Year : 2017

Rôle of host and fungal intra-species genetic variability in the symbiosis between the arbuscular mycorrhizal fungus Rhizophagus irregularis and cassava (Manihot esculenta)

Mateus Gonzalez Ivan Dario

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Département d'écologie et d'évolution

Role of host and fungal intra-species genetic variability in the symbiosis between the arbuscular mycorrhizal fungus *Rhizophagus irregularis* and cassava (*Manihot esculenta*)

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

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

par

Ivan Dario MATEUS GONZALEZ

Master de l'Université Claude Bernard Lyon 1, France

Jury

Prof. Paul Franken, Président
Prof. Ian Sanders, Directeur de thèse
Prof. Thomas Flatt, expert
Prof. Daniel Croll, expert

Lausanne 2016

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. Paul **Franken**
Directeur · rice de thèse	Monsieur	Prof. Ian **Sanders**
Experts · es	Monsieur	Prof. Thomas **Flatt**
	Monsieur	Prof. Daniel **Croll**

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

Monsieur Ivan Mateus Gonzalez

Master de l' Université Claude Bernard de Lyon, France

intitulée

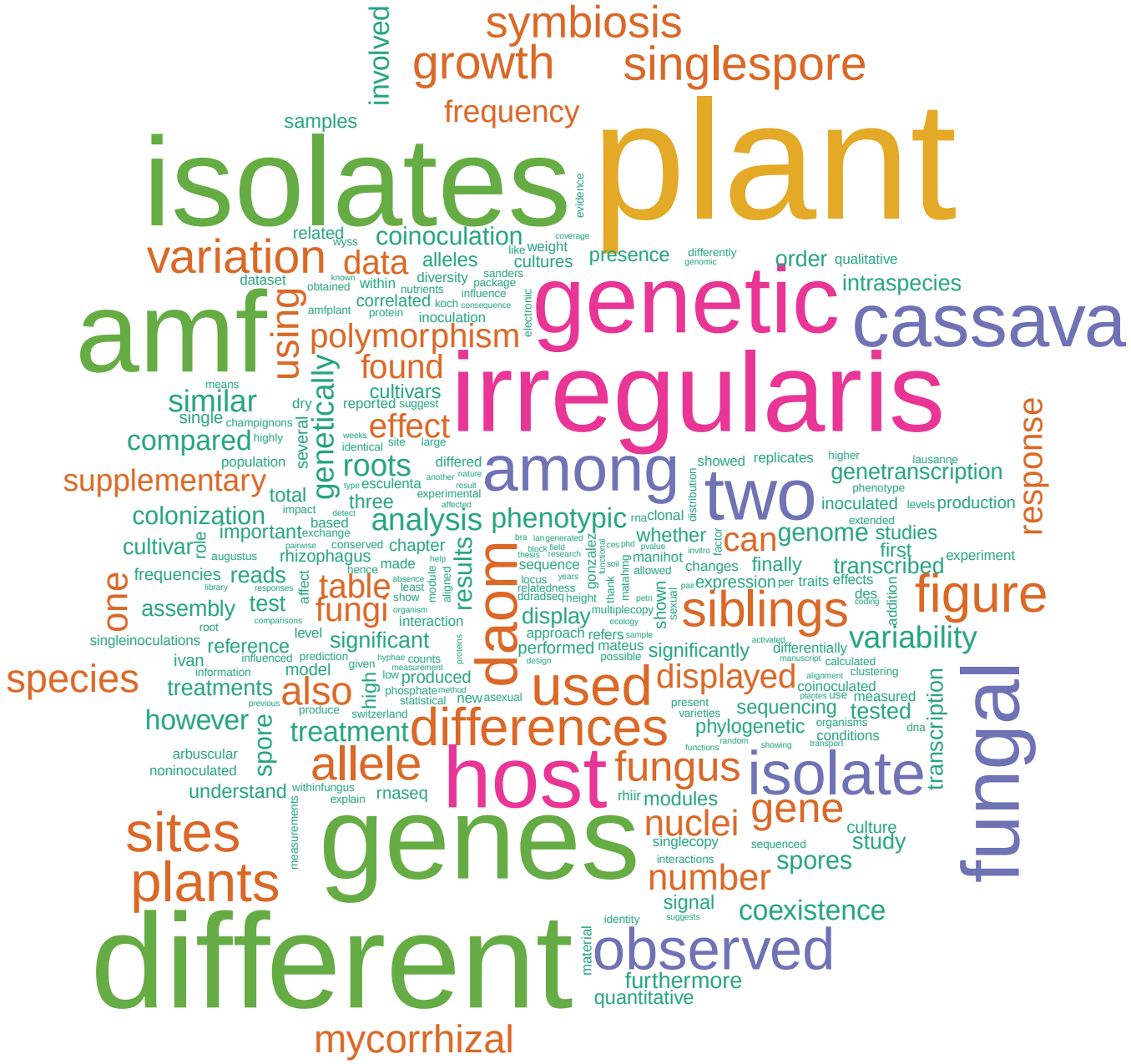
Role of host and fungal intra-species genetic variability in the symbiosis between arbuscular mycorrhizal fungus *Rhizophagus irregularis* and cassava (*Manihot esculenta*)

Lausanne, le 7 décembre 2016

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Paul Franken

Table of contents


symbiosis
involved
growth
singlespore
frequency
samples
plant
isolates
evidence
related
coinoculation
variation
data
alleles
cultures
presence
order
qualitative
intraspecies
genetic
cassava
amf
using
polymorphism
found
irregularis
similar
cultivars
effect
genetically
compared
among
two
response
roots
differed
showed
replicates
genetranscription
supplementary
phenotypic
inoculated
production
colonization
analysis
genome
studies
cultivar
important
whether
can
first
rhizophagus
manihot
transcribed
experiment
frequencies
reads
table
results
daom
siblings
figure
one
assembly
test
fungi
display
variability
species
significant
reference
used
displayed
fungal
treatments
also
produced
differences
however
treatment
phylogenetic
sequencing
tested
transcription
allele
host
fungus
isolate
gene
genes
nuclei
sites
plants
study
spores
number
different
observed
coexistence
signal
furthermore
quantitative
mycorrhizal

Acknowledgements

It has been five years since I started the adventure of a PhD thesis in the mycorrhizal symbiosis. It is very impressive to see the product of these five years of research in a PhD manuscript of 35000 words. Five years ago I did have very little knowledge in molecular biology, bioinformatics, fungi and plants. But now here we are, having fun designing experiments, asking new questions, making hypotheses, doing statistics, making publishable figures, writing, etc.

First and foremost I would like to thank my advisor Prof. Ian Sanders. I am grateful for his willingness to take me on in his lab without any prior knowledge of AMF fungi. He was confident that I could do a good job in the lab, and he gave me the liberty to surf between the different hot topics of AMF research. For me it was very important, the liberty and confidence, he has on me, I really appreciated that. I also would like to thank the lab members: Jeremy, Tania, Fred, Romain, Lucas, Cristian, Ricardo, Pawel and Consolée. Thanks for all the support, discussions and help during all this years. You were always there to discuss about AMF or other most important topics, it was really a pleasure to work with you guys.

I am very grateful with all the people that contributed and made this PhD possible. All the ones, which helped in culturing, pot the plants, making fungal colonizations, etc. Without your help this PhD would be much more difficult.

I also would like to thank the former lab members of the lab for advice at the beginning of this adventure: Caroline, Alex, Aurelien and Germain. Germain thanks for making the cold Switzerland a tropical environment for two years. I also would like to thank the Colombian group: Isabel, Daniela, Marcela and Alia for make me fell like in Colombia while you were here. I thank all the office mates; I appreciated all this important discussions that happened very often in the office. I also would like to thank all the people in the DEE that make this place a good place to work; the French speaking group, the English speaking group, the Spanish speaker group. I had a good time with all of you and also learned a lot. This was great from the DEE, the great diversity of very interesting people.

I would also like to thank my family that supported me on this adventure, you are the ones that make this happened. I am really grateful with all of you. Finally, I would like to thank Sarah that was there all this years, helping, cheering, and giving advice. I think you also deserved as me that this PhD comes to a happy end.

Abstract

Arbuscular mycorrhizal fungi (AMF) are strict endosymbionts of plant roots. They are important to the host plant because they provide nutrients to the host and have a large impact on plant growth. However, understanding how these fungi affect plant growth has been difficult because of the complexity of their genetic make-up and the difficulty to isolate these fungi. However, data from a population of the AMF *Rhizophagus irregularis* shows that genetic variation in the fungus occurs concurrently with variation in fungal phenotypic traits and in how plants grow. Different AMF can produce different effects on the host plant. As well, different host plants can be differently influenced by a single AMF.

In order to understand the effect of the fungal and host plant intra-species variability in different aspects of the symbiosis, we used the multi-locus sequencing techniques as restriction-site associated DNA sequencing (ddRad-seq) and RNA-seq on different isolates of *R. irregularis* and cassava (*Manihot esculenta*) cultivars to:

1) Confirm independently that single-spore siblings of two different *R. irregularis* isolates harbour different types of nuclei (Heterokaryosis), as shown by changes in allele frequency of several sites among the siblings.
2) Provide the first causal link between the genetic of the fungus and the plant response to the fungus. We show that genetically related *R. irregularis* isolates display similar phenotypes and similar effects on the host plant.
3) Show that the host plant intra-species variability influenced the fungal gene-transcription, that the fungal intra-species variability influenced the host plant gene-transcription and we reveal a first insight of gene-gene correlations between the host plant and the fungus.
4) Show that the host intra-species variability plays an important role in the coexistence of two *R. irregularis* isolates, that the coexistence of two *R. irregularis* isolates has different effects on different host plants and that two *R. irregularis* probably recognize each other.

Our results provide the first demonstrated link between genetic variation in the fungus, their phenotypic variation and plant growth response. We also reveal a first insight of gene-gene interactions between the fungus and the host plant. These results are essential to establish in our ultimate goal to use genetic variation in AMF to improve plant growth.

Résumé

Les champignons mycorhiziens (AMF) sont des endosymbiontes des racines des plantes. Ces champignons sont très importants pour les plantes hôtes car ils leur donnent des nutriments et donc provoquent un effet sur leur croissance. Comprendre comment ces champignons affectent la croissance des plantes a été très difficile car il est nécessaire d'isoler ces champignons du sol pour mieux étudier leurs effets. Malgré ces inconvénients, l'étude d'une population de champignons mycorhiziens a montré que des isolats de *Rhizophagus irregularis* ont des génotypes différents, des phénotypes différents et produisent différents effets sur leur plante hôte. Toutefois, il n'a jamais été prouvé que cette variation génétique des champignons soit responsable de la variation de la réponse des plantes à ces champignons.

Dans cette thèse, nous avons utilisé des techniques de séquençage à haut débit comme RAD-sequencing et RNA-seq sur des échantillons issus de différents isolats de *R. irregularis* ainsi que sur plusieurs variétés de manioc (*Manihot esculenta*) afin d'atteindre les buts suivants :

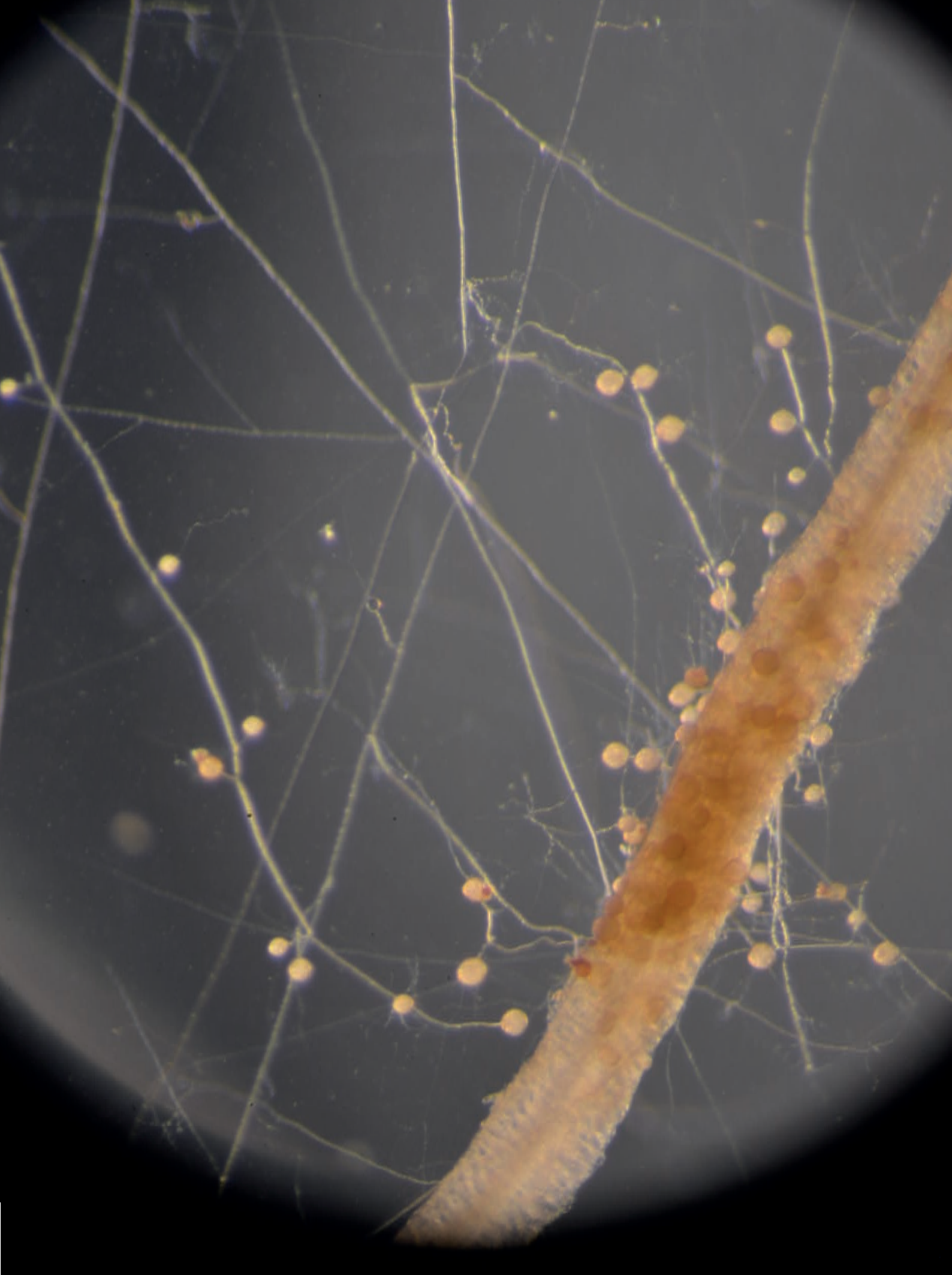
1) J'ai démontré que la descendance asexuelle d'un isolat présente différents types de noyaux. Ces différences sont probablement dues à des changements de fréquence des allèles au sein de la descendance.

