to amplify genomic regions containing several adjacent RAD loci.

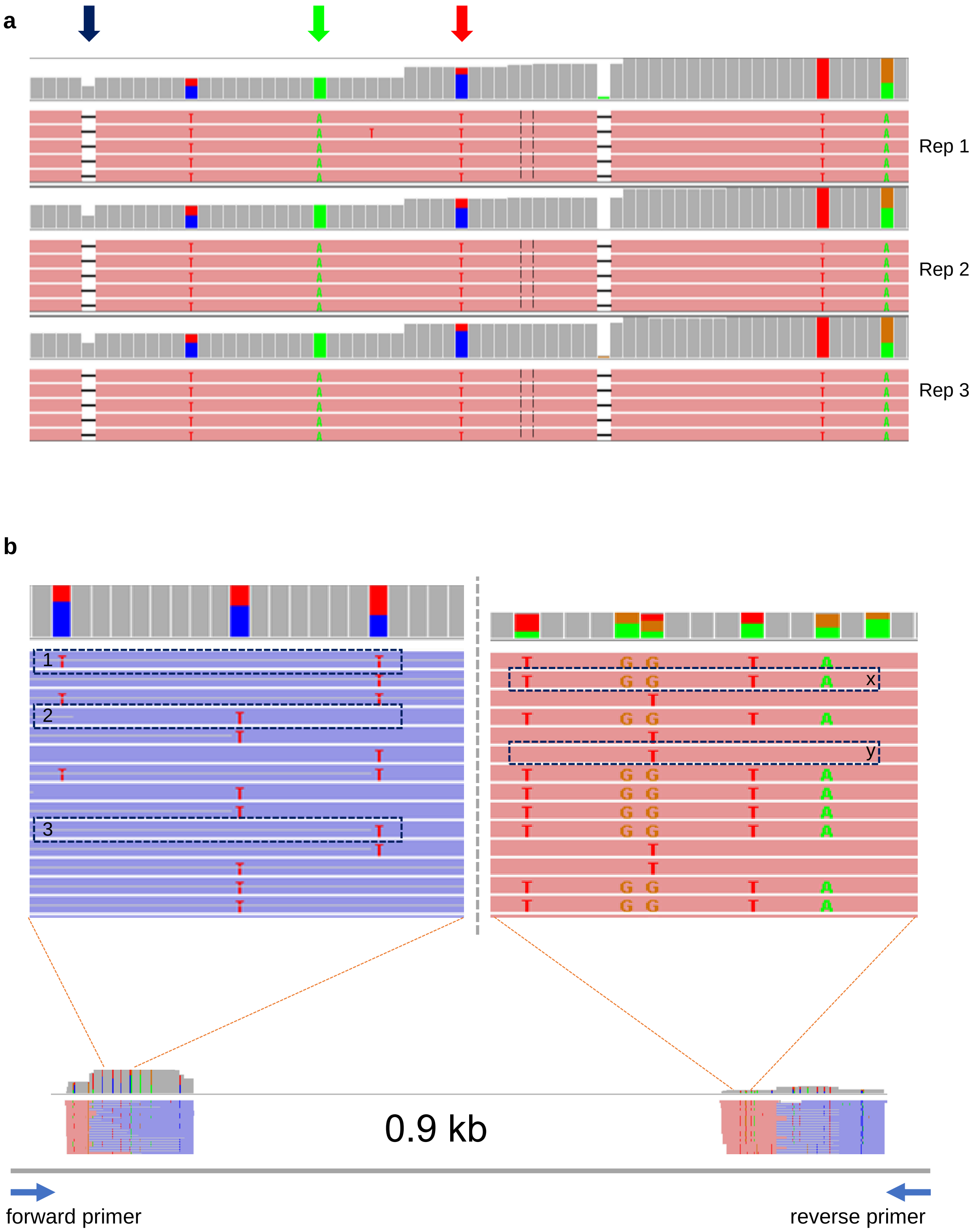


Figure 1. Schematic representation of the methodological approach used. **a)** Data was compared across three biological replicates of ddRADseq data. Pink horizontal bars represent the sequencing reads. Gray vertical bars correspond to nucleotide positions and its height to the coverage (coverage bar). Indels (blue arrow), SNPs (green arrow) and poly-allelic positions with more than one variant (red arrow) can be visualized. The colors in the coverage bar for a poly-allelic position (red arrow) represent which base was called and in which proportion. **b)** When two neighboring loci (< 2kb apart) have reads (pink and blue corresponding to forward and reverse reads in paired-end sequencing) showing poly-alleles, primers were designed to amplify the region

The amplification of these regions, containing several RAD loci, in turn containing multiple polymorphic sites, was used to obtain sequences more likely to be strain specific markers. Secondly, we tried to determine how these variants were organized in the genome. Reads from ddRAD-seq are very short (around 100 bp) and ddRAD-seq does not provide information on how variants in RAD loci are organized. By sequencing a single much longer DNA molecule (Figure 1) we could study the organization of these variants in the genome. Using this methodology we were able to identify haplotypes. We define haplotypes as the combination of poly-allelic sites at adjacent loci in a defined genomic region. Poly-allelic sites refer to those positions in which the difference, when compared to the reference genome, includes more than one variant (Figure 1).

The isolates used in this work are all representatives of the AMF species *R. irregularis*. This fungus is commonly used in field inoculation experiments (Buysens *et al.*, 2017; Akyol *et al.*, 2019; Ceballos *et al.*, 2019; Thioye *et al.*, 2019) and as it has been shown that it has a very wide geographical distribution (Savary *et al.*, 2018). Because of this, there is a high chance that different genotypes of the same species as the inoculated strain are already present in the soil where the experiments were performed. This makes the development of strain-specific markers, rather than just species specific markers, a necessity when trying to track introduced AMF strains in the field.

Materials and methods

Fungal isolates and source of published data

The species isolates used in this study were DAOM197198, C2, A2 and C3. Those isolates were chosen because they have been either used in field experiments (Ceballos *et al.*, 2019) or because they have been characterized using population genomics techniques (Savary *et al.*, 2018). DAOM197198, A2 and C2 have been characterized as being homokaryotic and C3 as heterokaryotic (harboring two nuclear genotypes, i.e. a dikaryon) (Ropars *et al.*, 2016).

Double-digest restriction site-associated sequencing (ddRAD-seq) was then performed on at least three biological replicates of these *R. irregularis* isolates by Wyss *et al*. (2016). The reference genome used was a single nucleus genome assembly of *R. irregularis* (DAOM197198) named N6 (Lin *et al.*, 2014).

Identification of target loci in ddRAD-seq data

The ddRAD-seq datasets were used to generate Variant Call Format (VCF) files that contained information about the sequence polymorphisms (indels and SNPs) of the ddRAD-seq compared to the reference genome. Since misalignment of repeated regions could lead to detection of false within-isolate polymorphism, repeated regions were defined with two complementary approaches as proposed by Wyss *et al.* (2016). First, repetitions were predicted and annotated with RepeatModeler Open-1.0 (Smit and Hubley, 2008) and RepeatMasker Open-3.0 (Smit *et al*., 1996). Second, the reference genome was *in silico* digested to predict fragments that could then potentially be amplified by PCR in the laboratory. Once these predicted fragments (referred to as RAD loci) were obtained they were subjected to pairwise comparisons using fasta-36.3.5e (Pearson & Lipman, 1988) to identify globally similar fragments (potential repeats). Predicted fragments tagged as repeated regions were then excluded from the data. GeneMark-ES (Ter-hovhannisyan *et al.*, 2008) was used to predict coding regions in the reference genome.

The information about the presence of polymorphism, their position in the genome and the “RAD loci” to which they belonged to was extracted from the VCF files using custom bash scripts. The intention was to select a group of regions in the genome that were suitable for PCR amplification and subsequent amplicon sequencing. The criteria to select those regions were: i) the resulting PCR product should be between 300 and 1800 bp; ii) the amplified product does not lie within a repeated region, iii) the variants must be contained in adjacent predicted RAD loci and iv) the variants should be present in all replicates and isolates. The VCF files were parsed to find regions matching the criteria and then those candidate regions were manually verified by visualizing the reference genome together with all the replicates of the ddRAD-seq datasets using IGV (Robinson *et al.*, 2017).

Primer design and amplification testing

Once the regions were manually validated, their sequences were used for primer design with primer3 (Untergasser *et al.*, 2012). Primer design software yields multiple possibilities, then each primer pair was tested *in silico* using the primersearch tool of the EMBOSS package (Rice *et al.*, 2000) allowing up to 10% mismatches to verify a unique priming site across the genome (thus, a single PCR product).

Primers were individually tested for PCR amplification. A range of PCR conditions were tested. The PCR was performed using *Taq* DNA Polymerase (New England Biolabs). The reaction mix contained 2 U μl^{-1} of polymerase, 1x PCR buffer, 250 µM dNTP mix, 25mM of $MgCl_2$ and 0.2 µM of each primer. Cycling conditions were the following: 2 minutes of initial denaturation at 94°C; 25 cycles of 1 minute denaturation at 94°C; 1 minute annealing at 50-55°C (primer set dependent); 1 minute extension at 72°C; and 10 minutes final extension at 72°C. Amplification was verified on 1% agarose gels. Only those primer sets yielding a PCR product with DNA from at least 3 out of the 4 isolates were subsequently used for amplicon sequencing (Table S1).

DNA extraction and PCR amplification

The AMF isolates were maintained in monoxenic plates with *Ri* T-DNA carrot roots which allowed us to have uncontaminated fungal material for DNA extractions. DNA was extracted using the DNeasy Plant Mini Kit (Qiagen) following the instructions of the manufacturer. DNA was eluted in 50 µl of ddH_2O. Quality and quantity of the DNA samples were assessed with a NanoDrop Spectrophotometer (Thermo Fisher Scientific).

PCR was then performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs). The reaction mix contained 0.02 U μl^{-1} of Q5 polymerase (high fidelity enzyme), 1x Q5 PCR buffer, 250 µM dNTP mix and 0.2 µM of each primer. Cycling conditions were the following: 1 minute of initial denaturation at 98°C; 30 cycles of 10 s denaturation at 98°C, 30 s annealing at 55°C; 30 s extension at 72°C; and 2 minutes final extension at 72°C. Amplification was verified on 1% agarose gels.

The forward primer was synthetized with additional nucleotides at the 5' end as a barcode to allow multiplexing of samples in a single library. Those barcodes were generated using a Phyton script (http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator) with the following parameters: 5 as barcode length; a minimum genetic distance between barcodes of 3 nucleotides, 10000 cycles of random attempts and GC content between 0 and 50%.

Library preparation and sequencing

Each library contained the PCR product of all the target loci from each of the four isolates pooled together in equimolar quantities. Two pools of PCR products were prepared into independent libraries and sequenced as technical replicates. Table S1 contains a list of the regions selected for amplicon sequencing, primer sequences and expected PCR product sequences. The libraries were prepared following a standard protocol using the Illumina TruSeq™ Nano Sample Preparation Kit but omitting the DNA fragmentation step. The quality of the libraries was assessed with a Fragment Analyzer and quantified with Qubit DNA assay quantification. The libraries were sequenced on an Illumina MiSeq device (giving up to 300 bp paired-end reads). Quality control was made with FastQC (version 0.10.1).

Sequences processing

All the computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the Swiss Institute of Bioinformatics. A series of custom Perl and R scripts were used to re-format, concatenate and analyze all the data. Initially, the raw reads were processed with TagCleaner to trim Illumina adapters (Schmieder *et al.*, 2010). Reads were quality-filtered and trimmed using Prinseq-lite version 0.20.4 (Schmieder & Edwards, 2011). Low quality 3'-ends were trimmed and then reads containing uncalled bases (N) were removed. Only reads longer than 250 bp were kept for further analyses. Subsequently, "demultiplexing" was carried out to separate samples by barcode. Paired reads were then joined either with PEAR v 0.9.6 (Zhang *et al.*, 2014) or with custom awk scripts for regions longer than 550bp (as there was no overlapping ends between the forward and reverse reads, for these regions the paired reads were simply concatenated after quality trimming) (Table 1). Sequences in each sample were clustered removing identical reads using Fastx toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The frequency of redundant sequences was recorded to be used for the detection of chimeras.

Haplotype characterization

Within each amplified region, we determined the number of haplotypes by identifying variants present on individual sequence reads. Because sequence reads contain errors that can be misinterpreted as variants (consequently overestimating the number of different haplotypes) a threshold filtering method was used. Firstly, an NJ tree was constructed in Jalview (Clamp *et al.*, 2004) for each one of the regions with the 20 most frequent haplotypes. If several sequences clustered in an unresolved node, those were considered as being the same allele variants and were then collapsed into a single haplotype. A relative frequency threshold was iteratively applied to obtain the same number of haplotypes as fully resolved tips on the trees. After the trials, the selected threshold was 5%, meaning that a sequence was considered a true haplotype only if its frequency corresponded to at least 5% of the total number of reads per region and per isolate. Once identified, the haplotypes were aligned to the N6 reference genome with MAFFT v7.305 (Katoh & Standley, 2013), together with RAD reads, to verify that variants detected with amplicon sequencing were consistent with variants found in the RAD seq data. To assess the potential of the haplotype to be used as a marker, the phylogenetic resolution of the obtained haplotypes was assessed by constructing Maximum Likelihood trees using 1000 Rapid Bootstraps as implemented in RAxML v8.2.9 (Stamatakis, 2014).

As a confirmation of the haplotypes obtained, and as an initial approach to the use of the haplotypes as markers for field conditions, six regions (Table 1) were amplified as described above and cloned using the StrataClone PCR Cloning Kit following the protocol instructions. Colony-PCR was performed with the GoTaq DNA Polymerase (Promega) with T_7 as vector primer. Twenty clones per haplotype were sent for Sanger sequencing at GATC Biotech (Germany).

Results

Twenty-one potential loci conforming to the criteria were identified in the ddRAD-seq dataset. Primers were designed, synthetized and tested and only those with successful amplification for at least three of the isolates were retained (Table S1). Gel electrophoresis of amplified DNA showed a single fragment per genomic region and per isolate for 13 regions, the Scaffold6059 presented multiple fragments and was then discarded. Only 10 of the regions were finally sequenced. Having amplicons with a large difference in length (i.e. 271 vs 1749 bp) might have affected the chemistry of the sequencing flow cell, therefore only regions shorter than 1000 bp were sequenced (Table 1).

The sequencing data was deposited at the ENA repository and can be retrieved under the accession number PRJEB36796. After demultiplexing and quality control the number of reads per locus and per isolate ranged from 25k to 850k (Table S2). Due to low quality base trimming at the end of the sequences, their effective length was diminished and loci longer than 550 bp had to be manually concatenated using the N6 genome as a reference.

Out of the 10 regions, 4 regions presented a single haplotype in each isolate (Scaffold1820, Scaffold319, Scaffold8233a and Scaffold8233a) whereas the other 6 revealed 2, 3 or 4 haplotypes in some isolates (Table 1). The observations were consistent among technical replicates except for two regions in which one of the replicates showed an additional haplotype (Figure 2, Isolate C3 Scaffold 8233a and Scaffold 5277). When multiple haplotypes were present, the proportion of the number of reads assigned to each haplotype also varied. In regions with two alleles, each was represented in approximately 50% of the reads. Regions comprising three haplotypes had a distribution of approximately 50%, 25% and 25%, respectively. Four haplotypes, were detected in one region (isolate A2, Scaffold 5172) with an equal distribution of about 25% for each haplotype (Figure 2). In isolate C3 (Scaffold 8434), the proportions of the haplotypes were approximately 45%, 40%, 8% and 7%. Haplotypes at the different loci were composed of between 4 and 63 variable sites (Table 1). In homokaryotic isolates (i.e. DAOM197198, C2, A2) a unique haplotype is expected per marker, a maximum of two in a dikaryon (C3), however, for several genomic regions more than the expected haplotypes were found. Verification of the found haplotypes was performed with cloning and Sanger sequencing. This process yielded between 8 and 19 high quality sequences per region. The sequences obtained via Sanger sequencing confirmed multiple haplotypes for the tested regions (Table 1, Figure S2).

Table 1. List of loci used for haplotype identification. GeneMark-ES (Ter-hovhannisyan *et al*., 2008)was used to predict coding regions. The haplotypes in bold denote which loci and in which isolates cloning and Sanger sequencing was performed to verify amplicon sequencing data (Figure S2). Variable sites were identified as those sites where two or more unique bases were found among the haplotypes. DAOM is an abbreviation for isolate DAOM197198.

Scaffold name	**Predicted coding region**	**PCR product size (bp)**	**Number of haplotypes**				**Number of variable sites**
			A2	C2	C3	DAOM	
Scaffold1820	No	592	1	1	1	1	4
Scaffold10105	Yes	336	***3***	2	1	2	63
Scaffold319	No	991	NA	1	1	1	47
Scaffold6810	No	499	1	1	2	1	13
Scaffold8434	No	400	NA	1	***4***	***3***	27
Scaffold5967	Yes	762	***3***	***3***	1	1	28
Scaffold5172	Yes	271	***4***	NA	1	2	21
Scaffold5277	Yes	597	2	1	2	1	38
Scaffold8233a	Yes	744	1	1	1	1	50
Scaffold8233b	Yes	699	1	1	1	1	37

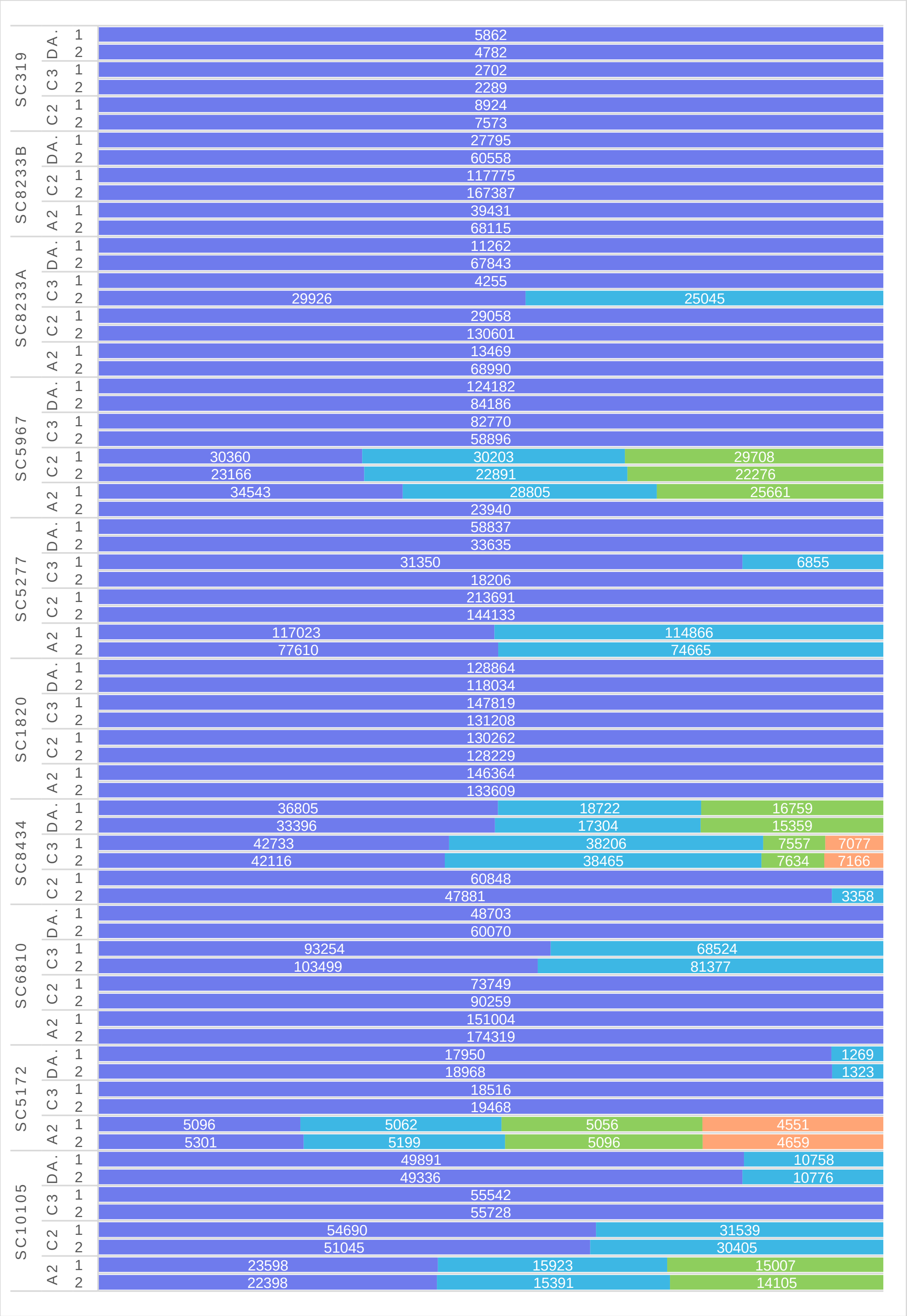


Figure 2. Proportion of haplotypes found in 10 loci within four isolates of *Rhizophagus irregularis*. Two technical replicates are shown for each locus in each isolate. A different color in a given row corresponds to a different haplotype. Numbers in the bars correspond to the total number of observed sequences of this haplotype. DA. stands for DAOM 197198. Not all the regions were successfully amplified in all the isolates

Maximum Likelihood (ML) trees showed relationships between the a given allele coming from different genomic regions. The ML trees were transformed to cladograms for visualization (Figure 1 and Figure S1). The resulting trees constructed from scaffolds 319, 5172, 8233a and 8233b showed a topology by which alleles of the same isolate grouped together. The technical replicates also clustered together. Each isolate and all its alleles was placed in an independent clade (Figure 1). This indicated that the variation in the regions examined was enough to distinguish isolates.

The remaining regions (10105, 1820, 5277, 5967, 6810 and 8434) presented either clades harboring alleles from different isolates or technical replicates located in independent clades (Figure S1). These characteristics make them unsuitable as strain-specific markers.

Discussion

The ultra-high coverage approach of this work, which was then confirmed by cloning and Sanger sequencing revealed that up to four haplotypes were found in the *R. irregularis* genome, where only one or two haplotypes were expected. This controverts the current thinking about these AMF isolates being strict homokaryons or dikaryons. Thus, what began as an approach to develop strain specific markers for tracking these AMF isolates in field conditions, ended up raising additional questions about the true homokaryotic or heterokaryotic nature of these isolates.

Masclaux *et al* (2019) found patterns of within-fungus genetic diversity in *R. irregularis* that were not consistent with the strict homokayon/dikaryon model. They compared sequencing data sets for several *R. irregularis* isolates and found discrepancies between methods (i.e. ddRAD-seq vs whole genome sequencing). This, highlights how the use of different data sets and analysis can bring to different conclusions. This work started using as a base the ddRAD-seq data generated by Wyss *et al.* (2016). The variant calling was then performed using N6 from Lin *et al.* (2014) as reference genome, as it was the only single nucleus genome assembly of this fungus at that date. However, with the advent of new available genomes re-analyzing the present data will likely show discrepancies when the variant calling will be performed using a different reference genome.