2) J'ai démontré que la variation génétique du champignon est responsable de sa variation phénotypique ainsi que de la variation de la réponse des plantes aux champignons.

3) J'ai démontré que la variation intra-spécifique de la plante hôte influence l'expression de plusieurs gènes du champignon, et que la variation intra-spécifique du champignon influence l'expression de plusieurs gènes de la plante hôte. J'ai également démontré que l'expression de plusieurs gènes de la plante hôte est corrélée à l'expression de plusieurs gènes du champignon.

4) J'ai démontré que la variation intra-spécifique de la plante hôte joue un rôle important dans la coexistence de deux isolats de champignons, que la coexistence de deux isolats influence la croissance de la plante hôte, et que probablement deux isolats se reconnaissent entre eux.

Ces résultats sont les premiers à démontrer qu'il existe un lien de causalité entre la variation génétique du champignon, la variation phénotypique et la variation dans la réponse de la plante au champignon. Celles-ci sont des découvertes très importantes car elles vont nous permettre d'utiliser la variation génétique des champignons pour améliorer la croissance des plantes.



Carrot root with hypha and spores of *R. irregularis*
Jeremy Bonvin

Introduction

Arbuscular mycorrhizal fungi (AMF)

Arbuscular mycorrhizal fungi are strict endosymbionts of plant roots. They provide nutrients such as phosphate and nitrate to the host plant in exchange for plant carbohydrates (Johnson *et al.* 1997). The mycorrhizal symbiosis is one of the oldest known symbioses and is thought to have evolved ~460 Million years ago (Redecker 2000). More than 80% of plants are thought to form the mycorrhizal symbiosis. This exchange of nutrients is facultative for plants but obligate for the fungi. AMF represent cosmopolitan taxa and intensive sampling of different taxa on different locations and soil types suggest that these microbes are present in all soils (Hazard *et al.* 2013).

Biology of AMF

Arbuscular mycorrhizal fungi are coenocitic (presenting no cell divisions). This probably allows the cytoplasm, mitochondria and nuclei to flow around the organism. During spore formation, different numbers of nuclei migrate into the spores. The consequence of this is the lack of a single-nucleus state at any time in the life cycle of the fungus. Furthermore, multi-nucleate hyphae can fuse with each other. This process is known as anastomosis and it has been described by (Giovannetti et al., 1999). Hyphae from the same, and also from genetically different isolates of the same AMF species, can fuse and exchange cytoplasm (Croll *et al.* 2009), consequently allowing the mixture of genetic material between the two fungi.

Arbuscular mycorrhizal fungi have been considered as ancient asexual scandals, because they are one of the oldest taxa that are thought to be asexual (Judson & Normark 1996). Indeed, no sexual structures have been described in AMF. As described previously, asexual recombination by means of anastomosis can allow genetic exchange among different hyphae. Confirming this, Croll et al., (2009) found signals of recombination in sequences of DNA in an AMF species. A genome survey of meiosis related genes in the genome of *R. irregularis* and transcriptome of different Glomeromycota species, showed the presence of several meiosis related genes, providing evidence of possible cryptic sex in these fungi (Halary *et al.* 2011; Tisserant *et al.* 2012). Furthermore, several homologues of mating type loci found in other fungal species have been found in the genome of *R. irregularis* (Riley *et al.*

2014). Finally, it has been shown that *R. irregularis* isolates can possess different alleles of a putative mat-locus and even some isolates contain combinations of two alleles of this mat-locus suggesting, without complete evidence, that *R. irregularis* posses and homokaryotic- dikaryotic life cycle (Ropars *et al.* 2016). These results together suggest that AMF should probably no longer be considered as ancient asexuals, like other taxa that were believed ancient asexuals, but display rare sex events (Croll 2016).

AMF within-species diversity

Genetic studies on single species of AMF have been possible because of the existence of *in-vitro* cultures. Notably, *Rhizophagus irregularis* has become the model species for genetics studies for its worldwide distribution and the ease of putting it into culture. *R. irregularis* is reported to be a very diverse AMF. Within the same location, a population of *R. irregularis* comprises isolates that are genetically different from each other (Koch *et al.* 2004). These different genotypes present differences in traits such as hyphal density, spore density and spore size (Koch et al., 2004; Ehinger et al., 2012). Moreover, the same species has been used to inoculate globally important crops such as rice and cassava, showing that different AMF isolates can produce different host growth responses (Angelard et al., 2010; Ceballos et al., 2013).

Genomic organization of AMF

The multinucleate status of these fungi could be an important source of genetic variation in AMF. However, this depends on whether the fungi contain nuclei that are identical (homokaryons) or that are different among each other (heterokaryons). Homokaryosis is the state where an organism possesses nuclei that are all identical. Heterokaryosis is the state where an organism possesses 2 or more nuclei that are genetically different and co-existing in the same cytoplasm. Some AMF, such as *Scutellospora pellucida* (Bever & Morton 1999), *Scutellospora castanea* (Kuhn et al., 2001) and *Glomus etunicatum* (Hijri & Sanders 2005) have been described as heterokaryons.

There is a considerable debate about if the nuclei found in the cytoplasm of *R. irregularis* are genetically identical or different. Tisserant et al., (2013) have published a 153-Mb genome of the *R. irregularis* isolate DAOM-197198. They compared sequenced regions containing more than 1000 bp

and they found very low polymorphism among the regions. Furthermore, Lin et al., (2014) have also published 4 single nucleus genomes of DAOM-197198. They compared 3 genes on the 4 nuclei with the objective of finding genetic polymorphism. However, they observe that there was no allelic variation among the nuclei. Subsequently, these two papers suggest that *R. irregularis* does not have multiple genomes and presents low polymorphism (Tisserant *et al.* 2013; Lin *et al.* 2014). On the other hand, isolates from the same species isolated from an agricultural field in Switzerland presented a high genetic polymorphism that suggests heterokaryosis. Isolates issued from crosses of individual spores (segregated lines) present different allele frequencies and have different effects on host plant growth (Angelard *et al.* 2010). Moreover, Ehinger et al., (2012) have shown in an independent study on *R. irregularis* (isolate C3) that single-spore sibling lines issued from the same parental culture, present different allele frequencies and germination-tube growth rate. In conclusion, these studies on the Swiss *R. irregularis* population are not in agreement with the studies made on the reference isolate DAOM 197198 concerning the genomic organization of *R. irregularis*. However, in 2016 two independent studies confirmed the heterokaryotic nature of *R. irregularis*. Restriction site-associated DNA sequencing (RAD-seq) was performed on 20 different isolates, showing a high level of polymorphisms that suggest the heterokaryotic nature of *R. irregularis* (Wyss *et al.* 2016). Furthermore, 6 different isolates of this species were sequenced using whole genome sequencing, showing that some isolates could be homokaryons, while others could be heterokaryons, as described by a drop in 50% of coverage in some genomics regions (Ropars *et al.* 2016).

Symbiosis functioning

In order to establish the symbiosis with the host plant, a molecular communication is made between the fungal endosymbiont and the host plant (Figure 1). First, the root produces exudates, as signal-molecules, which induce germination and hyphal branching of the fungus (Giovannetti *et al.* 1993; Buee *et al.* 2000). The molecules involved are phyto-hormones, among them strigolactones, which are responsible for the fungal branching (Akiyama *et al.* 2005). Second, the fungal hyphae form a structure in the root surface called hyphopodium. This fungal structure appears to be induced by a conserved mechanism in pathogens and symbiotic fungi (Gutjahr & Parniske 2013), which could be triggered by the perception of cutin monomers in the plant cell-wall (Dickman *et al.* 2003). The fungal

hyphae produce chitin molecules (chitooligosacharides) that induce the expression of several plant genes including calcium spiking (Genre *et al.* 2013).

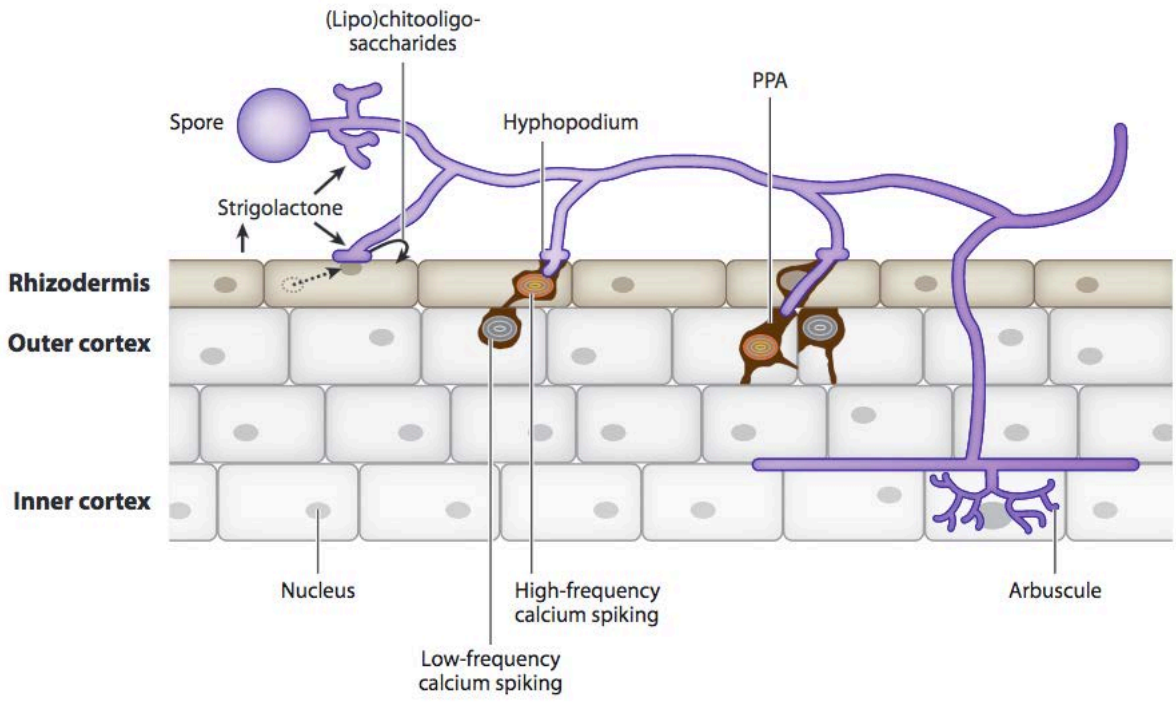


Figure 1. Schematic representation of the interaction between AMF and the host plant. The host plant exude strigolactones which induce the germination of the spores. The fungus produces (lipo)chitooligosaccharides that induces a calcium spiking of the host plant cells. AMF fungi form a hyphopodium structure in the surface of the rhizodermis. The host plant forms a prepenetration apparatus (PPA) that allows the fungal hyphae enter towards the roots inner cortex cells. In the inner cortex cells the fungus form arbuscules, which are the nutrient exchanging structure between the fungus and the host plant. Figure from Gutjahr & Parniske, 2013.

The plant response to fungal chitin molecules is dependent of the LysM receptor-like kinase that controls the formation of AM symbiosis (Op den Camp *et al.* 2011). After the formation of the hyphopodium, the plant forms its prepenetration apparatus (PPA) that shows the crossing path to the fungus inside the plant cells (Parniske 2008). Finally, the fungus forms 'arbuscules' in the inner roots cortical cells, where the nutrient transfer occurs. The plant develops a structure that surrounds the arbuscules, made by the plant endoplastic reticulum called peri-arbuscular membrane. The peri-arbuscular membrane is the plant structure that is involved in nutrient transfer from the plant side, and the peri-arbuscular space (between peri-arbuscular membrane and arbuscules) is the place where the exchange takes place between the two organisms.

Plant control over AMF

The host plant responds to fungal pathogens by producing salicylic acid, which is a molecule involved in a signal transduction pathway activated when the fungal hypha is in contact with the host plant (Blilou *et al.* 2000). However, when interacting with AMF the host plant can elicit a transient defence response as against fungal pathogens but with less intensity (García-Garrido & Ocampo 2002). Furthermore, phosphate inhibits the development of AMF by repressing essential symbiotic genes when exposed to high levels of exogenous phosphate(Breuillin *et al.* 2010). Finally, arbuscular mycorrhizal fungi are known to benefit the host plant by providing nutrients. However, it also has been observed that AMF can have a negative impact on host plant growth (Koide 1985). This demonstrates that AMF symbiosis effects are more situated in a mutualism-parasitism continuum (Johnson *et al.* 1997) rather than be only beneficial for the plant. Furthermore, the host plant possesses the mechanisms to avoid fungal strains that are less beneficial. It has been demonstrated, using radiolabeled phosphate and carbohydrate molecules, that plants give less carbohydrates to fungi that provide les phosphates to the host plant (Bever *et al.* 2009; Kiers *et al.* 2011).