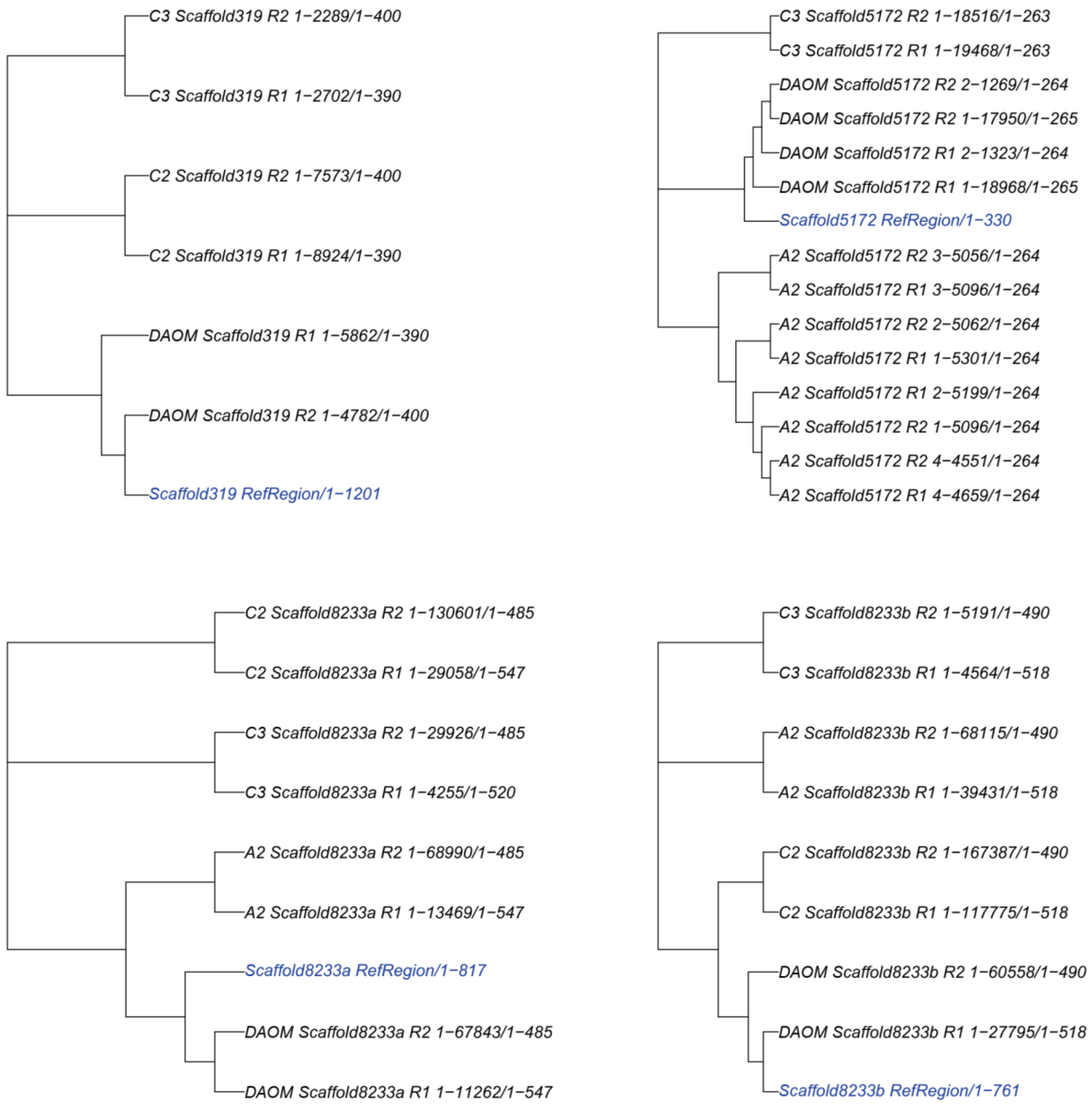

C3 Scaffold319 R2 1−2289/1−400
C3 Scaffold319 R1 1−2702/1−390
C2 Scaffold319 R2 1−7573/1−400
C2 Scaffold319 R1 1−8924/1−390
DAOM Scaffold319 R1 1−5862/1−390
DAOM Scaffold319 R2 1−4782/1−400
Scaffold319 RefRegion/1−1201
C3 Scaffold5172 R2 1−18516/1−263
C3 Scaffold5172 R1 1−19468/1−263
DAOM Scaffold5172 R2 2−1269/1−264
DAOM Scaffold5172 R2 1−17950/1−265
DAOM Scaffold5172 R1 2−1323/1−264
DAOM Scaffold5172 R1 1−18968/1−265
Scaffold5172 RefRegion/1−330
A2 Scaffold5172 R2 3−5056/1−264
A2 Scaffold5172 R1 3−5096/1−264
A2 Scaffold5172 R2 2−5062/1−264
A2 Scaffold5172 R1 1−5301/1−264
A2 Scaffold5172 R1 2−5199/1−264
A2 Scaffold5172 R2 1−5096/1−264
A2 Scaffold5172 R2 4−4551/1−264
A2 Scaffold5172 R1 4−4659/1−264
C2 Scaffold8233a R2 1−130601/1−485
C2 Scaffold8233a R1 1−29058/1−547
C3 Scaffold8233a R2 1−29926/1−485
C3 Scaffold8233a R1 1−4255/1−520
A2 Scaffold8233a R2 1−68990/1−485
A2 Scaffold8233a R1 1−13469/1−547
Scaffold8233a RefRegion/1−817
DAOM Scaffold8233a R2 1−67843/1−485
DAOM Scaffold8233a R1 1−11262/1−547
C3 Scaffold8233b R2 1−5191/1−490
C3 Scaffold8233b R1 1−4564/1−518
A2 Scaffold8233b R2 1−68115/1−490
A2 Scaffold8233b R1 1−39431/1−518
C2 Scaffold8233b R2 1−167387/1−490
C2 Scaffold8233b R1 1−117775/1−518
DAOM Scaffold8233b R2 1−60558/1−490
DAOM Scaffold8233b R1 1−27795/1−518
Scaffold8233b RefRegion/1−761

Figure 3. Cladograms of the regions were assessed to see which isolates grouped into independent clades. The notation indicates the isolate (DAOM, C2, C3 and A2), the replicate (R1 and R2), the allele and the frequency of this allele in the dataset (e.g. 1-11262) and the span of the aligned sequence (e.g. 1-547 bp). DAOM stands for DAOM 197198. The reference region from the N6 genome is showed in blue.

In view of this, in a homokaryotic state (e.g. DAOM, C2, A2) a unique haplotype should be found per marker and in a dikaryon (C3), a maximum of two haplotypes should be found.

However, we showed the presence of more than one haplotype in some regions in C2 and DAOM suggesting a heterokaryotic state. Furthermore, several loci, predicted as being in coding regions by GeneMark-ES (Table 1) exhibited up to 4 different haplotypes in C3. This showed that the observed within-fungus genetic diversity is not only present in regions prone to mutation accumulation but also those expected to be under purifying selection. We confirmed the existence of these alleles with cloning and Sanger sequencing which means they are extremely unlikely to simply be a result of sequencing error introduced in either RAD sequencing or amplicon sequencing.

Observing multiple alleles in coding regions suggests that the polymorphism within *R. irregularis* could potentially be functionally important and could be one of the factors contributing to the phenotypic plasticity of these fungi (Angelard *et al.*, 2014) and its broad adaptability to different soil types (Savary *et al.*, 2018). Chen *et al.* (2018b) reported high genome diversity and found this variation to have effects on protein domains. Masclaux *et al*. (2019) observed transcripts coming from two co-existing genomes in a heterokaryotic (dikaryon) isolate. This variation, and its implications in environmental adaptation and effects of AMF on plant host, remains to be studied. Additionally, the quality of the annotation in gene databases, remaining in most of the cases at ‘hypothetical proteins’ for the AMF case, hinders more precise conclusions about the biological significance of these allelic variants.

It has been suggested that the variation within one AMF individual is contained within highly similar nuclei and that this variation may be the result of genomic duplications. In this work, several approaches were used to ensure that the assessed genomic regions were indeed single-copy. The predictions with RepeatModeler and RepeatMasker, together with *in-silico* PCR, showed those regions to be predicted as non-repeated. There were no differences in the coverage of those regions in the whole genome sequencing (WGS) dataset that could be indicative of mis-assemblies of repeated regions. Additionally, WGS studies (Tisserant *et al.*, 2013; Lin *et al.*, 2014) revealed no evidence for extensive segmental duplications in the haploid genome. However, as the reference genome used to perform the variant calling in this study presents gaps, mis-assemblies cannot be ruled out. Wrongly assembled repetitions would show the kind of variation observed in this study. The assemblies of Lin *et al.* (2014), built using the typical short reads from Illumina sequencing (~150 bp) could be complimented by long read sequencing (e.g. Single Molecule, Real-Time (SMRT) Sequence reads ~10 kb) to improve the quality of the genomic assemblies and resolve this. Another potential source of variation comes from PCR-induced mutations and sequencing errors, although this is unlikely given that cloning and Sanger sequencing was performed. The ultra-high coverage obtained through amplicon sequencing and further validation through cloning and Sanger sequencing suggest that an artefactual origin of these variants is unlikely. Masclaux *et al.* (2019) re-analyzed published data across independent studies (using also different techniques) and concluded that the observed within-fungus genetic variation was not an artefact of sequencing errors and that such variation possibly exists. Thus, the unexpected within-isolate variation found in this study using ultra-high coverage amplicon sequencing remains a credible explanation of this variation.

Our goal was to design potential isolate specific markers of this fungal species. Recently, Savary *et al.* (2018), using a high-resolution population genomics approach, was able to classify *R. irregularis* isolates into 4 genetic groups. DAOM and C2 were placed in group 4 and C3 and C2 in group 3. This was achieved by concatenating 6888 SNPs. A high number of variants are needed to separate these isolates using molecular data. Finding genomic regions able to discriminate between isolates with the relatively short sequences used in this study indicates that this approach can potentially be used to develop strain-specific markers. However, both the existence and the nature of this variation remain to be confirmed. Additionally, the markers designed and tested in this work were performed on a small number of isolates. Ensuring that a marker is indeed strain-specific requires validation using a large number of isolates. However, these markers look promising.

Savary *et al.* (2018) showed that almost-clonal isolates of *R. irregularis* occurred in distant localities up to 4000 km apart. This corroborates the findings of Davison *et al.* (2015) of low endemism in AMF where “everyone is everywhere”. As *R. irregularis* likely has a large geographical distribution of almost-clonal isolates, it is probable that the alleles characterized in the present study, and intended as isolate-specific markers, are already present in the field before inoculation because clones or closely related AMF are already present. This, on the one hand, dismisses the concerns of *R. irregularis* being an exotic and potentially invasive fungus (Schwartz *et al.*, 2006; Hart *et al.*, 2017) and inoculation as a practice that will introduce new alleles in a community. However, on the other hand, the fact that those alleles are most likely already present in the field hinders the possibility to use them as tracking markers. The alleles already present in the field would most likely be identical to the alleles possessed by the inoculated AMF, making them undistinguishable. This still has to be assessed for these markers.

The extent of within-fungus variation in AMF is difficult to study given the particularities of their life cycle. Additionally, as an obligate symbiont, the culturing of AMF presents also challenges which although surpassed only allow for handful of isolates from few species can be cultured *in vitro*. Thus, investigating within-AMF genetic variation remains restricted to a small portion of the species described. Savary *et al.* (2018) called for isolation and study of isolates from agronomic and natural ecosystems from other continents such as Australia, Africa, Asia and South America. Independent studies using different techniques have detected within-fungus variation which was observed also at the transcriptomic level. Further studies are needed to determine the nature of the within-fungus genetic variation, its implications in environmental adaptation of AMF and effects of AMF on plant growth. The understanding of the within-fungus genetic variation will provide the basis for the development of strain-specific markers needed for an economically and ecologically safe use of AMF in agriculture.

References

Akyol TY, Niwa R, Hirakawa H, Maruyama H, Sato T, Suzuki T, Fukunaga A, Sato T, Yoshida S, Tawaraya K, et al. 2019. Impact of introduction of arbuscular mycorrhizal fungi on the root microbial community in agricultural fields. Microbes and Environments 34: 23–32.

Angelard C, Tanner CJ, Fontanillas P, Niculita-Hirzel H, Masclaux F, Sanders IR. 2014. Rapid genotypic change and plasticity in arbuscular mycorrhizal fungi is caused by a host shift and enhanced by segregation. The ISME Journal 8: 284–294.