AMF regulation of the symbiosis

It has been shown that AMF can interfere with the host defence signalling (Volpin *et al.* 1995), showing that AMF can also have a mediation role in the host-AMF symbiosis. AMF can also regulate the symbiosis by producing small molecules (Effector proteins) that will affect the defence response of the host plant (Kloppholz *et al.* 2011). Effector proteins are molecules that interfere with numerous functions to overcome the host plant defences (Kamoun 2006). Effector molecules appear to be conserved within the Glomeromycotan phylum (Sędzielewska Toro & Brachmann 2016), suggesting that this is an ancient mechanism in the plant-AMF interaction.

Nutrient exchange between the two organisms

Nutrient exchange between AMF and the host plant is a key feature to understand the symbiosis between the two partners. AMF provide nutrients to the plant such as nitrate and phosphate in exchange for plant carbohydrates (Johnson *et al.* 1997). Furthermore, AMF can enhance plant resistance to heavy metal polluted soils and protect the plant against water stress. This means that sugar, phosphate, nitrates, heavy metals and water transport should play an important role in the

symbiosis. Evidence of sugar export to the fungus has been shown by the expression of the fungal sugar transporter *MST2*, which is paired with the activity of the phosphate transporter *PT4* (Helber *et al.* 2011). Furthermore, the plant *SWEET* sugar transporter family has been shown to be also involved in the export of sugar to AMF (Manck-Götzenberger & Requena 2016).

Phosphate is a limiting nutrient for plants. Consequently it could be a main driver in the AM symbiosis. The host plant can import phosphate from AMF by the phosphate transporter *PT4*, which was first discovered in *Medicago trunculata* roots in presence of AMF (Harrison *et al.* 2002). Nitrogen has also been detected to be another nutrient that is transferred from the soil and allocated to the host-plant by AMF (Govindarajulu *et al.* 2005). The transfer of nitrate from AMF to the roots is made trough the nitrate transporter *NTR2* (Hildebrandt *et al.* 2002). Ammonium, which is another source of nitrogen, is imported into the host plant by the ammonium transporter *AMT2* (Guether *et al.* 2009). Potassium is another limiting nutrient for plants, which evidence of transfer from AMF has been shown in *Zea mays* (Kaldorf *et al.* 1999). Transporter family *TrK/Ktr/HKT* is involved in the potassium transport from the plant side (Corratgé-Faillie *et al.* 2010). Metals as zinc, cooper and Iron are also important nutrients for plant growth. However, they are toxic in high concentrations. AMF increase metal transport in low metal conditions, but decrease metal transport when the soil has toxic levels of metals (González-Guerrero *et al.* 2016). It has been shown that *ZIP* transporters, which are an ubiquitous family of metal transporters, are down regulated in presence of AMF (Burleigh *et al.* 2003). Finally, aquaporins are involved in water movement in plants. The downregulation of the aquaporins *PIP1* and *TIP* in presence of AMF suggest that AMF regulate plant aquaporin expression, and hence, the water flow in the plant (Ouziad *et al.* 2005). All the previous examples showed that the host plant has a large arsenal of transporters to achieve the exchange of molecules with AMF. However, this list is only a small number of transporters that have been reported and there are numerous homologues and other molecules that could be also involved specifically in the exchange of nutrients in the peri-arbuscular space.

Cassava (*Manihot esculenta*)

Cassava is considered a vital crop for food security that provides an important source of calories in tropical countries (Howeler 2013). This crop possesses a good capacity to resist drought stress (El-Sharkawy & Cock 1987). Few studies have tested the effect of AMF on the growth of cassava.

Cassava has been shown to respond strongly to AMF (Sieverding & howeler 1985). Furthermore, AMF have been shown to improve drought tolerance of cassava plants under water stress conditions (Oyetunji & Osonubi 2007). However, to better understand the mechanisms concerning the effect of AMF on Cassava, a good quality reference genome is needed. In 2016, a complete cassava reference genome (751 Mb, 97% of genes distributed in 18 chromosomes) was released (Bredeson *et al.* 2016). This genome has a contiguity of 27.7 kb (N50). The low diversity of chloroplast DNA suggests that the diversity in cassava is the result of a maternal bottleneck during domestication. However, it was found that introgression of at least 2 other *Manihot* especies on cassava occurred, having an impact on cassava genetic diversity (Bredeson *et al.* 2016).

Effect of AMF on plants

By providing nutrients to the host plant, AMF can have an important role in plant ecology. AMF colonize approximately 200'000 plant species (van der Heijden *et al.* 2015). It has been shown that AMF can affect plant community diversity (van der Heijden *et al.* 1998), and AMF identity affect plant growth (Koch et al., 2006). Furthermore, it has been shown that AMF can protect the host plant against, drought stress (Oyetunji & Osonubi 2007) and heavy metal toxicity (Ouziad *et al.* 2005). Finally, it has been shown that AMF can affect the host defence against pathogens, by producing a priming effect on the host plant (Jung *et al.* 2012). All these examples show the extent of effects that AMF could have on plant ecology.

Genetic variability and its effect on the host plant

Several results shown that AMF can have a positive impact on plant growth (Smith & Read 2008). However, there are also some results where AMF has no effect or even has a negative effect on plant growth (Koide 1985; Johnson *et al.* 1997; Grace *et al.* 2009; Veiga *et al.* 2011). All these evidence shows that there is a genotype-by-environment interaction in the AMF-plant symbiosis (Ehinger *et al.* 2009). Furthermore, an experiment using different grass plants single-inoculated with different species of AMF, showed that the mycorrhizal effect is dependent on the AMF species and the host plant species (Klironomos 2003), highlighting the complexity of the plant response to the AMF. Finally, the amount of genetic variability found at the within-species level in *R. irregularis* has resulted in different phenotypic responses of the host plant (Angelard *et al.* 2010). All this evidence suggests an important

role of fungal genetic variability on the response of the host plant to AMF. However, if fungal genetic variability creates important variation on the plant growth response to AMF, it is possible that the response of the host plant to AMF could be predictable. Then, the fungal genetic variability could be used as a tool to understand the host response to AMF.

PhD objectives

This PhD thesis aims to understand the role of the host and fungal intra-species variability in different aspects of the symbiosis. I aimed to understand the genetic make up of *Rhizophagus irregularis,* the interaction of different *R. irregularis* isolates in the same environment and the interaction and effect of *R. irregularis* on the host plant.

First, I tested whether the coexisting nuclei in *R. irregularis* are equal (homokaryosis) or different (heterokaryosis). I produced single-spore siblings of two *R. irregularis* isolates, I genotyped them by using the multi-locus sequencing approach ddRad-seq and tested if the single-spore siblings were phenotypically different, if they displayed qualitative genetic changes (presence/absence of alleles at a single locus) and if they displayed quantitative genetic changes (differences in allele frequency at a single locus).

Second, I tested if there is a genetic basis between the genetics of the fungus and the plant response to AMF inoculation. I grew different *R. irregularis* isolates and used them to inoculate cassava plants. I tested whether genetically related isolates displayed similar phenotypes and produced similar responses on the host plant.

Third, I evaluated the role of the fungus intra-species variability and host intra-species variability in the AMF symbiosis. I performed RNA-seq on two different *R. irregularis* isolates, inoculated independently, to five different cassava cultivars, to test: 1) if there are differences in gene-transcription between the two *R. irregularis* isolates. 2) if the two *R. irregularis* isolates produced different gene-transcription responses on the host plant. 3) if the host plant intra-species variability affects the gene-transcription of the *R. irregularis* isolates. 4) If the expression of *R. irregularis* genes can be correlated to the transcription of the cassava genes. 5) if the gene-transcription of cassava or *R. irregularis* genes correlates to the growth response of the plants.

Fourth, I tested whether the host plant identity has an effect on the coexistence of two different *R. irregularis* isolates within the same host roots by using RNA-seq. I also investigated the effect of the coexistence of two different *R. irregularis* isolates on the host growth response. Finally, I tested if the two different *R. irregularis* isolates interact directly to each other. I used two different isolates single-inoculated and co-inoculated on three different cassava cultivars.

Author's contributions

Chapter 1: Ivan Mateus Gonzalez and Ian Sanders designed the experiment. Ivan Mateus Gonzalez performed the experiment and the statistical analysis. Ivan Mateus Gonzalez and Consolée Aletti performed the ddRAD-seq libraries. Ivan Mateus Gonzalez, Ian Sanders, Fred Masclaux and Tania Wyss discussed the results and commented the manuscript. Ivan Mateus Gonzalez wrote the manuscript.

Chapter 2: Ivan Mateus Gonzalez designed the experiment. Ivan Mateus Gonzalez and Romain Savary performed the experiment. Ivan Mateus Gonzalez performed the statistical analysis. Fred Masclaux and Tania Wyss provided unpublished ddRAD-seq data. Ivan Mateus Gonzalez, Ian Sanders, Fred Masclaux and Tania Wyss discussed the results and commented the manuscript. Ivan Mateus Gonzalez wrote the manuscript. *This chapter represents a short report containing the analysis of the phylogenetic signal in the fungal phenotype and plant response to inoculation with R. irregularis. These analyses will be part of a larger manuscript including data on cassava responses to different genetic lines of R. irregularis in the field and genetic analysis of the genetic differences among the lines. The other parts of that manuscript were not my work and, therefore, not presented here.*

Chapter 3: Ivan Mateus Gonzalez designed the experiment. Ivan Mateus Gonzalez, Edward Rojas and Romain Savary grew the plants, measured plant growth and colonization. Ivan Mateus Gonzalez performed the RNA extraction and RNA-seq libraries with the help of Consolée Aletti. Ivan Mateus Gonzalez performed the statistical analysis. Ian Sanders and Ivan Mateus Gonzalez wrote the manuscript. Fred Masclaux helped in the statistical analysis and revision of the manuscript.

Chapter 4: Ivan Mateus Gonzalez designed the experiment. Ivan Mateus Gonzalez, Edward Rojas and Romain Savary grew the plants, measured plant growth and colonization. Ivan Mateus Gonzalez performed the RNA extraction and RNA-seq libraries with the help of Consolée Aletti. Ivan Mateus Gonzalez performed the statistical analysis. Ivan Mateus Gonzalez wrote the manuscript. Fred Masclaux and Ian Sanders helped in the revision of the manuscript.

R. irregularis over carrot root tip

Jeremy Bonvin

Chapter 1. Quantitative genetic and phenotypic differences generated among clonal siblings of the symbiotic fungus Rhizophagus irregularis

Ivan D. Mateus-Gonzalez[1], **Frédéric G. Masclaux**[1,2] , **Tania Wyss**[1], **Consolée Aletti**[1] **& Ian R. Sanders**[1]

[1]*Department of Ecology and Evolution, University of Lausanne, Biophore building, 1015 Lausanne, Switzerland*
[2] *Vital-IT, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*

Abstract

Asexual organisms are at a disadvantage compared to sexual organisms because of the lack of recombination allowing the generation of genetic variation. However, the multinucleated arbuscular mycorrhizal fungi (AMF) produce genetically variable clonal progeny that can greatly affect plant growth. This suggests a potentially high ecological impact of such variation. We aimed to understand the nature of the within-fungus genetic polymorphism in clonally produced progeny of the AMF species *Rhizophagus irregularis*. We observed that clonal single-spore siblings of two isolates did not display qualitative genetic differences (i.e. presence or absence of alleles) but displayed significant quantitative genetic differences observed as differences in allele frequency. These quantitative genetic changes were distributed genome-wide rather than located in specific regions. The results were consistent using different reference genome assemblies. We also observed that the single-spore siblings differed phenotypically. These results suggest that variation in structural arrangements of DNA sequences and the possession of genetically different nuclei are the most likely explanations for the within-fungus genetic polymorphism. Our results highlight the importance of possessing different nuclei as a mechanism to generate genetic variation that could have a consequence on the phenotype of a clonal organism.

Introduction

Arbuscular mycorrhizal fungi (AMF) are strict endosymbionts of plant roots that form symbioses with approximately 200'000 different plant species (van der Heijden *et al.* 2015). They provide the plant with nutrients such as phosphate and nitrate in exchange for plant-derived carbohydrates. Arbuscular mycorrhizal fungi affect plant growth (Koch *et al.* 2006), and plant community diversity and productivity (Heijden *et al.* 1998), suggesting an important role of these fungi in plant ecology. They also have the potential to be used as 'biological fertilizers' to improve crop production (Ceballos *et al.* 2013). These fungi have been considered as one of the oldest taxa that are thought to be asexual (Judson & Normark 1996). To date, no sexual structures have been described in AMF. Despite the presence of meiosis genes in AMF genomes that are highly conserved in eukaryotes (Riley & Corradi 2013), the evidence of nuclear exchange between different hyphae (Croll *et al.* 2009), and recently the discovery

of mating-type like alleles in *R. irregularis* (Ropars *et al.* 2016), no obvious stage of meiosis has ever been observed.

From an evolutionary perspective, clonal organisms should be at disadvantage compared to sexual organisms because of the inability to generate genetic diversity by recombination and the accumulation of deleterious mutations over generations; a process known as Muller's Ratchet (Felsenstein & Yokoyama 1976). Nevertheless, clonal reproduction is common in bacteria, fungi, plants and animals, demonstrating a wide presence in the tree of life. In AMF, siblings issued from asexual spores that did not have the possibility to recombine with other isolates exhibited different phenotypes (Ehinger *et al.* 2012) and produced strongly differential effects on host plant growth (Angelard *et al.* 2010). This indicates that there could be a mechanism in AMF allowing the generation of phenotypic and functional diversity in clonally produced asexual offspring and this could affect their success in different environments as well as affecting their plant hosts.

Arbuscular mycorrhizal fungi are coenocytic, meaning that they have a continuous cytoplasm without the separation of nuclei into individual cells. This could allow nuclei to move around freely in the mycelium. In addition, during spore development, several nuclei migrate into spores (Marleau *et al.* 2011). The consequence of this is the lack of a known single-nucleus stage at any time in the life cycle of AMF (Sanders & Croll 2010). The generation of small differences among nuclei, and the inheritance of several nuclei at each generation, could result in the maintenance of a heterokaryotic state (coexistence of different nuclei). Furthermore, segregation of genetically diverse nuclei into individual spores could occur and influence spore shape (Bever & Morton 1999). Thus if AMF nuclei are not identical, the multinucleate status of these fungi could be an important source of genetic variation in a seemingly clonal organism.

There is a debate about whether the nuclei of AMF are genetically identical (homokaryotic) or genetically different (heterokaryotic). Tisserant *et al.* (Tisserant *et al.* 2013) published a 115 Mb genome sequence of *Rhizophagus irregularis* (isolate DAOM 197198) where low levels of polymorphism within the genome were found. Furthermore, Lin *et al.* (Lin *et al.* 2014) sequenced the genomes of four single nuclei of the same *R. irregularis* isolate. They compared the genomes of the

four nuclei and observed that there was very low allelic polymorphism among the nuclei. Consequently, these two studies suggest that *R. irregularis* (DAOM 197198) does not harbour genetically different nuclei and exhibits low within-fungus polymorphism (Tisserant *et al.* 2013; Lin *et al.* 2014). However, some isolates of the same AMF species from a population in Switzerland presented a higher level of within-fungus genetic polymorphism (Wyss *et al.* 2016). This polymorphism was evaluated by analysing the SNP density in short sequences of the genomes of different isolates of *R. irregularis*, using the multi-locus sequencing approach called double-digest restriction site associated DNA sequencing (ddRAD-seq) (Parchman *et al.* 2012). Polymorphism within *R. irregularis* isolates was ascribed to either genetic differences among nuclei or polymorphism caused by copy-number variations (CNV) between the different isolates (Wyss *et al.* 2016). Copy-number variation has been proposed as a mechanism that can increase genetic variability and cause phenotypic variation by modifying gene expression levels and affect individual survival (Tang & Amon 2013). This type of polymorphism, manifested as a variation in the structural arrangement of DNA sequences, has previously been shown in an AMF population (Corradi *et al.* 2007). Hence, this type of structural arrangement could be an alternative to the heterokaryosis explanation for the existence of within-fungus genetic polymorphism in AMF. Furthermore, *R. irregularis* (isolate DAOM 197198) is haploid, as determined by flow cytometry measurements of nuclear DNA content that were compared to genome size estimates from sequencing data (Sedzielewska *et al.* 2011; Tisserant *et al.* 2013). As a consequence, polyploidy is not a likely explanation of the within-fungus polymorphism of *R. irregularis* isolate DAOM 197198.

Several studies attempting to understand the within-fungus genetic polymorphism in *R. irregularis* adopted a particular experimental design, but the nature of the within-fungus genetic polymorphism remains unclear (Taylor *et al.* 2015). The design involved producing new cultures, each from a single-spore of a given AMF parental isolate, and then looking to see if those progeny were genetically identical or different. Qualitative measurements (i.e. presence or absence of alleles), and quantitative measurements (i.e. differences in frequency of alleles) among the siblings were used as proxies to determine if the coexisting nuclei were identical or different. The assumption in these experiments was that if the nuclei were all the same, then all clonal progeny would also be genetically identical.

However, if nuclei were different then segregation of genetically different nuclei could lead to observable differences among siblings.

Using the above-mentioned experimental design, three studies attempted to detect the source of within-fungus genetic polymorphism in AMF. However, each study has serious limitations that could compromise their conclusions. First, Pawlowska *et al.* analysed the *POL1*-like sequence (PLS1) locus in single-spore siblings of the AMF *Glomus etunicatum*, showing the presence of all 13 variants of the PLS1 site in all siblings (Pawlowska & Taylor 2004), and concluded that all nuclei were identical. However, the authors analysed a single locus, and did not look for quantitative differences among the single-spore siblings. Furthermore, it cannot be excluded that the parental isolate possessed genetically different nuclei, but that all the different nuclei were transmitted in equal frequencies to the single-spore siblings; thus, generating identical heterokaryotic siblings that were qualitatively the same. Second, Angelard *et al.* used isolates that were crosses between two parental isolates to produce single-spore siblings. The single-spore siblings exhibited differences in the presence/absence of alleles at multiple sites using AFLP, and differed in relative allele frequency at a locus, *Bg112* (Angelard *et al.* 2010), supporting the hypothesis that different nuclei coexisted. However, the use of progeny originating from crossed AMF isolates could have inflated the observed diversity because it could be influenced by the disproportionate inheritance of nuclei originating from the different parental isolates. In addition, the authors used AFLP; a method that does not allow sequence verification. Additionally, only 1 locus was used to measure allele frequency differences. Third, Ehinger *et al*. showed changes in the relative frequency of alleles at one site (*Bg112* locus) among single-spore siblings (Ehinger *et al.* 2012), supporting the hypothesis that this AMF comprised genetically different nuclei. The premise in these two last studies is that differences in relative frequency of alleles could only have occurred if each nucleus contained one of the alleles and the progeny inherited different proportions of the nuclei carrying each allele. But the authors did not look for qualitative genetic differences and also only analysed a single locus as a measurement of quantitative differentiation. In those last studies, allele frequency differences among siblings were only observed at the *Bg112* locus, which Lin *et al.* suggest is an inappropriate marker for understanding whether genetic variation is partitioned among nuclei or not (Lin *et al.* 2014). An alternative explanation is that the nuclei harboured variable numbers of copies of the alleles; thus differences in allele frequency are due to the

structural arrangement of the alleles among nuclei. However, in either scenario, it is difficult to explain differences in allele frequency among siblings if the nuclei do not harbour either different alleles or different numbers of copy of the alleles. The only other possible explanation is that some unknown mechanism causes a change in allele copy number in all nuclei during spore formation; although this is not a highly parsimonious explanation, as no such mechanism is known.

To understand the generation of genetic variation during clonal propagation of this fungus we developed a workflow that could allow us to discriminate between the different possibilities explaining its genomic organization (figure 1*a*). Hence, four possible scenarios could be expected from the experimental design where clonal single-spores siblings are produced from the same parental isolate:

(1) The clonal siblings, each initiated from a single asexual spore, are genetically identical. Thus, no qualitative or quantitative differences are observed among siblings (figure 1*b*-1).

(2) The siblings inherit the same composition of genetically different nuclei. Therefore, the single-spore siblings do not display either qualitative or quantitative differences, even though they are heterokaryotic (figure 1*b*-2).

(3) Single-spore siblings display qualitative genetic differences, detected as the presence and absence of alleles at some positions. By chance, some nuclear genotypes are lost in some single-spore siblings giving rise to qualitatively genetically different siblings. Random loss of alleles is only likely to happen in one generation if some nuclear genotypes are present in low frequency (figure 1*b*-3).

(4) Single-spore siblings display quantitative genetic differences.

    a. The observed differences in relative frequency of alleles are found at single-copy sites among siblings. The nuclei differ genetically and each new spore receives the same nuclei but in different relative frequencies. If the knowledge of the genome of the tested isolate is good, then we can assume that such differences are at true single-copy sites, where each locus only exists once in each nucleus. Thus, changes in allele frequency should be due to a change in the relative frequency of nuclei carrying a given allele at a given locus (figure 1*b*-4*a*).

    b. The differences in the relative frequency of alleles observed at multiple-copy sites could be the result of CNVs generated during spore development (by an unknown

mechanism) or by existing differences among nuclei and the subsequent unequal inheritance of these nuclei among siblings. In this case, the nuclei are divergent and the CNV polymorphism could explain the within-fungus genetic polymorphism. Additionally, if sequence reads are aligned to a reference assembly of a different AMF isolate, or aligned to a poorly assembled genome, sites assumed to be single-copy in the reference assembly could be in fact multiple-copy; hence the quantitative differences in allele frequency among siblings could be explained by CNV among nuclei (figure 1*b*-4*b*).

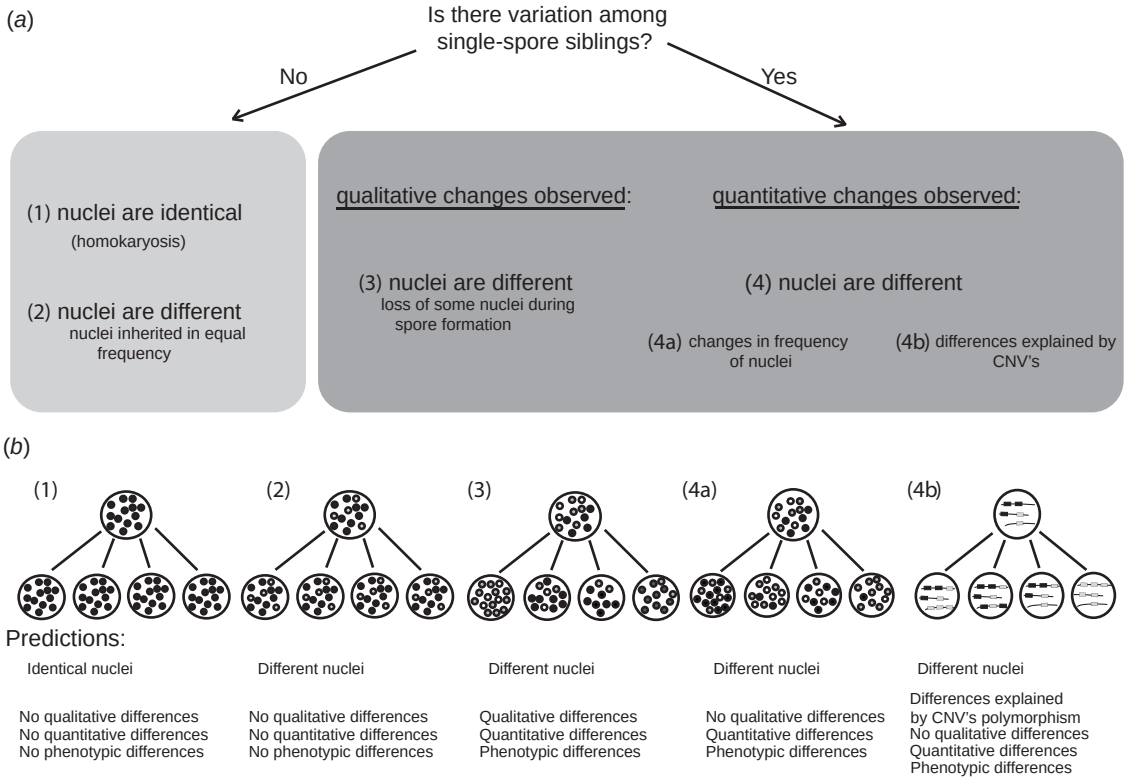


Figure 1. (*a*) A decision tree workflow for this study. (*b*) Scenarios for the outcome of the experimental design. Large circles represent spores and small circles with different grey shading represent genetically different nuclei.

In this experiment, we aimed to experimentally test which of the above scenarios is more likely to explain the within-fungus genetic polymorphism. We took several asexual spores from a single parental culture of *R. irregularis* (isolates DAOM 197198 and B4) and grew each spore separately in new Petri dishes (preventing possible recombination between the different spores). Consequently, we measured whether there were qualitative or quantitative genetic differences among asexual single-

spore siblings originating from a parental AMF culture. We also produced replicates of each single-spore culture by sub-culturing in a way that allowed us to apply statistical tests on qualitative and quantitative genetic data generated from each culture. Additionally, in one of the isolates, we determined whether significant phenotypic differences existed among the single-spore siblings or not.

Methods

Biological material and experimental design

We used *R. irregularis* isolates B4 and DAOM 197198 in this study. Isolate B4 was introduced into *in vitro* culture in 2000 starting with a single-spore of this fungus taken from a pot culture that originated from an agricultural field in Tänikon, Switzerland (Jansa *et al.* 2002). Isolate DAOM 197198 is the reference *R. irregularis* isolate used in many laboratories worldwide and was the first AMF genome to be sequenced (Tisserant *et al.* 2013; Lin *et al.* 2014). The fungi were grown in association with Ri T-DNA transformed carrot roots in an *in vitro* system (Bécard & Fortin 1988). We cultivated these fungi in split Petri dishes, where one compartment only contained the host roots with the fungus and the other compartment contained hyphae and spores of the fungus (St-Arnaud *et al.* 1996). We created 9 single-spore cultures of the isolate B4 and 3 single-spore cultures of isolate DAOM 197198. Each of these single spores germinated, colonized the roots, proliferated and produced new clonal spores. In order to obtain biological replicates of each single-spore culture, after proliferation of the material, we divided each single-spore culture into three equal parts (each containing roots, hyphae and more than 3000 spores) and propagated these parts in new split Petri dishes (electronic supplementary material, figure S1). It was not possible for any of the siblings cultures to fuse with any of the other cultures during propagation as each sibling was cultivated in a separate Petri dish. All cultures were maintained for 6 months.

ddRAD-sequencing and data analysis

The DNA was extracted from fungal material of all replicates using the DNeasy Mini Plant Kit (QIAGEN) and the DNA concentration was normalized in order to use the same amount of DNA in each sample. A ddRad-seq protocol was used to prepare libraries for Illumina sequencing. We chose ddRAD-seq for genotyping over other multi-locus sequencing techniques, such as whole-genome sequencing, because of the relatively high coverage per locus that can be achieved for a lower price in a large number of individuals. This technique allowed us to obtain a genome-wide estimation of

genetic differences among sibling cultures. It also allowed us to identify rare alleles that can only be detected with a deep coverage. In addition, a previous study using ddRAD-seq obtained results on genetic polymorphism in *R. irregularis* that were consistent with another independent sequencing technique, namely amplicon sequencing (Wyss *et al.* 2016), thus confirming the reliability of the ddRAD-seq protocol.