Backer R, Rokem JS, Ilangumaran G, Lamont J, Praslickova D, Ricci E, Subramanian S, Smith DL. 2018. Plant growth-promoting rhizobacteria: Context, mechanisms of action, and roadmap to commercialization of biostimulants for sustainable agriculture. Frontiers in Plant Science 871: 1–17.

Berg G. 2009. Plant-microbe interactions promoting plant growth and health: Perspectives for controlled use of microorganisms in agriculture. Applied Microbiology and Biotechnology 84: 11–18.

Buysens C, Alaux PL, César V, Huret S, Declerck S, Cranenbrouck S. 2017. Tracing native and inoculated *Rhizophagus irregularis* in three potato cultivars (Charlotte, Nicola and Bintje) grown under field conditions. Applied Soil Ecology 115: 1–9.

Ceballos I, Mateus ID, Peña R, Peña-Quemba DC, Masso C, Vanlauwe B, Rodriguez A, Sanders IR. 2019. Using variation in arbuscular mycorrhizal fungi to drive the productivity of the food security crop cassava: 1–21.

Ceballos I, Ruiz M, Fernández C, Peña R, Rodríguez A, Sanders IR. 2013. The in vitro mass-produced model mycorrhizal fungus, *Rhizophagus irregularis*, significantly increases yields of the globally important food security crop cassava. PLoS ONE 8.

Chen ECH, Morin E, Beaudet D, Noel J, Yildirir G, Ndikumana S, Charron P, St-Onge C, Giorgi J, Krüger M, et al. 2018. High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. New Phytologist 220: 1161–1171.

Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. Bioinformatics 20: 426–427.

Davison J, Moora M, Öpik M, Adholeya A, Ainsaar L, Bâ A, Burla S, Diedhiou AG, Hiiesalu I, Jairus T, et al. 2015. Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. Science 127: 970–973.

Hart MM, Antunes PM, Abbott LK. 2017. Unknown risks to soil biodiversity from commercial fungal inoculants. Nature Ecology and Evolution 1: 1.

van der Heijden MGA, Martin FM, Selosse MA, Sanders IR. 2015. Mycorrhizal ecology and evolution: The past, the present, and the future. New Phytologist 205: 1406–1423.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30: 772–780.

Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DGO, Mu D, Pang E, Cao H, Cha H, Lin T, et al. 2014. Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. PLoS Genetics 10.

Masclaux FG, Wyss T, Pagni M, Rosikiewicz P, Sanders IR. 2019. Investigating unexplained genetic variation and its expression in the arbuscular mycorrhizal fungus *Rhizophagus irregularis*: A comparison of whole genome and RAD sequencing data. PLoS ONE 14: 1–20.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences of the United States of America 85: 2444–8.

Pellegrino E, Turrini A, Gamper HA, Cafà G, Bonari E, Young JPW, Giovannetti M. 2012. Establishment, persistence and effectiveness of arbuscular mycorrhizal fungal inoculants in the field revealed using molecular genetic tracing and measurement of yield components. New Phytologist 194: 810–822.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16: 276–277.

Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. 2017. Variant review with the integrative genomics viewer. Cancer Research 77: e31–e34.

Ropars J, Toro KS, Noel J, Pelin A, Charron P, Farinelli L, Marton T, Krüger M, Fuchs J, Brachmann A, et al. 2016. Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. Nature Microbiology: 16033.

Savary R, Masclaux FG, Wyss T, Droh G, Cruz Corella J, Machado AP, Morton JB, Sanders IR. 2018. A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus *Rhizophagus irregularis*. ISME Journal 12: 17–30.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863–864.

Schmieder R, Lim YW, Rohwer F, Edwards R. 2010. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. BMC bioinformatics 11: 341.

Schwartz MW, Hoeksema JD, Gehring CA, Johnson NC, Klironomos JN, Abbott LK, Pringle A. 2006. The promise and the potential consequences of the global transport of mycorrhizal fungal inoculum. Ecology Letters 9: 501–515.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Research 18: 1979–1990.

Thioye B, Sanguin H, Kane A, de Faria SM, Fall D, Prin Y, Sanogo D, Ndiaye C, Duponnois R, Sylla SN, et al. 2019. Impact of mycorrhiza-based inoculation strategies on *Ziziphus mauritiana* Lam. and its native mycorrhizal communities on the route of the Great Green Wall (Senegal). Ecological Engineering 128: 66–76.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, Frei dit Frey N, Gianinazzi-Pearson V, et al. 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proceedings of the National Academy of Sciences of the United States of America 110: 20117–22.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Research 40: 1–12.

Wyss T, Masclaux FG, Rosikiewicz P, Pagni M, Sanders IR. 2016. Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus *Rhizophagus irregularis*. The ISME journal: 1–13.

Young JPW. 2015. Genome diversity in arbuscular mycorrhizal fungi. Current Opinion in Plant Biology 26: 113–119.

Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30: 614–620.

Zhang S, Lehmann A, Zheng W, You Z, Rillig MC. 2019. Arbuscular mycorrhizal fungi increase grain yields: a meta-analysis. New Phytologist 222: 543–555.

Table S1. List of loci used for haplotype identification. The primer sequences and product size correspond to the output of Pirmer3. The haplotypes in bold denote for which genomic regions amplification was obtained for at least 3 of the isolates.Scaffold6059 (underlined) presented unspecific bands after PCR optimization and was therefore discarded from the analyses. DAOM is an abbreviation for isolate DAOM197198.

Region	**Forward primer (5'-3',without barcode)**	**Reverse primer (5'-3')**	**Product Size**	**Amplification**			
				DAOM	C2	A2	C3
Scaffold10105	TCAGATGCAGAGCACCTACAG	CGAATAAAACAAGTCGGTCCATAG	336	+	+	+	+
Scaffold10997	GGCAGTCCCCTAATCGAAC	TTGCAGAAGTTTCCCTCCTG	998	+	-	-	-
Scaffold1595	CGTCCTGTGTTCAACTACCC	TCGCCATCCACTAGTTTGTC	1580	+	+	+	+
Scaffold1706	CAGGGAAAAGACGTCTCCAG	GGACTATTTCGACCAGTGAGG	1157	+	-	-	-
Scaffold1820	AATTTCAACGTACGGATCATAGAG	TTCGGGGCAAAACTTGTTAG	592	+	+	+	+
Scaffold248	TCCGTGACTGGGAAGGTATC	AACCGTTCCACACATTACCG	1336	+	-	-	+
Scaffold3025	TCCGATATTGCACTTGAAGC	TCAGGTTCTATTTGCCGGTTAC	1749	+	+	+	+
Scaffold319	CAACGGCTCTTATGTTCTTGT	TCACGTAGTTAAAAATTGAATCAGA	991	+	+	-	+
Scaffold3438	GGATTCACACGGAAACATTC	AAATTTGAGTTACCGACGAAAC	1187	+	+	+	+
Scaffold3546	CCAAAAGTTGAGGTAAGCTTCTAC	AACTTTTAAACCAAATTATGTCATACC	1031	+	-	-	-
Scaffold357	TTGAGATCTATAAATGGACTTTTCAAT	CGTATGGGATTACTTGGAGGAA	899	+	-	-	+
Scaffold374	TTGGAATTTCTCTTGAAAGTATTTG	AGCGCAAGTTTTATCGAAAGG	699	+	-	+	-
Scaffold4559	TTTCGATCATAATTCTGTTCATTG	GGTGATGATGAGGGGAACG	464	+	-	-	+
Scaffold5172	ATCACATCTACCAGTTCTTGGTC	AATTGCAGAAGAAGCTTGGAC	271	+	-	+	+
Scaffold5277	GGCCTTATAATGTAAGTAATAATCCTG	AAATATTCTTGACTCATTATCCATTCC	597	+	+	+	+
Scaffold5967	TGATGTGTAAACCAAATTCCTG	GTGGTGGGAATTCTGACTG	762	+	+	+	+
Scaffold6059	TTTAACAGGTTGGTGAATTG	CAATTTATAAAAGATATCAAGTAAAGG	367	+	+	+	+
Scaffold6810	TGGTCTCTCCATTAACGTATTTC	AAATTTTATGAAAACATTCTTTGATCC	499	+	+	+	+
Scaffold8233a	CCAGGTTTTTATGCTGATCG	TTCACTAAATGGAGAAACTGAAAATG	744	+	+	+	+
Scaffold8233b	GGTTTTGGTGATAATGTTTTTCC	CAACGCTTGTTGCTTAAGATCC	699	+	+	+	+
Scaffold8434	CGTAGTAAGGTCTTTAACGGGTTG	TTTCTTTGACGTTCTTGTTCCAG	400	+	+	-	+

Table S2. List of loci used for haplotype identification. The figures correspond to the number of reads passing the quality filter. The percentual variation between technical replicates is presented. DAOM is an abbreviation for isolate DAOM197198.

Region	***Isolate***	***Replicate 1***	***Replicate 2***	***Variation***
Scaffold10105	A2	88,562	88,985	0.48%
	C2	121,865	123,283	1.16%
	C3	74,215	70,476	5.17%
	DAOM	81,705	78,421	4.10%
Scaffold1820	A2	433,091	441,626	1.95%
	C2	369,457	398,498	7.56%
	C3	418,857	414,038	1.16%
	DAOM	364,264	366,341	0.57%
Scaffold319	C2	141,806	150,495	5.95%
	C3	52,972	53,565	1.11%
	DAOM	85,191	93,248	9.03%
Scaffold5172	A2	27,595	26,438	4.28%
	C3	23,829	22,074	7.65%
	DAOM	25,720	23,855	7.52%
Scaffold5277	A2	852,197	805,164	5.68%
	C2	647,580	636,214	1.77%
	C3	141,551	127,768	10.24%
	DAOM	187,126	177,457	5.30%
Scaffold5967	A2	508,906	503,331	1.10%
	C2	429,525	453,130	5.35%
	C3	427,434	423,794	0.86%
	DAOM	425,190	416,367	2.10%
Scaffold6810	A2	449,154	435,393	3.11%
	C2	388,942	399,490	2.68%
	C3	447,733	424,285	5.38%
	DAOM	257,297	267,499	3.89%
Scaffold8233a	A2	511,166	504,388	1.33%
	C2	569,564	579,991	1.81%
	C3	441,946	423,933	4.16%
	DAOM	517,279	520,118	0.55%
Scaffold8233b	A2	731,306	675,097	7.99%
	C2	763,738	739,508	3.22%
	C3	291,000	266,351	8.85%
	DAOM	565,999	547,700	3.29%
Scaffold8434	C2	67,255	81,834	19.56%
	C3	137,569	131,466	4.54%
	DAOM	100,537	105,658	4.97%