The libraries were sequenced using Illumina Hi-Seq 100nt paired-end technology. We first processed the raw reads with Tagcleaner-0.14 to trim the Illumina adapters (Schmieder *et al.* 2010). Quality-filtering and trimming was carried out using PrinSeq-lite-0.20.4 (Schmieder & Edwards 2011) and the reads were then demultiplexed by process_radtags-1.21 from Stacks (Catchen *et al.* 2011). We used the aligner Novoalign-3.02 to align the sequence reads against the reference assembly (DAOM 197198) of a single nucleus of *R. irregularis*, N6 (Lin *et al.* 2014). The size of this assembly is 115 Mb and comprises 12567 scaffolds, N50 = 20776. In addition, we also mapped the reads against the Rhiir2 assembly, an unpublished improved assembly constructed from sequences originating from spores and mycelia of DAOM 197198. This assembly spanned 137 Mb in 1123 scaffolds, N50 = 336373 (F. Martin, personal communication). Samtools-1.2 (Li *et al.* 2009) was first used to record sequences with a minimum mapping quality of 30, then we used it to extract the count of number of reads per allele per site sequenced. We then used Popoolation2-1201 (Kofler *et al.* 2011) in order to obtain a '.sync' file of the different read counts per samples. Interspersed repeats were identified with RepeatModeler Open-1.0/RepeatMasker open-3.0 (Smith *et al.* 2010) and multiple copy sites were identified with homology-based method as described in Wyss *et al*.. Coding regions were defined with the *ab initio* predictor GeneMark-ES (Ter-Hovhannisyan *et al.* 2008). We filtered the data by selecting only sites that displayed a minimum coverage per site of 10×, minimum allele coverage of 6× and alleles with frequencies greater or equal to 0.1. Because of the low quality of sequences obtained from DNA of 3 single-spore cultures of isolate B4, 3 out of the 9 single-spore cultures were excluded from the analysis.

Phenotypic measurements

We measured spore production in each replicate of 9 single-spore siblings of isolate B4 after 6 months of growth. To do this, we took photographs of 6 areas of 2 cm^2 in the hyphal compartment of each *in vitro* culture using a Leica stereoscope (MZ125) with a camera attached. An automated measurement

of spore production was then made for each image with the open source software ImageJ (Abràmoff *et al.* 2004). In addition, we developed a method to detect whether the spores were clustered together or if they were regularly spaced, using the R package Spatstat (Baddeley & Turner 2005). To do this, we measured the spatial organisation of the spores produced by the different single-spore siblings by measuring the nearest distance to spores from random points chosen within each image.

Statistical analysis

We conducted two types of analysis of genetic polymorphism depending on the nature of the genetic data. First, we conducted a qualitative analysis that used the allele identity to distinguish whether qualitative differences in the presence or absence of alleles occurred among the siblings. Second, we conducted a quantitative analysis to test for differences in allele frequencies among the siblings. In order to qualitatively assess differences among siblings, we calculated a genome-wide pairwise fixation index F_{ST}. We calculated the genome-wide F_{ST} differentiation among different pairs of single-spore siblings with the software Popoolation2 (Kofler *et al.* 2011). The F_{ST} metric is a widely used metric that allows the measurement of the differentiation among populations (Wright 1951; Nei 1973). The values of F_{ST} range from 0 to 1, where a value close to 0 means that there is no genetic differentiation among the single-spore siblings, and a value higher than 0.2 means that some degree of differentiation exists among siblings.

For the quantitative analyses, we used Samtools and Popoolation2 to extract allele counts. Then we selected the sites that displayed more than one allele (poly-allelic sites) among all the replicates and proceeded to measure the allele frequencies at each site in each sibling. We only selected the sites that displayed more than one allele and where there were no missing data among replicates. We then performed three independent statistical tests on the data at each site. First, we tested whether differences existed in allele frequency among the three replicates of each single-spore culture with the Fisher exact test (FET). We then selected the sites that did not display differences in allele frequencies among replicates. We compared the single-spore siblings in pairwise comparisons to detect statistical differences between each pair of single-spore siblings with the Cochran-Mantel-Haenszel test (CMH). The CMH test was used to test the independence of two variables, while controlling for repetitive measurements. Here, we were interested in the potential difference in allele

counts between pairs of single-spore siblings. The test analysed whether differences in counts (in this case allele frequencies) existed among single-spore cultures, by comparing counts at a given site of one replicate chosen from each treatment. The test was then repeated successively for each possible pair of replicates giving a final probability of whether the allele frequencies at each site were different between single-spore cultures. In this analysis, we split the sites that were tested into different categories, according to whether they were predicted as single-copy or multiple-copy positions and if they were predicted as being in coding or in non-coding regions. We also kept sites in a separate category if they could not be assigned to any of the above categories. We repeated this analysis for each pairwise comparison among the single-spore siblings of isolate B4 and among single-spore siblings of isolate DAOM 197198 separately. In total, this gave 15 pairwise comparisons among single-spore siblings of isolate B4 and 3 pairwise comparisons among single-spore siblings of DAOM 198198. Therefore, we used an analysis where we were able to measure the total number of sites at which we saw significant differences in at least one pair of single-spore siblings. Then, we also calculated the average number of sites that displayed differences in allele frequency between any given pair of single-spore siblings. Finally, we plotted the allele frequency distribution of poly-allelic sites by randomly choosing one of the alleles per site to calculate its frequency.

In order to test whether there were differences in phenotypic measurements between the single-spore siblings of isolate B4, we used a mixed generalised linear model with the R Package Lme4 (Bates *et al.* 2012). The model took into account the *Petri* dish as a random effect. Significant differences among single-spore cultures were tested by comparison of the model to the null model (Single-spore culture + random factor (*Petri* dish)) to the null model (1 + random factor (*Petri* dish)). All the graphical outputs were made with the open source software R (R 2016).

Results

Sequence data

Using the ddRAD-seq data and the two different DAOM 197198 assemblies to align sequence reads, we observed that on average there were 11% fewer reads that mapped uniquely to the unpublished Rhiir2 genome assembly compared to the N6 assembly. On average, the total level of within-fungus polymorphism found in isolate B4 with the N6 assembly was 3.60 poly-allelic sites/kb. This value was

1.79 poly-allelic sites/kb when we used the Rhiir2 assembly. The average estimation of polymorphism in DAOM 197198 when mapped to N6 was 1.66 poly-allelic sites/kb. When mapped to the Rhiir2 assembly, it was 0.64 poly-allelic sites/kb (electronic supplementary material, table S1). We also observed that the level of polymorphism was lower in DAOM 197198 compared to B4. This was in accordance with previous observations (Wyss *et al.* 2016). Data generated from replicates of each single-spore culture were very similar, indicating a high reliability of the dataset. On average, 91% of sites exhibited no difference in allele frequency among replicates of each single-spore culture, as tested by the Fisher exact test (electronic supplementary material, table S2). More information about coverage of the different replicates of each single-spore culture can be found in the supplementary material (electronic supplementary material, figure S2).

Qualitative analysis of genetic polymorphism

There were very few sites (less than 0.05%) among the single-spore cultures that displayed an F_{ST} value greater than 0.2. The average genome-wide F_{ST} between pairs of cultures was below 0.015, confirming that there was no significant qualitative genetic differentiation among the single-spore siblings of isolate B4 (electronic supplementary material, table S3). A similar lack of qualitative genetic differentiation among single-spore siblings was observed in DAOM 197198 (electronic supplementary material, table S4).

Quantitative analysis of genetic polymorphism

In order to see the global extent of sites that displayed differences in allele frequency among the B4 siblings and among the DAOM 197198 siblings, we measured the total number of sites at which we saw significant differences between at least one pair of isolates out of the total pairwise comparisons. When the N6 genome assembly was used to align reads, we observed that 33.26 % out of 5261 sites in B4 and 13.35% out of 3251 sites in DAOM 197198 displayed differences in allele frequency between at least two single-spore siblings (table 1; results from alignment to the N6 assembly). We observed that 32.27% of single-copy sites in B4 and 15.93% of single-copy sites in DAOM 197198 differed significantly in allele frequency. In multiple-copy sites, results were similar, where 33.76% of sites in B4 and 12.09% of sites in DAOM 197198 displayed significant differences in allele frequency (table 1; results from alignment to N6 assembly). We found similar values of differences in allele

frequencies in sites that did not have a copy number prediction (neither classifiable as single or multiple-copy; electronic supplementary material, table S5). Despite the fact that the number of sites analysed differed according to the assembly used for mapping reads, the percentage of sites that differed in allele frequency for the different classes were remarkably consistent between the two genome assemblies (table 1).

	ISOLATE B4						ISOLATE DAOM 197198					
	All categories		Single-copy		Multiple-copy		All categories		Single-copy		Multiple-copy	
	N6	**Rhiir2**	**N6**	**Rhiir2**	**N6**	**Rhiir2**	**N6**	**Rhiir2**	**N6**	**Rhiir2**	**N6**	**Rhiir2**
Total number of sites	5261	3554	1751	1865	3510	1689	3251	1492	1067	734	2184	758
Number of significant sites (p-value > 0.05)	1750	1060	565	545	1185	515	434	223	170	147	264	76
% significant sites	**33.26%**	**29.83%**	**32.27%**	**29.22%**	**33.76%**	**30.49%**	**13.35%**	**14.95%**	**15.93%**	**20.03%**	**12.09%**	**10.03%**
Number of coding sites	2967	2016	850	920	2117	1096	1870	945	530	430	1340	515
Number of significantly different coding sites (*p-value* >0.05)	1026	665	267	322	759	343	259	147	78	87	181	60
% significant coding sites	**34.58%**	**32.99%**	**31.41%**	**35.00%**	**35.85%**	**31.30%**	**13.85%**	**15.56%**	**14.72%**	**20.23%**	**13.51%**	**11.65%**
Number of non-coding sites	2294	1538	901	945	1393	593	1381	547	537	304	844	243
Significantly different non-coding sites (*p-value* >0.05)	724	395	298	223	426	172	175	76	92	60	83	16
% significant non-coding sites	**31.56%**	**25.68%**	**33.07%**	**23.60%**	**30.58%**	**29.01%**	**12.67%**	**13.89%**	**17.13%**	**19.74%**	**9.83%**	**6.58%**

Table 1. Global summary of sites that were tested for differences in allele frequenca in all pairwise comparisons among single-spore cultures of *R. irregularis,* isolates B4 (n=15) and DAOM 197198 (n=3). Values represent the number of sites that displayed quantitative genetic differences in allele frequencies among the single-spore siblings. For each comparison we independently used the sequencing reads aligned to the N6 and Rhiir2 genome assemblies. The contingency table contains information about the number of sites analysed and the number of significant differences in allele frequency (CMH test). The table shows the sites analysed in 4 categories (coding, non-coding, single-copy and multiple-copy).

The sites that significantly differed in allele frequencies among single-spore siblings were not restricted to a specific region of the genome. The sites that differed in allele frequencies in B4 were located on 50.83% of the tested scaffolds. In isolate DAOM 197198 the sites that differed in allele frequency were located in 23.86 % of the tested scaffolds (electronic supplementary material, table S6; results from alignment to N6 assembly).

When we looked at the average number of sites that displayed significant differences in allele frequency between a given pair of siblings, we observed that 6.79% of sites in B4 and 6.56% of sites

in DAOM 197198 displayed significant differences in allele frequency among any two single-spore siblings (table 2; results from alignment to the N6 assembly). On average, in single-copy sites, we detected 6.06% of sites in B4 and 7.96% of sites in DAOM 197198 that differed in allele frequency among single-spore siblings (table 2; results from alignment to N6 assembly). Similar results were found when using the Rhiir2 assembly (table 2). We observed that some pairs of cultures displayed a lower number of sites that displayed differences in allele frequency (e.g. B4 pair of cultures ssc5-ssc6: 2.78% of sites; DAOM 197198 pairs ssc2-ssc3: 4.45% of sites) compared to other pairs where the number of sites was higher (e.g. B4 pair of cultures ssc1-ssc5: 11.12% of sites; DAOM 197198 pairs ssc1-ssc2: 7.63% of sites; electronic supplementary material, tables S7 and S8; results from alignment to N6 assembly).

	Isolate	B4 (n=15)		DAOM 197198 (n=3)	
	Reference assembly	N6	Rhiir2	N6	Rhiir2
Single-copy	Total	1201.13	1284.00	850.00	628.67
	Significant	72.40	66.47	67.67	66.67
	%	6.06%	5.11%	7.96%	10.57%
Multiple-copy	Total	2489.40	1124.27	1758.67	596.00
	Significant	166.00	71.60	104.33	29.67
	%	6.66%	6.32%	5.92%	4.98%
No prediction	Total	950.53	401.13	629.33	167.67
	Significant	77.33	26.00	41.00	11.00
	%	8.12%	6.38%	6.51%	6.58%
Total	Total	4641.07	2809.40	3238.00	1392.33
	Significant	315.73	164.07	213.00	107.33
	%	6.79%	5.78%	6.56%	7.71%

Table 2. Mean results of the pairwise comparisons of quantitative genetic differences in allele frequencies between any pair of single-spore cultures of isolate B4 (n=15) and DAOM 197198 (n=3). For each comparison we used the data aligned to the N6 and Rhiir2 genome assembly independently. We reported the total number of sites tested and the number of sites showing significant differences in allele frequency (CMH test). All these values were calculated for: Single-copy sites, multiple-copy sites, sites that did not have any prediction regarding the number of copies, and the total number of sites tested.