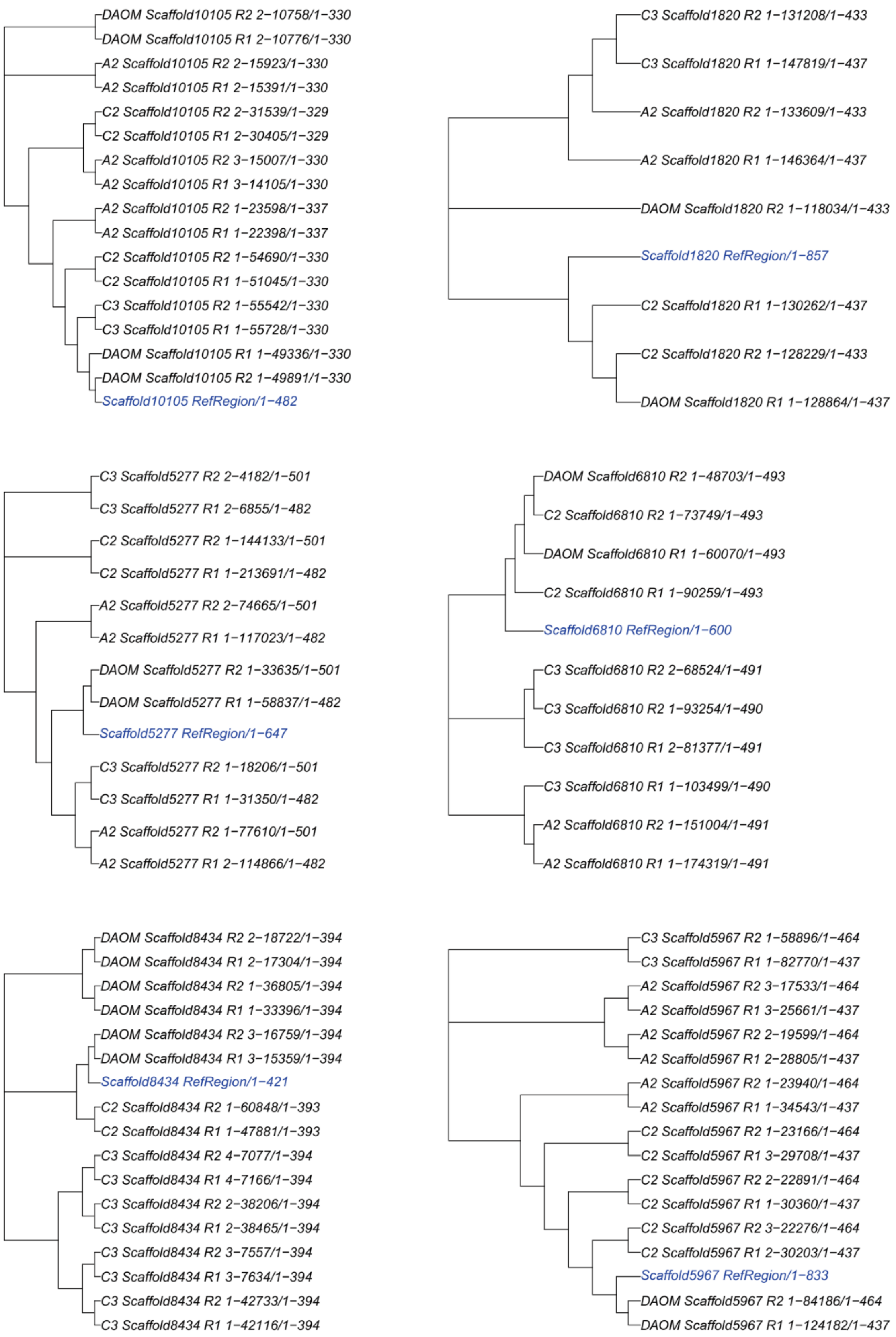
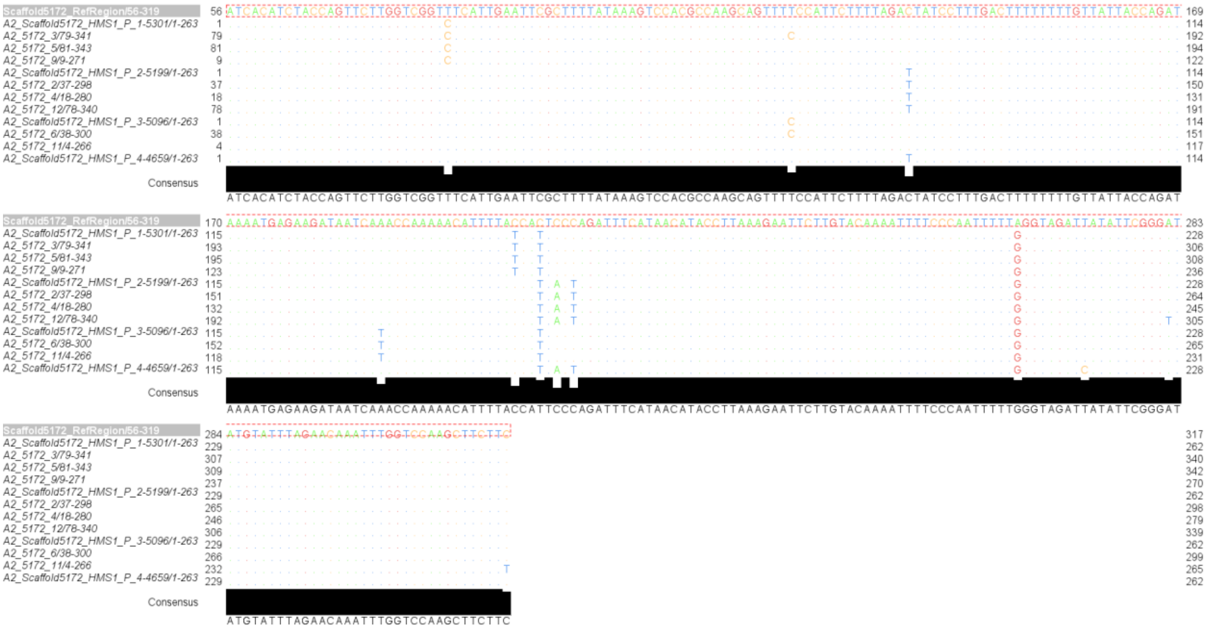

DAOM Scaffold10105 R2 2−10758/1−330
DAOM Scaffold10105 R1 2−10776/1−330
A2 Scaffold10105 R2 2−15923/1−330
A2 Scaffold10105 R1 2−15391/1−330
C2 Scaffold10105 R2 2−31539/1−329
C2 Scaffold10105 R1 2−30405/1−329
A2 Scaffold10105 R2 3−15007/1−330
A2 Scaffold10105 R1 3−14105/1−330
A2 Scaffold10105 R2 1−23598/1−337
A2 Scaffold10105 R1 1−22398/1−337
C2 Scaffold10105 R2 1−54690/1−330
C2 Scaffold10105 R1 1−51045/1−330
C3 Scaffold10105 R2 1−55542/1−330
C3 Scaffold10105 R1 1−55728/1−330
DAOM Scaffold10105 R1 1−49336/1−330
DAOM Scaffold10105 R2 1−49891/1−330
Scaffold10105 RefRegion/1−482
C3 Scaffold1820 R2 1−131208/1−433
C3 Scaffold1820 R1 1−147819/1−437
A2 Scaffold1820 R2 1−133609/1−433
A2 Scaffold1820 R1 1−146364/1−437
DAOM Scaffold1820 R2 1−118034/1−433
Scaffold1820 RefRegion/1−857
C2 Scaffold1820 R1 1−130262/1−437
C2 Scaffold1820 R2 1−128229/1−433
DAOM Scaffold1820 R1 1−128864/1−437
C3 Scaffold5277 R2 2−4182/1−501
C3 Scaffold5277 R1 2−6855/1−482
C2 Scaffold5277 R2 1−144133/1−501
C2 Scaffold5277 R1 1−213691/1−482
A2 Scaffold5277 R2 2−74665/1−501
A2 Scaffold5277 R1 1−117023/1−482
DAOM Scaffold5277 R2 1−33635/1−501
DAOM Scaffold5277 R1 1−58837/1−482
Scaffold5277 RefRegion/1−647
C3 Scaffold5277 R2 1−18206/1−501
C3 Scaffold5277 R1 1−31350/1−482
A2 Scaffold5277 R2 1−77610/1−501
A2 Scaffold5277 R1 2−114866/1−482
DAOM Scaffold6810 R2 1−48703/1−493
C2 Scaffold6810 R2 1−73749/1−493
DAOM Scaffold6810 R1 1−60070/1−493
C2 Scaffold6810 R1 1−90259/1−493
Scaffold6810 RefRegion/1−600
C3 Scaffold6810 R2 2−68524/1−491
C3 Scaffold6810 R2 1−93254/1−490
C3 Scaffold6810 R1 2−81377/1−491
C3 Scaffold6810 R1 1−103499/1−490
A2 Scaffold6810 R2 1−151004/1−491
A2 Scaffold6810 R1 1−174319/1−491
DAOM Scaffold8434 R2 2−18722/1−394
DAOM Scaffold8434 R1 2−17304/1−394
DAOM Scaffold8434 R2 1−36805/1−394
DAOM Scaffold8434 R1 1−33396/1−394
DAOM Scaffold8434 R2 3−16759/1−394
DAOM Scaffold8434 R1 3−15359/1−394
Scaffold8434 RefRegion/1−421
C2 Scaffold8434 R2 1−60848/1−393
C2 Scaffold8434 R1 1−47881/1−393
C3 Scaffold8434 R2 4−7077/1−394
C3 Scaffold8434 R1 4−7166/1−394
C3 Scaffold8434 R2 2−38206/1−394
C3 Scaffold8434 R1 2−38465/1−394
C3 Scaffold8434 R2 3−7557/1−394
C3 Scaffold8434 R1 3−7634/1−394
C3 Scaffold8434 R2 1−42733/1−394
C3 Scaffold8434 R1 1−42116/1−394
C3 Scaffold5967 R2 1−58896/1−464
C3 Scaffold5967 R1 1−82770/1−437
A2 Scaffold5967 R2 3−17533/1−464
A2 Scaffold5967 R1 3−25661/1−437
A2 Scaffold5967 R2 2−19599/1−464
A2 Scaffold5967 R1 2−28805/1−437
A2 Scaffold5967 R2 1−23940/1−464
A2 Scaffold5967 R1 1−34543/1−437
C2 Scaffold5967 R2 1−23166/1−464
C2 Scaffold5967 R1 3−29708/1−437
C2 Scaffold5967 R2 2−22891/1−464
C2 Scaffold5967 R1 1−30360/1−437
C2 Scaffold5967 R2 3−22276/1−464
C2 Scaffold5967 R1 2−30203/1−437
Scaffold5967 RefRegion/1−833
DAOM Scaffold5967 R2 1−84186/1−464
DAOM Scaffold5967 R1 1−124182/1−437

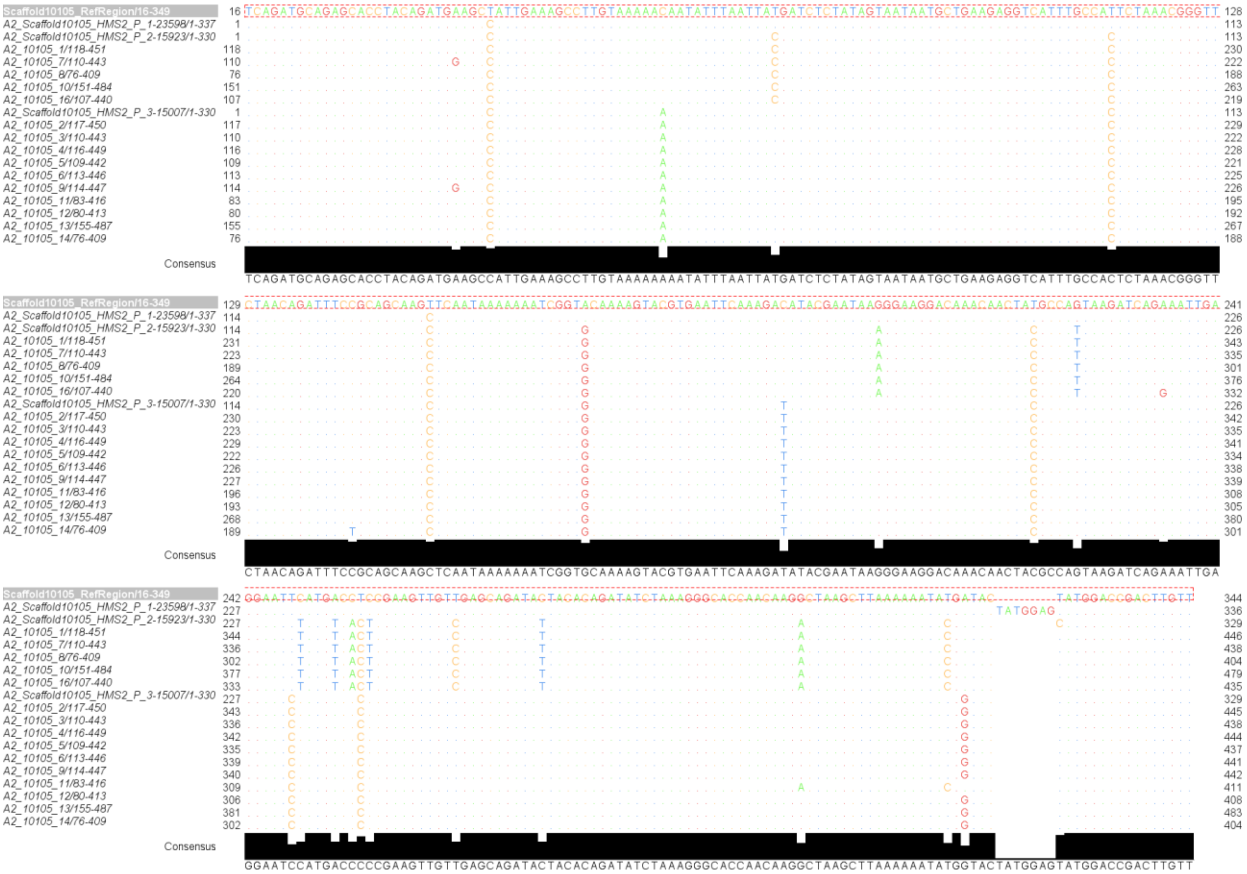
Figure S1. Cladograms of the regions assessed for which isolated were not grouped in independent clades. The notation indicates the isolate (DAOM, C2, C3 and A2), the replicate (R1 and R2), the allele and the frequency of this allele in the dataset (e.g. 1-42116) and the span of the aligned sequence (e.g. 1-394 bp). DAOM stands for DAOM 197198. The reference region from the N6 genome is showed in blue.

Figure S2 Aligments of haplotypes found in *R.irregularis* isolates. The alignment includes the sequence of N6 as the reference genome (marked with 'RefRegion'), the haplotypes detected with amplicon sequencing (e.g. A2_Scaffold5172_HMS) and sequences derived from cloning and Sanger sequencing (e.g. A2_5172_3). Only variable positions respect to the reference are shown. The black bar at the bottom shows the conservation percentage. Alignments were built using MAFFT with the G-INS-i preset (higher accuracy) and then visualized with Jalview

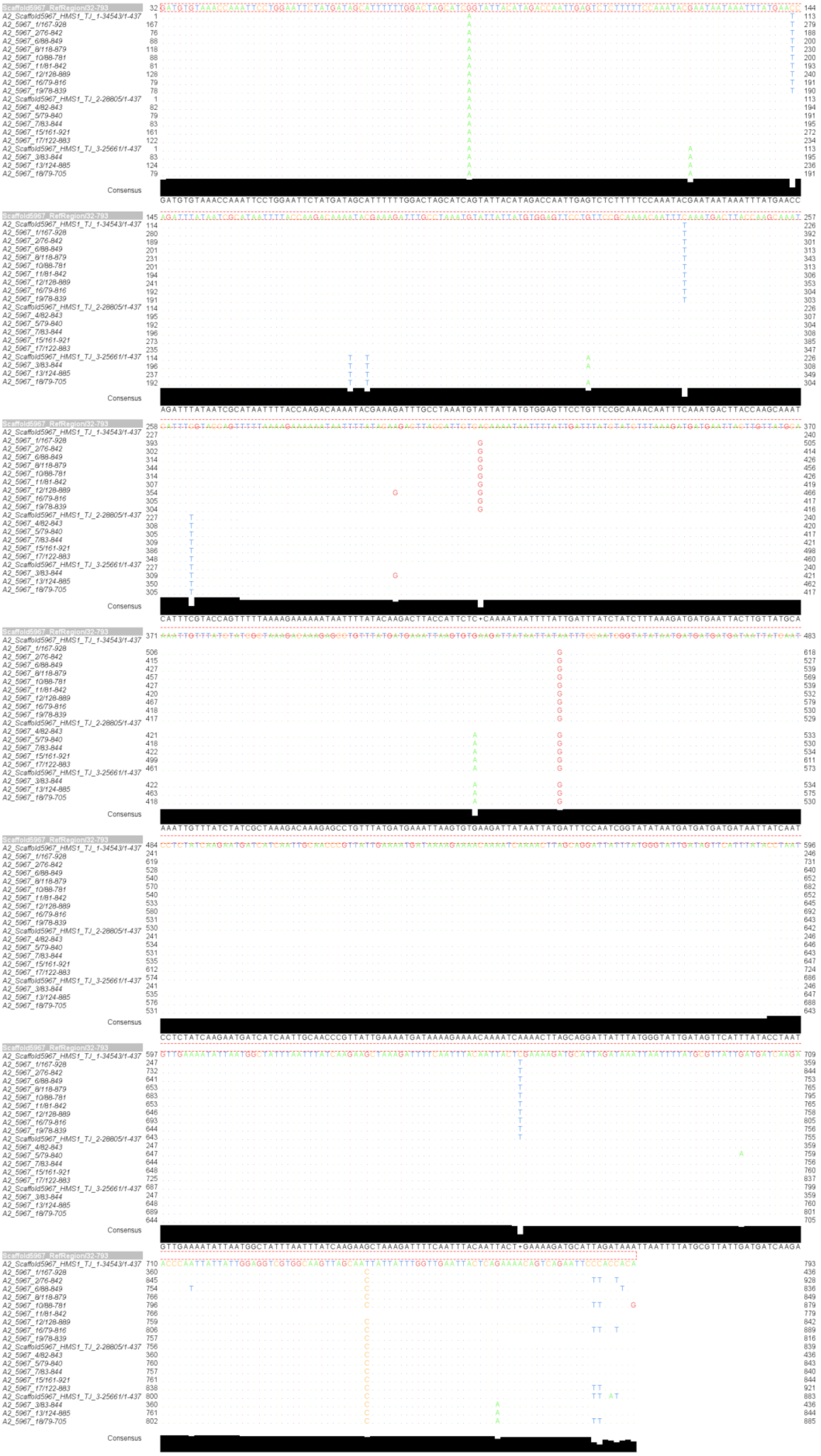
a) Scaffold5172 for isolate A2


Scaffold5172_RefRegion/56-319
A2_Scaffold5172_HMS1_P_1-5301/1-263
A2_5172_3/79-341
A2_5172_5/81-343
A2_5172_9/9-271
A2_Scaffold5172_HMS1_P_2-5199/1-263
A2_5172_2/37-298
A2_5172_4/18-280
A2_5172_12/78-340
A2_Scaffold5172_HMS1_P_3-5096/1-263
A2_5172_6/38-300
A2_5172_11/4-266
A2_Scaffold5172_HMS1_P_4-4659/1-263
Consensus
ATCACATCTACCAGTTCTTGGTCGGTTTCATTGAATTCGCTTTTATAAAGTCCACGCCAAGCAGTTTTCCATTCTTTTAGACTATCCTTTGACTTTTTTTTGTTATTACCAGAT
AAAATGAGAAGATAATCAAACCAAAAACATTTTACCATTCCCAGATTTCATAACATACCTTAAAGAATTCTTGTACAAAATTTTCCCAATTTTTGGGTAGATTATATTCGGGAT
ATGTATTTAGAACAAATTTGGTCCAAGCTTCTTC

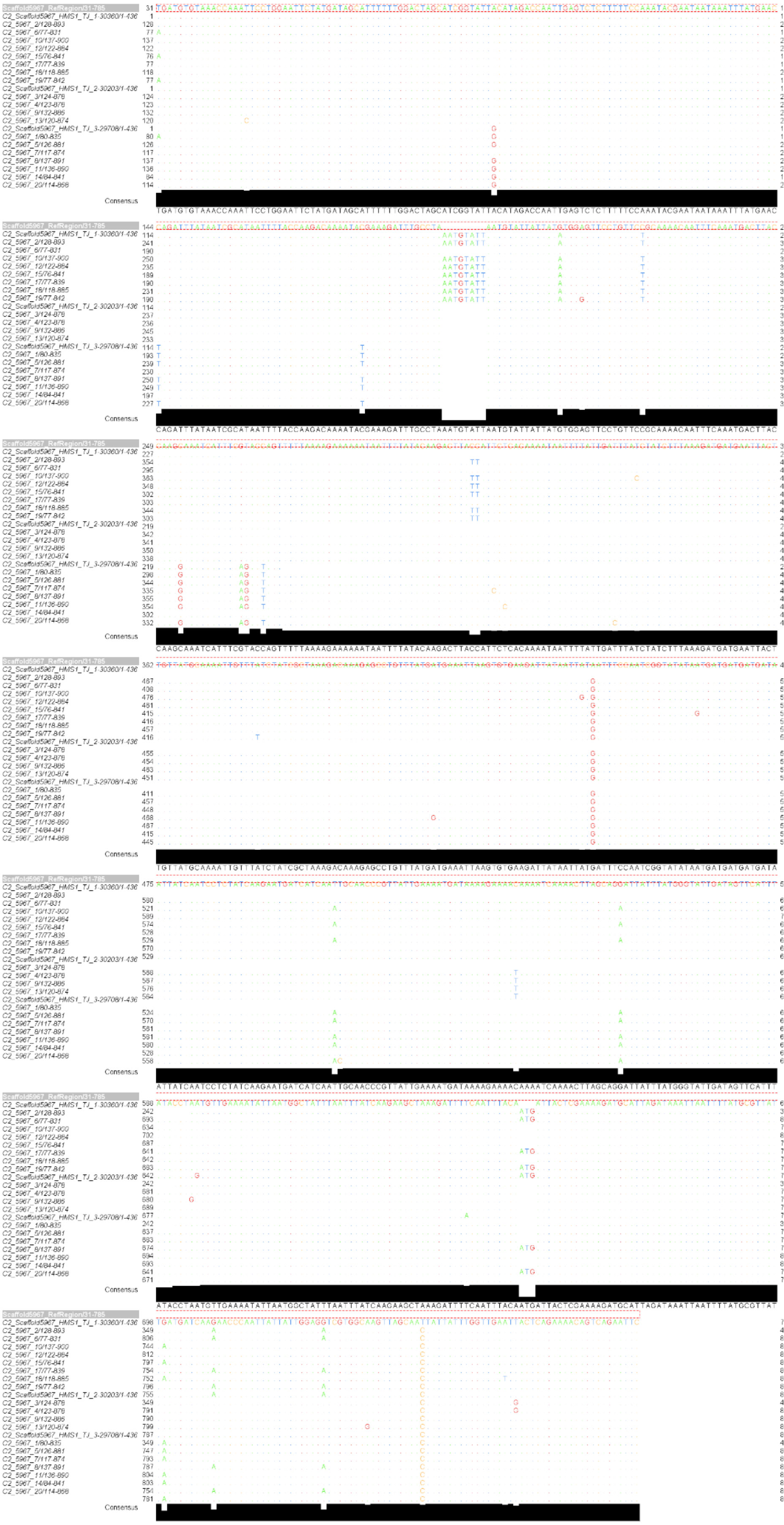
b) Scaffold10105 for isolate A2


Scaffold10105_RefRegion/16-349
A2_Scaffold10105_HMS2_P_1-23598/1-337
A2_Scaffold10105_HMS2_P_2-15923/1-330
A2_10105_1/118-451
A2_10105_7/110-443
A2_10105_8/76-409
A2_10105_10/151-484
A2_10105_16/107-440
A2_Scaffold10105_HMS2_P_3-15007/1-330
A2_10105_2/117-450
A2_10105_3/110-443
A2_10105_4/116-449
A2_10105_5/109-442
A2_10105_6/113-446
A2_10105_9/114-447
A2_10105_11/83-416
A2_10105_12/80-413
A2_10105_13/155-487
A2_10105_14/76-409
Consensus
TCAGATGCAGAGCACCTACAGATGAAGCCATTGAAAGCCTTGTAAAAAAAATATTTAATTATGATCTCTATAGTAATAATGCTGAAGAGGTCATTTGCCACTCTAAACGGGTT
CTAACAGATTTCCGCAGCAAGCTCAATAAAAAAAATCGGTGCAAAAGTACGTGAATTCAAAGATATACGAATAAGGGAAGGACAAACAACTACGCCAGTAAGATCAGAAATTGA
GGAATCCATGACCCCCGAAGTTGTTGAGCAGATACTACACAGATATCTAAAGGGCACCAACAAGGCTAAGCTTAAAAAAATATGGTACTATGGAGTATGGACCGACTTGTT