Finally, the allele frequency distribution plots showed that isolates B4 and DAOM 197198 displayed a uni-modal distribution with a peak in frequencies where alleles are present in a 50:50 proportion (figure 2; electronic supplementary material, figure S3)

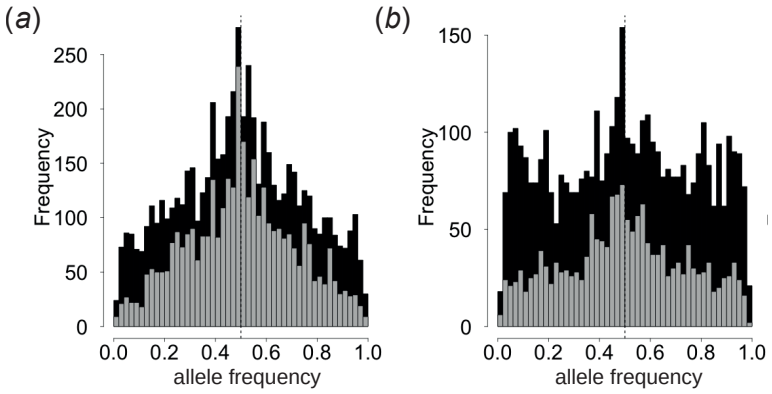


Figure 2. Allele frequency distribution plot of poly-allelic sites in isolates (*a*) B4 and (*b*) DAOM 197198. In grey, data aligned to N6 assembly. In black, data aligned to the Rhiir2 assembly. Each figure shows the first replicate of one of the single-spore cultures of each isolate. For the other figures showing all replicates of each isolate, see electronic supplementary material, figure S3.

Phenotypic differences among single-spore siblings of isolate B4

The number of spores produced differed significantly among the single-spore siblings (generalised linear mixed model (GLMM): Single-spore effect *vs.* NULL model: χ^2 = 32551, *df* = 8, $p < 0.001$; figure 3). In addition, the spatial organization was significantly different among the single-spore siblings with some cultures exhibiting a more regular spatial distribution than others (GLMM): Single-spore effect *vs.* NULL model*:* χ^2 = 1273, *df* = 8, $p < 0.001$; electronic supplementary material, figure S4)

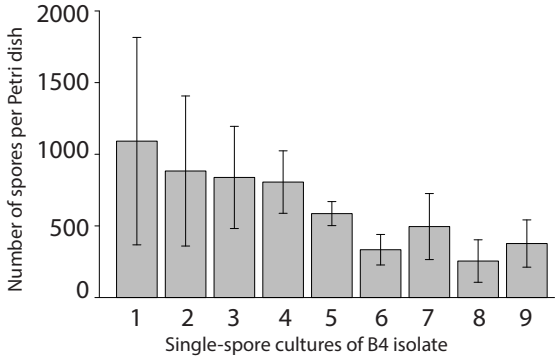


Figure 3. Spore production of 9 single-spore sibling cultures of *R. irregularis*, isolate B4. For each single-spore sibling culture, 6 different areas of 2 cm^2 on three different Petri dishes were used for spore counts. Mean value per single-spore culture and standard deviation are shown in the figure.

Discussion

In this study, we observed genetic variation produced among clonal single-spore siblings of the *R. irregularis* isolates B4 and DAOM 197198 after a single generation. We did not observe any significant qualitative genetic changes, but we observed significant quantitative genetic changes among single-spore siblings. We also observed that quantitative genetic differences among siblings occurred in coding and non-coding regions, and were randomly distributed across the genome rather than being clustered in a single region of the genome. We observed that the use of a different reference assembly did not influence the detection of allele frequency differences among single-spore siblings. Finally, we observed that the clonal single-spore siblings of isolate B4 displayed phenotypic differences after a single generation, demonstrating that within-fungus genetic polymorphism could have important consequences for the biology of AMF and their interaction with host plants.

Compared to other studies using similar experimental designs, our study contains novel features that describe the inheritance of genetic variation in clonal offspring in a more extensive way. First, we used a reliable multi-locus genotyping approach that was consistent among biological replicates (electronic supplementary material, table S2). Second, we included true replication for each of the single-spore cultures allowing statistical tests, thereby, controlling for false-positives. Third, we achieved a high coverage per site, giving us high confidence for allele identification. Finally, we aligned the sequence data to two different genome assemblies of isolate DAOM 197198 that allowed us to control for a possible bias caused by the quality of the reference assembly.

We detected quantitative genetic variation among siblings. Consequently, scenarios 1 (homokaryosis) and 2 (heterokaryosis with nuclei inherited in equal frequencies) cannot explain the observed data. We also exclude scenario 3 (loss of given nuclear genotypes during spore formation) because we did not find evidence of qualitative genetic differences among the siblings. However, we cannot discard scenario 4 because we observed quantitative genetic differences among the siblings.

Given that *R. irregularis* is considered haploid, the changes in allele frequency in single-copy sites in DAOM 197198 suggest that more than one type of nuclear genotype should coexist in this isolate.

However, in the analysis of the data for isolate B4, we may have overestimated the number of single-copy sites. We used the closely related isolate DAOM 197198 as the reference genome assembly to predict single-copy and multiple-copy sites in B4. These sites classified, as single-copy in B4 because they were defined as single-copy in the DAOM 197198 genome, could potentially be multiple-copy in isolate B4. Thus variation at a number of these sites could potentially represent CNV. Therefore, CNV could be an alternative hypothesis to explain the differences in allele frequency between the single-spore siblings in B4 (scenario 4b). In DAOM 197198, the prediction of single-copy and multiple-copy sites is much more accurate because we mapped the reads to the reference assembly of the same isolate. In this isolate, there were still a significant number of single-copy sites that displayed differences in allele frequencies among single-spore siblings. This was confirmed by using the greatly improved assembly Rhiir2. Hence, the results from the DAOM 197198 sequence data suggest that CNV polymorphism could be one source of variation as we observed differences in allele frequencies at multiple-copy sites (scenario 4b). In this case significant differences in allele frequencies among the siblings cannot easily be explained unless CNV occurs among nuclei. However, heterokaryosis is the most likely hypothesis to explain the differences in allele frequencies in single-copy sites in DAOM 197198 (scenario 4a). This was supported by the 50:50 ratio peak found in the allele frequency distribution observed with data aligned to the two assemblies (N6 and Rhiir2) that suggest that at least 2 dominant nucleotypes exist in these isolates.

We observed changes in allele frequencies between single-spore siblings in coding regions, suggesting that these changes could potentially have functional consequences. We also found that the genetically different siblings displayed significantly different phenotypes; a feature that would not be expected if progeny arising from one parent were genetically identical. As the single-spore siblings were cultured in the same controlled environment, it is likely that the phenotypic polymorphism observed is due to genetic variation in the siblings.

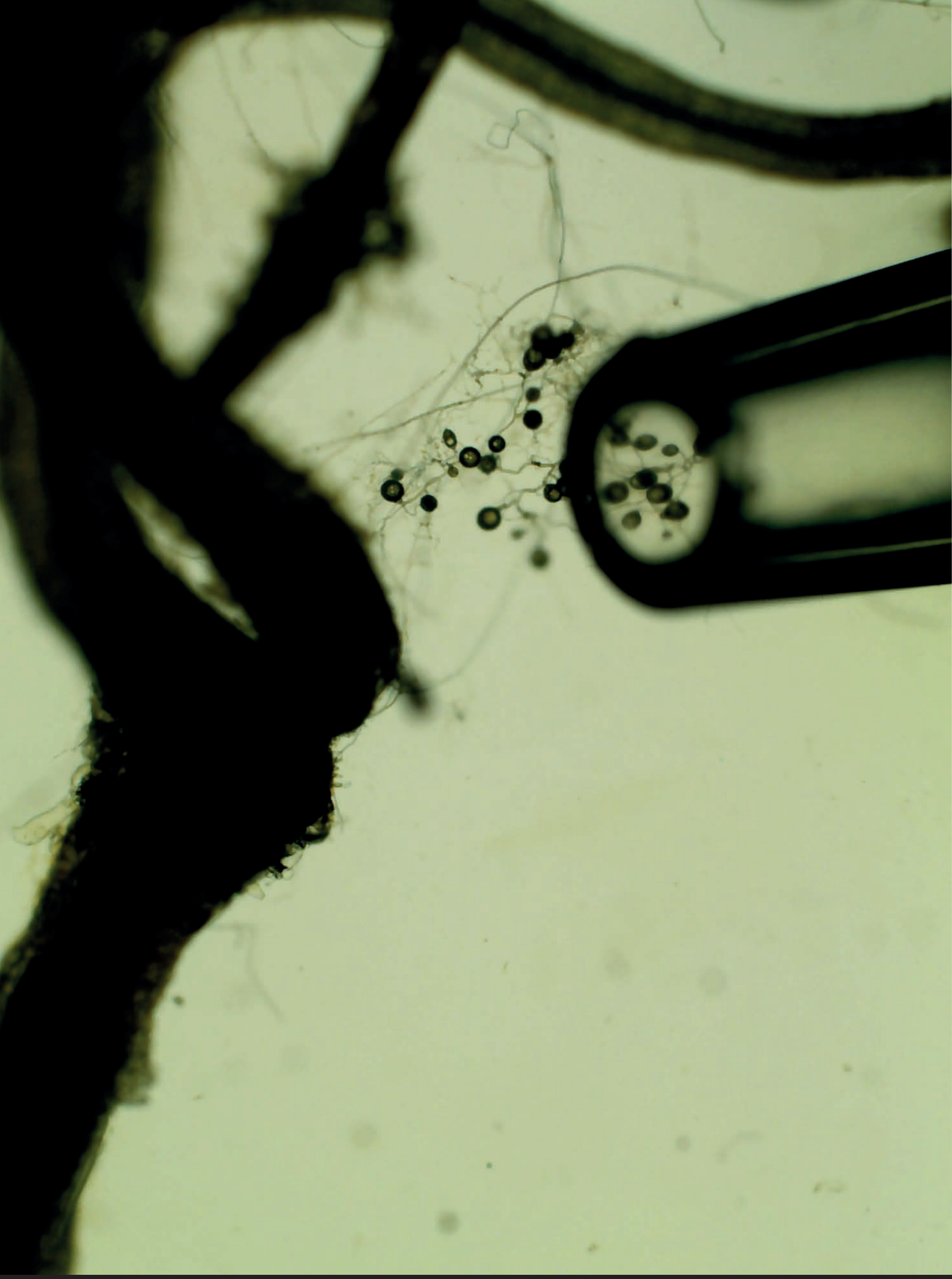
In nature, genetic variation in clonal organisms has been reported in different organisms such as aphids (Vorburger 2006), plants (Ellstrand & Roose 1987), and Daphnia (Hebert *et al.* 1989), among others. In *R. irregularis*, the coexistence of at least 18 clonal genotypes in a single agricultural field has been reported (Croll *et al.* 2008). Hence, harbouring genetic variation from different nuclei, copy

number polymorphisms and somatic mutations could be possible sources of variation that can lead to the differentiation of different clonal genotypes in *R. irregularis* in the nature.

In conclusion, in this study we highlighted two mechanisms by which *R. irregularis,* a clonal organism, generates genetic and phenotypic variation in a single generation from asexual spores. These results are evolutionarily and ecologically relevant for three reasons: First, the production of genetic variation from asexual spores will allow the siblings to possess a certain degree of polymorphism that could make the population more resilient against stochastic processes. Second, the fact of possessing different nuclei opens a new range of questions, such as understanding mechanisms of cooperation or competition between the nuclei, as well as whether there is genetic exchange between them. Third, as AMF are strict endosymbionts of plant roots, genetically variable asexual offspring could interact differently with host plants, influence the outcome of the symbiosis (Angelard *et al.* 2010), and thus impact plant ecology (van der Heijden *et al.* 2003, 2015). As a consequence, the quantitative analysis of genotypic polymorphism should be considered of vital importance when analysing the genetic polymorphism and its ecological and evolutionary consequences in arbuscular mycorrhizal fungi.

Acknowledgments

We are grateful to Jeremy Bonvin for help in the culturing the isolates and Francis Martin for providing us an unpublished genome assembly of DAOM 197198 (Rhiir2). We also thank Keith Harshman and Johann Weber (Lausanne University Genomic Technologies Facility) for the sequencing of the ddRad-seq libraries. This research was funded by a Swiss National Science Foundation grant (3100 3A_162549).



Collecting spores of *R. irregularis*
Jeremy Bonvin

Chapter 2. Phylogenetic signal shows that intra-specific genetic variation in Rhizophagus irregularis causes variation in fungal traits and growth of a globally important plant

Ivan D. Mateus-Gonzalez[1], **Frédéric G. Masclaux**[1,2] , **Tania Wyss**[1], **Romain Savary**[1] **& Ian R. Sanders**[1]

[1]*Department of Ecology and Evolution, University of Lausanne, Biophore building, 1015 Lausanne, Switzerland*
[2] *Vital-IT, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*

Abstract

Arbuscular mycorrhizal fungi (AMF) are important because of the large impact they have on plant growth. However, understanding how these fungi affect plant growth has been difficult because of the genetic complexity of these fungi. Data from a population of *Rhizophagus irregularis* show that genetic variation in the fungus occurs concurrently with variation in fungal phenotypic traits and in how plants grow. However, it has never been tested whether this genetic polymorphism is responsible for the phenotypic variation observed between isolates.

In order to test if genetically similar isolates of *R. irregularis* display similar phenotypes and similar plant responses, we analysed a population of *R. irregularis* that was isolated from an agricultural field in Switzerland. First, we characterized the phenotype of these isolates by measuring different traits such as spore production and mycelium density. Then, we inoculated cassava (*Manihot esculenta*) plants with each isolate and we measured the phenotypic response of these plants.

We observed that genetically similar isolates, displayed significantly similar spore production. In addition, we observed that plants inoculated with similar isolates displayed a significantly similar dry-weight compared to more distant isolates. Our results provide the first demonstrated link between genetic variation in the fungus, their phenotypic variation and plant growth response. This is an essential link to establish in our ultimate goal to use genetic variation in AMF to improve plant growth.