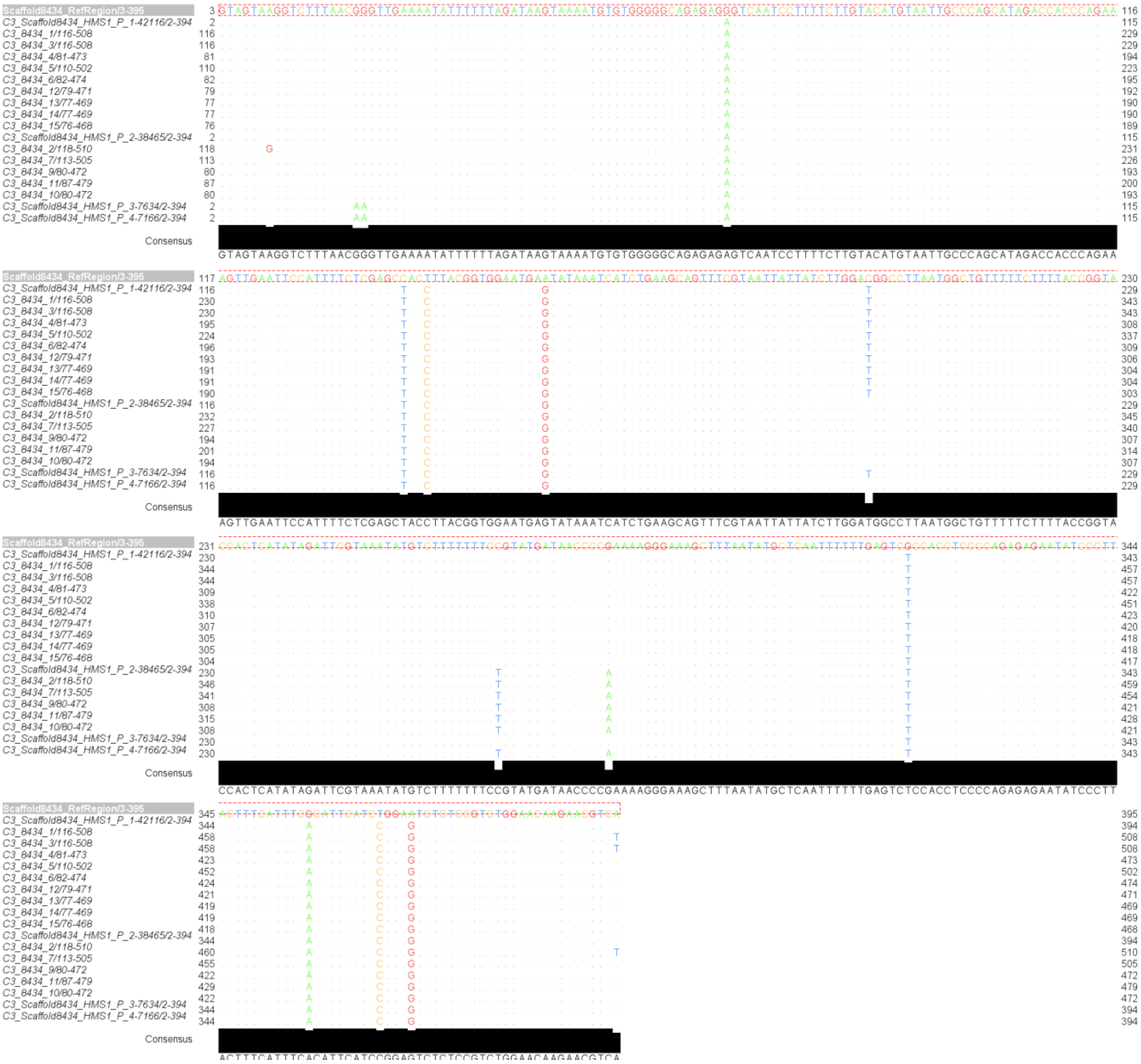
c) Scaffold5967 for isolate A2


Scaffold5967_RefRegion/32-793
A2_Scaffold5967_HMS1_TJ_1-34543/1-437
A2_5967_1/167-928
A2_5967_2/76-842
A2_5967_6/88-849
A2_5967_8/118-879
A2_5967_10/88-781
A2_5967_11/81-842
A2_5967_12/128-889
A2_5967_16/79-816
A2_5967_19/78-839
A2_Scaffold5967_HMS1_TJ_2-28805/1-437
A2_5967_4/82-843
A2_5967_5/79-840
A2_5967_7/83-844
A2_5967_15/161-921
A2_5967_17/122-883
A2_Scaffold5967_HMS1_TJ_3-25661/1-437
A2_5967_3/83-844
A2_5967_13/124-885
A2_5967_18/79-705
Consensus
GATGTGTAAACCAAATTCCTGGAATTCTATGATAGCATTTTTTGGACTAGCATCAGTATTACATAGACCAATTGAGTCTCTTTTTCCAAATACGAATAATAAATTTATGAACC
AGATTTATAATCGCATAATTTTACCAAGACAAAATACGAAAGATTTGCCTAAATGTATTATTATGTGGAGTTCCTGTTCCGCAAAACAATTTCAAATGACTTACCAAGCAAAT
CATTTCGTACCAGTTTTTAAAAGAAAAAATAATTTTATACAAGACTTACCATTCTC-CAAAATAATTTTATTGATTTATCTATCTTTAAAGATGATGAATTACTTGTTATGCA
AAATTGTTTATCTATCGCTAAAGACAAAGAGCCTGTTTATGATGAAATTAAGTGTGAAGATTATAATTATGATTTCCAATCGGTATATAATGATGATGATAATTATCAAT
CCTCTATCAAGAATGATCATCAATTGCAACCCGTTATTGAAAATGATAAAAGAAAACAAAATCAAAACTTAGCAGGATTATTTATGGGTATTGATAGTTCATTTATACCTAAT
GTTGAAAATATTAATGGCTATTTAATTTATCAAGAAGCTAAAGATTTTCAATTTACAATTACT-GAAAAGATGCATTAGATAAATTAATTTTATGCGTTATTGATGATCAAGA

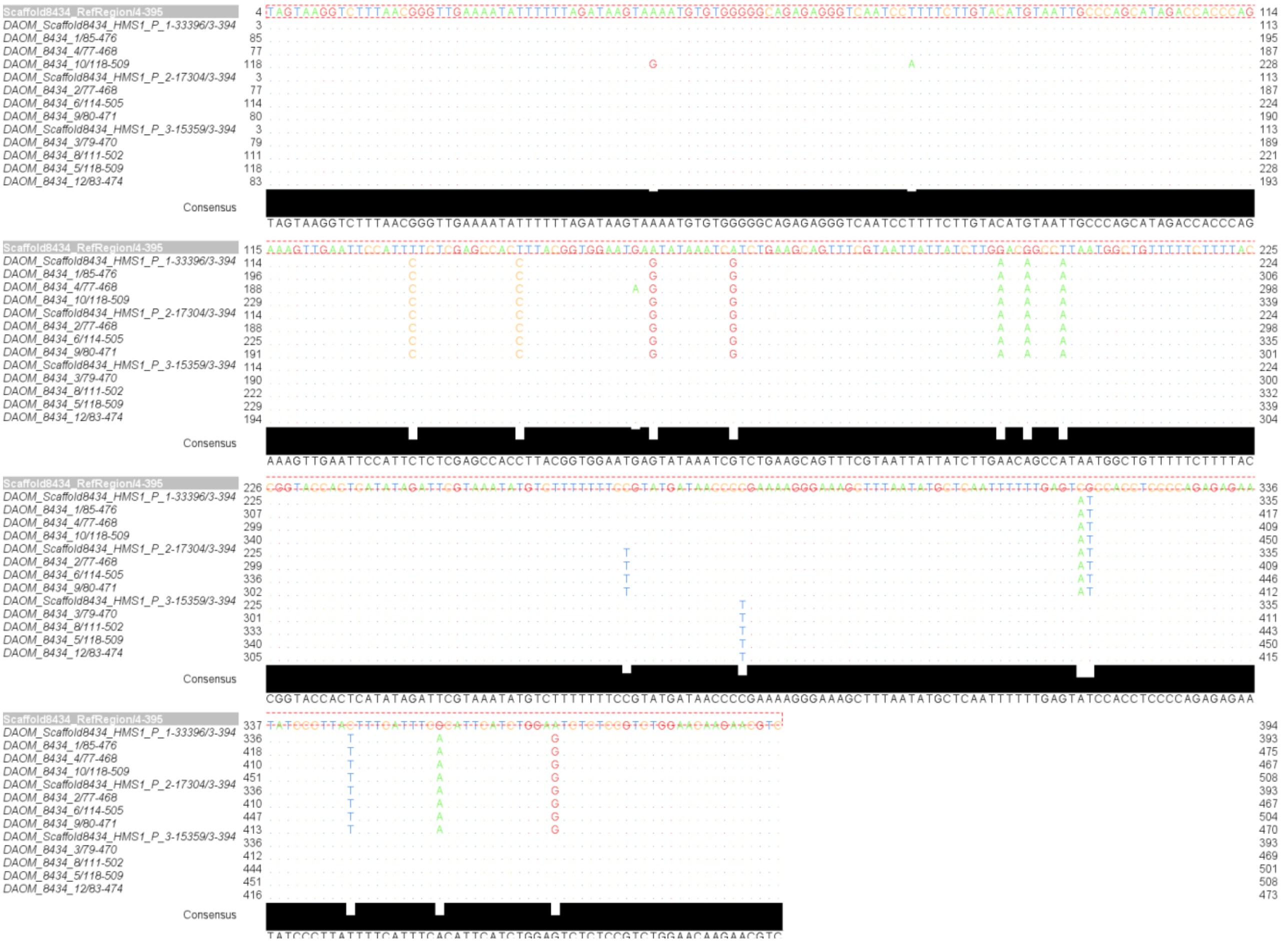
d) Scaffold5967 for isolate C3



e) Scaffold8434 for isolate C3


Scaffold8434_RefRegion/3-395
C3_Scaffold8434_HMS1_P_1-42116/2-394
C3_8434_1/116-508
C3_8434_3/116-508
C3_8434_4/81-473
C3_8434_5/110-502
C3_8434_6/82-474
C3_8434_12/79-471
C3_8434_13/77-469
C3_8434_14/77-469
C3_8434_15/76-468
C3_Scaffold8434_HMS1_P_2-38465/2-394
C3_8434_2/118-510
C3_8434_7/113-505
C3_8434_9/80-472
C3_8434_11/87-479
C3_8434_10/80-472
C3_Scaffold8434_HMS1_P_3-7634/2-394
C3_Scaffold8434_HMS1_P_4-7166/2-394
Consensus
GTAGTAAGGTCTTTAACGGGTTGAAAATATTTTTTAGATAAGTAAAATGTGTGGGGGCAGAGAGAGTCAATCCTTTTCTTGTACATGTAATTGCCCAGCATAGACCACCCAGAA
AGTTGAATTCCATTTTCTCGAGCTACCTTACGGTGGAATGAGTATAAATCATCTGAAGCAGTTTCGTAATTATTATCTTGGATGGCCTTAATGGCTGTTTTTCTTTTACCGGTA
CCACTCATATAGATTCGTAAATATGTCTTTTTTTCCGTATGATAACCCCGAAAAGGGAAAGCTTTAATATGCTCAATTTTTTGAGTCTCCACCTCCCCAGAGAGAATATCCCTT
ACTTTCATTTCACATTCATCCGGAGTCTCTCCGTCTGGAACAAGAACGTCA

f) Scaffold8434 for isolate DAOM197198



General discussion and perspectives

The main goal of this PhD thesis was to understand the effects of experimental manipulations (modifying either biotic or abiotic components) on the soil microbiome. In this work I aimed to use the advances in sequencing technologies and bioinformatic algorithms to provide insights into the complexity of soil microbiomes.