Introduction

Arbuscular mycorrhizal fungi are plant endosymbionts that provide nutrients to the plant in exchange for plant carbohydrates (Smith & Read 2008). This means that these fungi can be used in agriculture to promote crop growth. Recently, it has been shown that AMF can enhance the productivity of cassava (*Manihot esculenta*). Previous studies show that cassava is a crop that displays a high response to AMF (Sieverding, 1985). Furthermore, in general terms, the outcome of the response of the host plant to AMF depends on the soil composition (Zaller *et al.* 2011), the identity of the AMF and the host identity (Helgason *et al.* 2002; Klironomos 2003).

Despite the multiple factors influencing the plant response to AMF inoculation, then plant responses could be predicted by taking into account the phylogenetic inertia of the AMF fungi and the host plant.

All living species have a common ancestor, and by consequence more related species generally display more similar traits. Trait conservatism can influence ecological and evolutionary processes and can be a tool to predict the behaviour or niche use of similar species (Gilbert & Parker 2016). The phylogenetic signal describes, the trait conservatism within related species. This has been shown in plant pathogens where genetically related species have similar impacts on the host plant (Gilbert *et al.* 2015). Furthermore, similar pathogens can have similar host ranges, as the likelihood of a pathogen to infect two plant species decreases with the phylogenetic distance between these two plants (Gilbert & Webb 2007). Therefore, the detection of a phylogenetic signal between AMF genetic variation and plant growth could as well exist.

It has been shown that similar AMF species within the Glomeromycota phylum display similar colonization levels of the host plant, confirming that complex traits as colonization of an endosymbiotic fungus are conserved within families (Powell *et al.* 2009). Within the Glomeromycota phylum the species *Rhizophagus irregularis* has been well studied and reported to display a high among-isolates genetic polymorphism (Wyss *et al.* 2016), high phenotypic diversity (Koch *et al.* 2004) and differential effects on the host plant (Koch *et al.*, 2006; Angelard *et al.*, 2010). The previous features, means that *R. irregularis* could be a good candidate to detect intra-species trait conservatism. In this case, such a test would be highly relevant because a causality link between the *R. irregularis* genetic variation and the plant growth response to *R. irregularis* has never been proved.

In addition to among isolates genetic variation in *R. irregularis*, within fungus genetic variation has been reported in this species (Wyss *et al.* 2016). It has been shown that phenotypically distinct *R. irregularis* siblings derived from the same parental isolate, displayed genetic differences in allele frequencies (Angelard *et al.* 2010; Ehinger *et al.* 2012; Mateus-Gonzalez et al. *Chapter 1*). The genetic differences in terms of allele frequency could be the result of segregation of different nuclei into newly produced asexual spores (Mateus-Gonzalez et al. *chapter 1*). Thereby, the multinucleated status of AMF describes the within-isolate genetic polymorphism of *R. irregularis* isolates. In consequence, the within-isolate genetic polymorphism could also be influenced by the genetic relationship between these fungi and their quantitative traits or effects on plant growth.

The high genetic diversity found in this species allows us to test whether if there is conservatism of phenotypic traits at the intra-species level. In addition, we also tested if there is conservatism in the plant response to AMF. In order to investigate this, we grew different isolates of *R. irregularis* in Petri dishes to measure different phenotypic measurements. In parallel we inoculated clones of the cassava cultivar NGA-16 with different isolates of this fungus and then measured the growth response of these plants to the different AMF isolates. We used SNP data on a population of *R. irregularis* (Wyss *et al.* 2016) to construct the genetic relationship between the different isolates. Finally, we used two metrics that describe the genetic relationships between the isolates to calculate the phylogenetic signal of fungal traits and plant response to *R. irregularis*. The first metric used describes the among-isolates genetic polymorphism (presence/absence of SNP) (Wyss *et al.* 2016). The second metric describes the genetic differences and similarities among AMF isolates based on the within isolate genetic polymorphism, more precisely in allele frequencies. These two metrics were used in order to infer the genetic relationship between the isolates and calculate the phylogenetic signal of fungal and plant response traits.

Materials and methods

Fungal growth

We used several *R. irregularis* isolates (A4, A5, B3, B4, B10, B15, C1, C2, C4, D1 & D4), which are representative of different genetic clusters reported in an *R. irregularis* population (Croll *et al.* 2008). We grew the isolates in split-plates Petri dishes (St-Arnaud *et al.* 1996) in association with Ri T-DNA transformed carrot roots (Bécard & Fortin 1988). We phenotypically characterized five replicates of the different isolates with the same methodology as in the chapter 1 of this thesis (Mateus-gonzalez et al., 2016 *Chapter 1*). We measured the spore production of each sample after 6 months of growth. To do this, we took photographs of 6 areas of $2cm^2$ in the fungal compartment of each *in-vitro* culture with the camera device (DFC290) of a Leica stereoscope (MZ125). An automated measurement of spore production was then made for each image with the open source software ImageJ (Abràmoff *et al.* 2004). In addition, we measured the spatial distribution of spores in the Petri dish. We measured whether the spores were clustered together or if they were produced in a more regular distribution, using the R package Spatstat (Baddeley & Turner, 2005; CHAPTER 1). To do this, we measured the spatial organisation of the spores produced by the different single-spore siblings by measuring the

nearest distance to the spores from random points chosen within each image. We also measured the hyphae produced by counting the number of hyphae that crossed two transects of 1.44 cm length. We took 5 independent pictures for each dish to take this measurement.

Plant responses to inoculation with different AMF isolates

We propagated *in-vitro* the cassava variety NGA-16. The plantlets were grown in a growth chamber (25°C, 14 hours light, 90% RH) in essay tubes on MS medium for 1 month. Then, we placed the seedlings in a steam sterilized (180° 25-min) soil substrate (Klassman seedling substrate:perlite 1:1). After 1 month of growth ex-vitro, we transferred the plantlets to the final steam sterilised (120° 40 min 2x) substrate (Klassman substrate 4:sand:clay:perlite (4:2:1:1). The plants were then kept in the greenhouse at conditions of 28°C, 16 hours light, 70% RH.

We inoculated each plant with 300 spores of the *R. irregularis* isolates (A3, B4, B10, C1, C2, C3, C4, C5, D1, D4, G1). We produced 15 replicates for each treatment. The plants were harvested after 8 months of growth. We measured the height and dry weight of aboveground and belowground parts of the plants. Dry weight was obtained after the plants were dried at 72°C for 6 days.

Root colonization

Root colonization was determined using at least 10 replicates of each treatment by the grid line intersect method (Giovannetti & Mosse 1980), after clearing roots with 10% KOH for 4 hours, acidified with HCl (1%) during 5 minutes and staining with trypan blue (0.10% in a lactic acid-glycerol solution) overnight.

Genetic relatedness among AMF isolates

We used published SNP data on a population of *R. irregularis* (Wyss *et al.* 2016) and unpublished SNP data obtained from other *R. irregularis* isolates issued by the same methodology developed by Wyss et al., to construct the genetic relationship among the different isolates.

A traditional genetic relatedness analysis (comparing presence/absence of SNP's) that describe the among-isolate genetic polymorphism, of the Tänikon population was done using the same methods as in Wyss *et al*..

We also calculated an alternative metric of genetic relatedness among isolates by using the information of the within-isolate genetic polymorphism, as described by comparing the allele frequencies of sites displaying more than one allele between two pairs of isolates. To do this, we selected sites that displayed more than one allele. We then calculated allele frequencies per site. We applied a Fisher exact test for each pair of isolates in order to see, if at a given site, there were differences in allele frequency between the two given isolates. We then calculated a value of relatedness between two isolates by measuring how many sites differed in allele frequency out of the total number of sites tested. This allowed us to perform pairwise comparisons on the isolates and describe the genetic relatedness among the isolates in terms of their differences in within-isolate genetic polymorphism as measured by differences in allele frequency changes. Then for the two different genetic relatedness analyses, we calculated Euclidean distances among the different isolates and performed a hierarchical clustering with the complete linkage method, to obtain dendrograms of the *R. irregularis* isolates genetic relationships.

Dendrogram comparison

We used the R packages 'ape' (Paradis *et al.* 2004) and 'dendextend' (Galili 2015) to compare the dendrograms obtained by the two different methods. We computed a Mantel test in order to test if there was a correlation between the two dissimilarity matrices.

Statistical differences between treatments

We used a generalized linear mixed-effect model analysis (Bates *et al.* 2012) and the non-parametric Kruskal-Wallis test in order to see whether the fungal phenotypic measurements were significantly different among the fungal isolates. We used a mixed-model with the block factor as random in order to test whether the plants inoculated with different isolates grew significantly different.

Phenotypic traits phylogenetic signal

We used the R package 'phylosignal' (Keck *et al.* 2016a) in order to test whether there was a phylogenetic signal in the *R. irregularis* phenotype and in the response of the host plant to AMF. We calculated two different phylogenetic signal metrics; Abouheif's C_{mean} (Abouheif 1999) and Moran's I (Moran 1948). These two methods are not based on an evolutionary model and use an autocorrelation

approach (Münkemüller *et al.* 2012). We calculated the phylogenetic signal metrics to the quantitative fungal traits (Spore production, Extra-radical mycelia and spore clustering) and the plant growth response to AMF (Height, Aboveground dry-weight, belowground dry-weight). We performed these analyses using the two different genetic relatedness measurements calculated.

Results

We observed significant phenotypic differences among the different *R. irregularis* isolates on the spore production, production of extra-radical mycelia and spore clustering measurements (Table 1a). Notably, isolates A4 and C4 produced half of the amount of spores produced by isolate C2 (Figure 1).

(a)

Glmer(poisson): Full model: X ~treatment + (1|dish):

Null model: X ~ 1 + (1|dish):

	Chisq	df	*p-value*
Spore production	41141	10	< 0.001
Extra radical mycelia	1114	10	< 0.001
	Kruskal-Wallis: X ~ treatment		
Spore clustering	100	10	< 0.001

(b)

Glmer(binomial): Full model: X ~treatment:

Null model: X ~ 1

	Res.deviation	*df*	*null res.dev*	*df*	*df model - null*	*p-value*
Fungal colonization	560	126	829	136	10	<0.0001

(c)

Model: X ~ Treatment +(1| Block)

	numDF	denDF	*F-value*	*p-value*
Height	10	139	3	0.0082
Aboveground DW	10	139	2	0.0174
Underground DW	10	139	3	0.0005
Total DW	10	139	4	0.0003

Table 1. (a) Statistical analysis of the fungal phenotypic measurements among the isolates. (b) Statistical analysis of the fungal colonization on plant roots. (c) Statistical analysis of the host response to AMF.

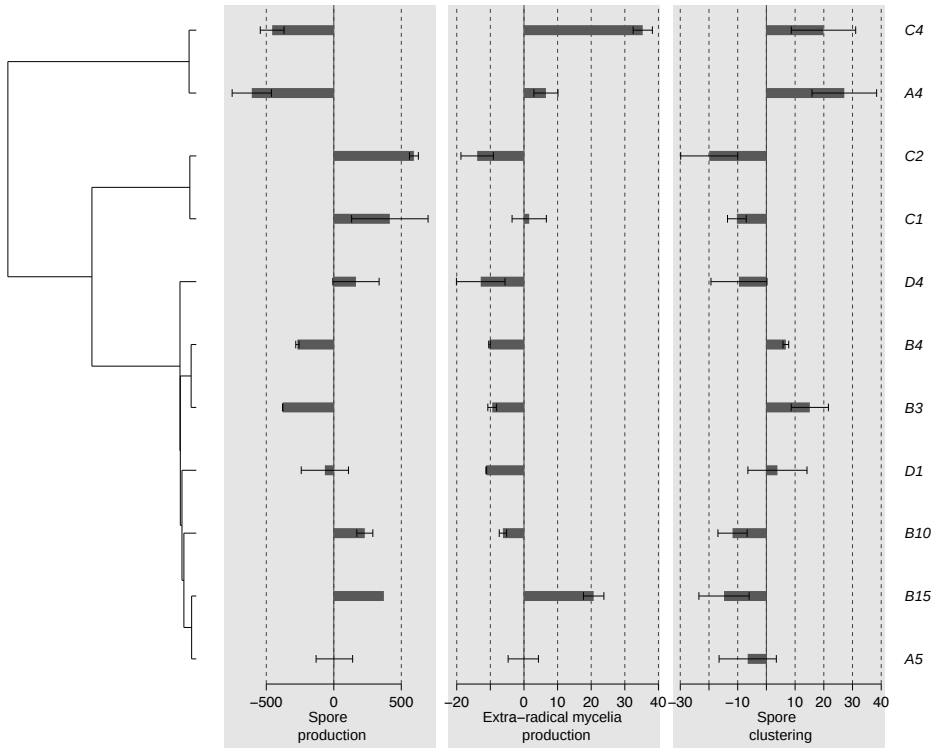


Figure 1. Mean spore production, extra-radical mycelia production and spore clustering next to the dendrogram based on the genetic divergence among different *R. irregularis* isolates. The trait values are centred (value – mean).

We observed that there was a significant difference in the colonization rate among the different isolates in cassava roots (Table 1b), where isolates C3 and C4 colonized 27 % less the plant roots compared to isolate C1 (Figure 2). In addition, we observed that the *R. irregularis* isolates induced significantly different plant growth responses. Plant height, aboveground and belowground dry weight and total dry weight, significantly differed among plants inoculated with the different fungi (Table 1c). Remarkably, the inoculation with the isolate C2 increased by 40 % the plant dry weight compared to plants inoculated with isolates B10 and D1 (Figure 2).

We constructed two dendrograms: a dendrogram based in differences in the presence/absence of alleles among the *R. irregularis* isolates and a dendrogram based on pairwise allele frequency differences between all the fungal isolates. We observed that the dendrogram based on the within-isolate genetic polymorphism was not significantly different from the traditional dendrogram based on the differences in presence/absence of alleles among the fungal isolates (Supplementary Figure 1).