In chapter 1, using the framework of the Transplant network, we started a collaboration with the group of Jake Alexander (now in Zurich). In this project, we aimed to see whether or not there were patterns in the soil metagenome changes induced by transplantation. This, across a series of regions in the northern hemisphere with multiple localities within each region. The results showed that changes in the soil metagenome were not the same across the different regions and localities assessed.

In chapter 2, an experiment was set in an individual location, meaning the abiotic conditions were the same. In this case, the goal was to better understand the changes induced in the soil metagenome by the introduction of an organism. This change in a biotic component of the community was again studied using metagenomics. We demonstrated that inoculation of cassava with the AMF *Rhizophagus irregularis* mostly increased gene richness and gene alpha diversity of the microbial metagenome compared to the mock-inoculated treatment while there was not an observable shift in the taxonomic richness and alpha diversity of the microbiome. In addition, the observed changes were different depending on the genetics of the *R. irregularis* isolates.

In chapter 3, as it was shown that inoculation with AMF changes the soil metagenome and that it has been shown to also change the abundance of local AM fungi (e.g. Akyol *et al.*, 2019), we aimed to develop easy use strain-specific markers to track isolates in the field. Small genomic regions containing polymorphisms were identified using ddRAD-seq reads, then, neighboring small regions were amplified together into longer sequences containing many more variable sites. The amplicon sequencing and posterior validation by cloning and Sanger sequencing revealed patterns of within isolate variation inconsistent with current knowledge about the genetics of these fungi.

In chapter 1, the collaborators of the project aimed understand broad geographic scale patterns of the impacts of novel plant-soil interactions on ecosystem processes. These novel interaction would arise from low elevation communities migrating upwards to track their current climate in a situation of global warming. Among the data collected, there is information about the changes in composition of the plant community to assess the pace and predictability of plant species establishment in transplanted turfs and measurements of carbon and nitrogen reservoirs to estimate the influence of novel plant-soil interactions on ecosystem nutrient cycling. Integrating all these data with the results of chapter 1 will provide valuable information concerning the interactions between plants and soil microbial communities in a climate warming scenario.

In different studies, along an altitudinal gradients, some of the factors found to be the main determinants of abundance and diversity of microbial communities were: pH and C/N (Siles & Margesin, 2016); plant diversity and density (Porazinska *et al.*, 2018); micro-topography (Frindte *et al.*, 2019); nutrient concentration (Bahram *et al.*, 2018); vegetation diversity (Yang *et al.*, 2014). This shows that the influence of climatic conditions and plant communities on soil microbial communities are difficult to disentangle.

Soil metagenome changes were observed in chapter 2 in a clonally propagated crop, thus, without alterations of the plant community. Despite this, it is reasonable to expect that in a climate warming scenario, soil metagenome will be influenced by changes in the plant community (Yang *et al.*, 2014; Porazinska *et al.*, 2018). Plant migration and microbe-mediated soil processes develop at different time scales (Ding *et al.*, 2015; Rumpf *et al.*, 2018; Walker *et al.*, 2018). It is likely that microbial communities will adapt faster to warmer conditions (Graham *et al.*, 2016), meaning that, in conjunction with the potential alteration of nutrient cycling (Walker *et al.*, 2018), it is belowground processes that are more likely to constrain the rate of vegetation changes in mountains following climate warming (Hagedorn *et al.*, 2019).

As mentioned above, microbial communities acclimate to warmer conditions. Notably, this acclimatization process is expected to occur without substantial changes in their community structure (Hagedorn *et al.*, 2019). However, as seen in chapter 2, the apparent lack of a taxonomic changes of the soil microbiome due to an environmental perturbation can mask underlying changes in the metabolic capabilities of the soil metagenome.

The experimental set up in chapter 2, where all the samples shared the climatic conditions, was used to study the effects of a change in a biotic element (i.e. effects of AMF inoculation). The approach followed in chapter 1, even if seen as a modification of abiotic conditions, undeniably implies also changes in the biotic components. In fact, this is one of the advantages of transplantation experiments (compared to, for example, open top chambers) as these biotic changes are very likely to occur in a global warming scenario. One of the most interesting findings in chapter 2 was the fact that gene richness, diversity and composition were observed without associated changes in the taxonomic composition of the microbial community.

We hypothesize that the observed decoupling of taxonomic and gene diversity could be due to horizontal movement of genes among bacterial taxa (thus spreading functions across taxonomic and phylogenetic barriers) in response to a change of environment (in this case inoculation with AMF). It is important to emphasize that non-inoculated treatments in chapter 2 are still mycorrhizal as this is a wide spread symbiont. It is expected that the abundance of the inoculated fungus will increase and it has been shown that this abundance augmentation is the most significant factor determining the observed plant responses (Niwa *et al.*, 2018). However, this was not measured in chapter 2. Additional to the importance of developing strain-specific markers for the inoculated fungus highlighted in chapter 3, the study of functional or taxonomic resilience in soil microbial communities has to consider that the responses will vary depending on whether the inoculation imposes a press (long-term) or pulse (short-term) disturbance (Shade *et al.*, 2012).

The time frame in which the disturbance (either abiotic or biotic) is assessed is a key factor as it can change the interpretation of the findings. In chapter 1, the time between the transplantation and the sampling varied across localities. However, it was mostly locality specific, thus when including the locality in the models, the variation in the so-called 'year range' was indirectly taken into account. In chapter 2, the time between the disturbance (i.e. AMF inoculation) and the sampling corresponded to the time in which the symbiotic association starts to have a positive effect on plant growth (Ceballos, 2016). The inoculation can constitute either a press disturbance (the inoculated AMF establishes and persists generating a long term effect) or a pulse disturbance (the inoculated AMF is outcompeted and the community starts a recovery process). To understand whether inoculation constitutes a press or pulse disturbance, it is necessary to be able to track the inoculated AMF. This will allow to obtain ecological insights on community stability and response to perturbations that cannot be gained otherwise.

The development of markers greatly relies on the accuracy of the sequencing information available. At the same time, the quality of the genome assembly is crucial for accurately identifying the polymorphism intended to be used in the markers. Early work towards a genome of the model AMF reported it as being an especially arduous challenge even after 4 years of work (Martin *et al.*, 2008). Later, the first published genome of an AMF had more than 12000 scaffolds (Tisserant *et al.*, 2013). Posterior efforts either using single nuclei sequencing or genomic DNA approaches reported about 30000 and 11000 scaffolds respectively (Lin *et al.*, 2014; Ropars *et al.*, 2016). This amount of scaffold indicates that these drafts are far from being complete genome assemblies. The characterization of polymorphism within AMF isolated need re-assessing once better genome assemblies become available. The development of markers presented in chapter 3 will greatly benefit of this information.

Even if strain-specific markers are not yet available, the results of chapter 2 present interesting worth developing. Among the enriched metabolic pathways found with the differential gene abundance approach, it is possible to find several times pathways related to the degradation of xenobiotics. The place where the experiment was set up is an agricultural field in which pesticides where used in the past. It has been shown how in the mycorrhizosphere (defined here as the zone of soil surrounding both the root and the extraradical mycorrhizal fungal hyphae) harbors not only a distinct bacterial community (e.g. Akyol *et al.*, 2019) but also stimulates microbial activity by releasing energy-rich organic compounds (Barea *et al.*, 2002). This can in turn increase the metabolic capacity of this communities (Uroz *et al.*, 2007). This could be developed into potential applications of microbial degradation of persistent soil pesticides accelerated using AMF.

References

Akyol TY, Niwa R, Hirakawa H, Maruyama H, Sato T, Suzuki T, Fukunaga A, Sato T, Yoshida S, Tawaraya K, et al. 2019. Impact of Introduction of Arbuscular Mycorrhizal Fungi on the Root Microbial Community in Agricultural Fields. Microbes and Environments 34: 23–32.

Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, et al. 2018. Structure and function of the global topsoil microbiome. Nature 560: 233–237.

Barea JM, Azcón R, Azcón-Aguilar C. 2002. Mycorrhizosphere interactions to improve plant fitness and soil quality. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology 81: 343–351.

Ceballos Rojas IC. 2016. Funcionalidad de la interacción simbiótica entre variedades de yuca y genotipos de *Rhizophagus irregularis* en la Orinoquía Colombiana. Doctoral Thesis. Available from http://bdigital.unal.edu.co/54313/7/IsabelCeballos2016.pdf.

Ding J, Zhang Y, Deng Y, Cong J, Lu H, Sun X, Yang C, Yuan T, Van Nostrand JD, Li D, et al. 2015. Integrated metagenomics and network analysis of soil microbial community of the forest timberline. Scientific Reports 5: 1–10.

Frindte K, Pape R, Werner K, Löffler J, Knief C. 2019. Temperature and soil moisture control microbial community composition in an arctic–alpine ecosystem along elevational and micro-topographic gradients. ISME Journal: 2031–2043.

Graham EB, Knelman JE, Schindlbacher A, Siciliano S, Breulmann M, Yannarell A, Beman JM, Abell G, Philippot L, Prosser J, et al. 2016. Microbes as engines of ecosystem function: When does community structure enhance predictions of ecosystem processes? Frontiers in Microbiology 7: 1–10.

Hagedorn F, Gavazov K, Alexander JM. 2019. Above- and belowground linkages shape responses of mountain vegetation to climate change. Science 365: 1119–1123.

Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DGO, Mu D, Pang E, Cao H, Cha H, Lin T, et al. 2014. Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. PLoS Genetics 10.

Martin F, Gianinazzi-Pearson V, Hijri M, Lammers P, Requena N, Sanders IR, Shachar-Hill Y, Shapiro H, Tuskan GA, Young JPW. 2008. The long hard road to a completed Glomus intraradices genome. New Phytologist 180: 747–750.

Niwa R, Koyama T, Sato T, Adachi K, Tawaraya K, Sato S, Hirakawa H, Yoshida S, Ezawa T. 2018. Dissection of niche competition between introduced and indigenous arbuscular mycorrhizal fungi with respect to soybean yield responses. Scientific Reports 8: 2–5.

Porazinska DL, Farrer EC, Spasojevic MJ, Bueno de Mesquita CP, Sartwell SA, Smith JG, White CT, King AJ, Suding KN, Schmidt SK. 2018. Plant diversity and density predict belowground diversity and function in an early successional alpine ecosystem. Ecology 99: 1942–1952.

Ropars J, Toro KS, Noel J, Pelin A, Charron P, Farinelli L, Marton T, Krüger M, Fuchs J, Brachmann A, et al. 2016. Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. Nature Microbiology: 16033.

Rumpf SB, Hülber K, Klonner G, Moser D, Schütz M, Wessely J, Willner W, Zimmermann NE, Dullinger S. 2018. Range dynamics of mountain plants decrease with elevation. Proceedings of the National Academy of Sciences of the United States of America 115: 1848–1853.

Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JBH, et al. 2012. Fundamentals of microbial community resistance and resilience. Frontiers in Microbiology 3: 1–19.

Siles JA, Margesin R. 2016. Abundance and Diversity of Bacterial, Archaeal, and Fungal Communities Along an Altitudinal Gradient in Alpine Forest Soils: What Are the Driving Factors? Microbial Ecology 72: 207–220.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, Frei dit Frey N, Gianinazzi-Pearson V, et al. 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proceedings of the National Academy of Sciences of the United States of America 110: 20117–22.

Uroz S, Calvaruso C, Turpault MP, Pierrat JC, Mustin C, Frey-Klett P. 2007. Effect of the mycorrhizosphere on the genotypic and metabolic diversity of the bacterial communities involved in mineral weathering in a forest soil. Applied and Environmental Microbiology 73: 3019–3027.

Walker TWN, Kaiser C, Strasser F, Herbold CW, Leblans NIW, Woebken D, Janssens IA, Sigurdsson BD, Richter A. 2018. Microbial temperature sensitivity and biomass change explain soil carbon loss with warming. Nature Climate Change 8: 885–889.

Yang Y, Gao Y, Wang S, Xu D, Yu H, Wu L, Lin Q, Hu Y, Li X, He Z, et al. 2014. The microbial gene diversity along an elevation gradient of the Tibetan grassland. ISME Journal 8: 430–440.