We then tested if there was a significant phylogenetic signal in the fungal growth traits and the plant response to the different fungi. The two different phylogenetic signal statistics supported the hypothesis that genetically more related isolates displayed more similar spore production and spore

clustering (Figure 1, Table 2a). This was not the case for the extra-radical mycelia (Figure 1, Table 2a).

(a)

Statistics

	Cmean	I
Spore production	0.377	0.45
Extra radical mycelia	0.123	0.127
Spore clustering	0.393	0.44

p-value

	Cmean	I
Spore production	0.0167	0.0031
Extra radical mycelia	0.1122	0.0817
Spore clustering	0.0115	0.0038

(b)

Statistic

	Cmean	I
Fungal colonization	0.269	0.26
Height	0.272	0.121
Aboveground DW	0.272	0.118
Belowground DW	0.526	0.37
Total DW	0.534	0.375

p-value

	Cmean	I
Fungal colonization	0.0564	0.056
Height	0.0495	0.1146
Aboveground DW	0.0461	0.1321
Belowground DW	0.0032	0.0191
Total DW	0.0015	0.0158

Table 2. (a) Phylogenetic signal of fungal traits. (b) Phylogenetic signal of fungal colonization and plant response to the mycorrhizal fungus. Summary statistic of Abouheif's C_{mean} and Moran's I. Metrics statistics (above) and significance (*p-value*) are shown.

There was no significant phylogenetic signal in the roots colonization by the fungal isolates (Figure 2, Table 2b), however the *p-values* at 0.05 suggest that there was a weak trend of an effect in this measurement. We observed a significant phylogenetic signal in the plant height, belowground dry weight and total dry weight of the host plant (Figure 2, Table2b). This result is supported for the two different statistics defining the phylogenetic signal (Table 2b). These results were consistent but less significant when we used the genetic information of the within-isolates genetic polymorphism on allele frequency changes to estimate the relationship between the traits and the genetic relatedness of the isolates (Supplementary Figure 2).

Discussion

In this study we found that the different *R. irregularis* isolates differed in their spore production, production of extra-radical mycelia and spore clustering. We also observed that they differed in their colonization inside cassava roots. We also observed that cassava displayed significantly different growth responses to the *R. irregularis* isolates. We found that there was a high correlation among the genetic relatedness of the *R. irregularis* population measured in terms of presence/absence of

mutations (between-isolates polymorphism), and the measurement of genetic relatedness made with the allele frequency changes (within-isolates polymorphism).

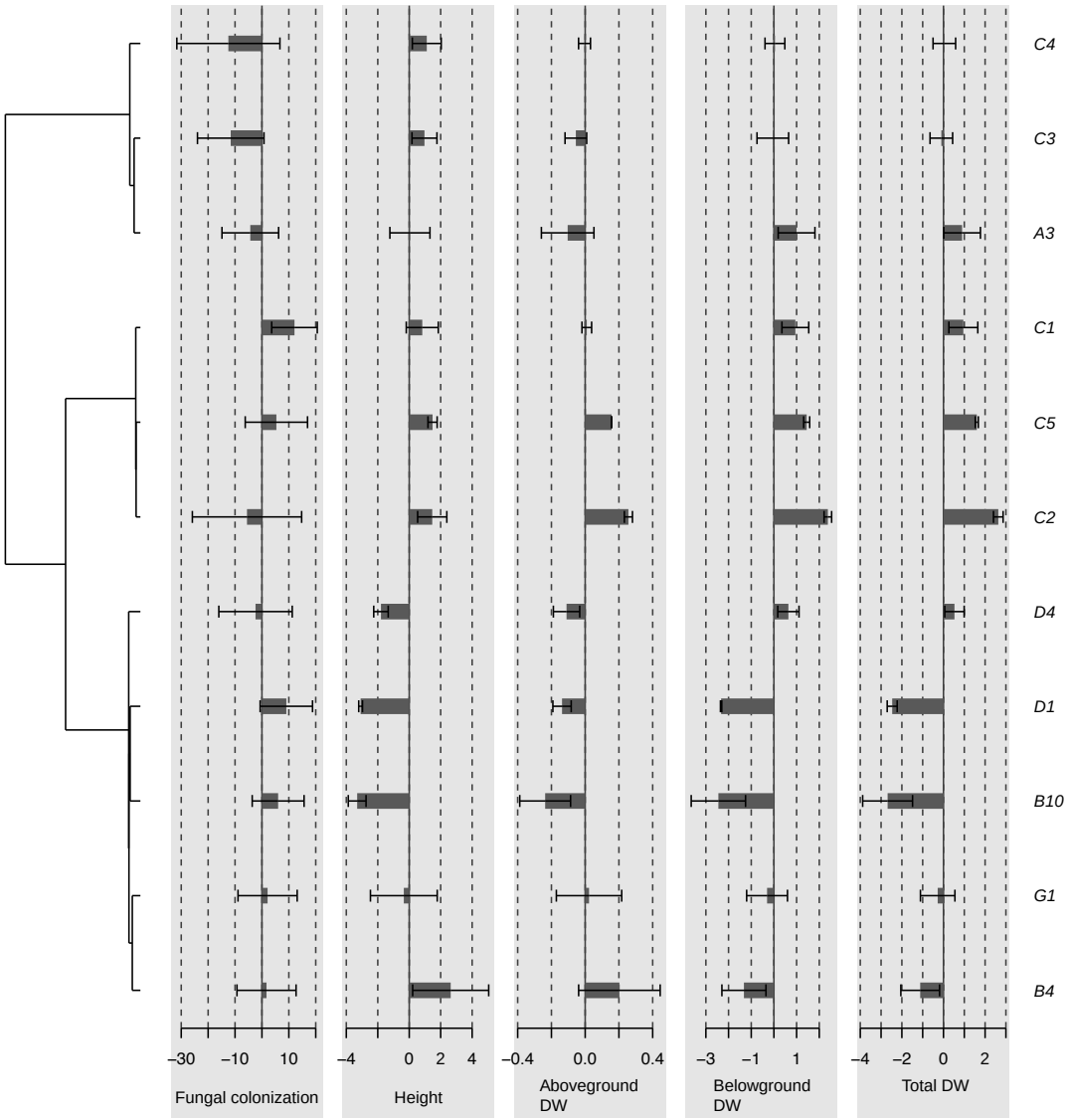


Figure 2. Representation of fungal traits (fungal colonization) and plant traits (height, aboveground dry weight, belowground dry weight and total dry weight) along with the dendrogram based on the genetic divergence among different *R. irregularis* isolates. The trait values are centred (value – mean).

We found that fungal traits such as spore-production and spore clustering displayed a phylogenetic signal. Finally, we found a phylogenetic signal between the *R. irregularis* genetic variation and the host plant response to *R. irregularis*.

The phylogenetic signal observed in the fungal phenotype and host plant response, could be also the result of the effect of other confounding factors such as: 1) a genetic bottleneck resulting from the selection of a single-spore, which was then used to start the initial *in-vitro* cultures and 2) the identity of the initial host-plant used as primary trap-culture for these isolates. However, Koch *et al.* 2004,

showed that the initial genetic drift resulting from the selection of single-spores to start *in-vitro* cultures could not explain the genotypic and phenotypic diversity of the population. Moreover, although the initial trap-cultures influenced the frequency of genotypes that could be isolated from the field (Croll *et al.* 2008), there is no explicit evidence that there is a correlation of the initial trap-culture and the three genetic clades obtained by the multi-locus genotyping ddRad-seq (Supplementary figure 3).

We observed a strong correlation between the genetic relatedness among-isolates and the genetic relatedness among isolates based in the within-isolates genetic polymorphism. The changes in allele frequency between isolates is a genetic measurement taking into account, only changes in frequency between sites which have more than one allele. Different isolates of *R. irregularis* have been shown by means of flow cytometry to be haploid (Sedzielewska *et al.* 2011; Ropars *et al.* 2016). Hence, changes in allele frequency in sites that display more than one allele on haploid organism, describes the within-isolate genetic polymorphism of the multinucleated fungus (Mateus-Gonzalez et al*. chapter 1*). We then observed that genetically more related isolates, display also more similar levels of within-isolates genetic polymorphism. This suggest that the divergence of the within-isolate polymorphism measured in terms of changes in allele-frequency between isolates, is the result of the evolutionary differentiation between the different isolates.

We observed that genetically similar isolates displayed similar phenotypes. However, evolutionary processes such as local adaptation could affect the phenotypic traits of species. All the isolates tested in this study, except DAOM197198, originated from one agricultural field in Tänikon (Switzerland) that was part of a long-term experiment to study the impact of tillage in agriculture. There is the possibility that the different isolates adapted specifically to fields were tillage was used or not as an agricultural practice and the result that we observed in this study could be the consequence of the adaptation to these environments. However, the genotypic and phenotypic diversity observed in the *R. irregularis* population could not be explained by this long-term agricultural practice (Koch *et al.* 2004).

It has been shown, that the use of a commercial *R. irregularis* isolate to inoculate globally important crops like cassava, could reduce the need to apply phosphate fertilizer to agricultural soils and still obtain the same amount of production at harvest (Ceballos *et al.* 2013). To date, research in AMF has

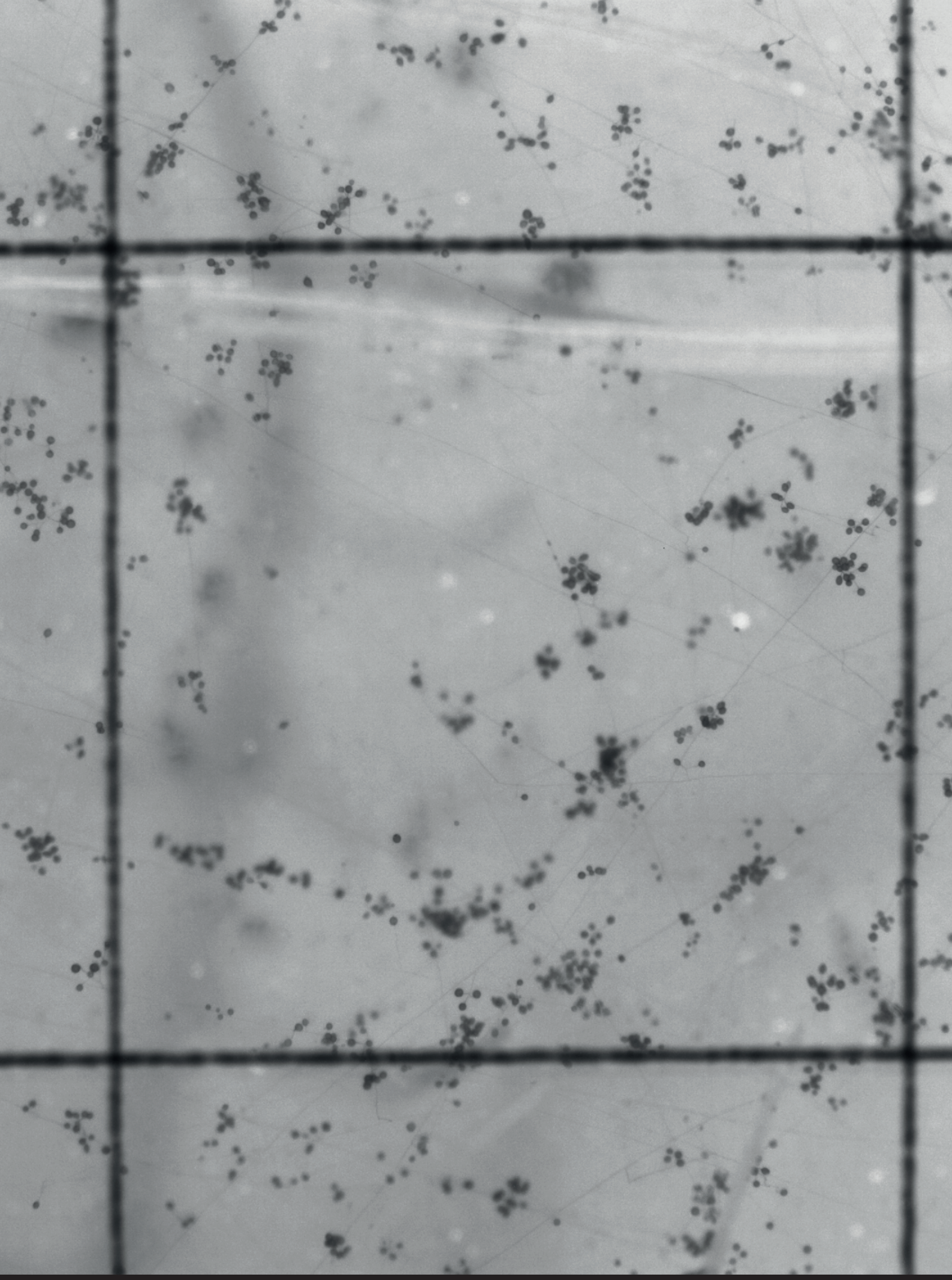
found that genetic variation in AMF occurs concomitantly with phenotypic variation in the host plant (Koch *et al.* 2006). This study is the first time where a causality link is demonstrated between the genetic variation in *R. irregularis* and the plant growth response *to R. irregularis*. Because of the causality link found in this study between the genotype of the fungus and the plant growth response, we could expect that similar isolates of this *R. irregularis* strain could produce similar effects on the host plant on similar conditions. Finally, these results open a new series of possibilities to improve crop growth by screening representative AMF species or isolates within a clade, instead of testing isolates without taking into account their phylogenetic information.

Conclusion

In this study we reported a link between the fungus genetic variation, phenotypic variation and the variation of the host response of an AMF population. The use of molecular tools to infer the relationships between different taxa, coupled with well phenotyping of the symbiotic fungus and the host plant response, could help to unravel the relationships between genotypes and phenotypes on the AMF-plant system. Furthermore, similar approaches using the phylogenetic signal could be performed, in order to determine if there is a link between the presence of different taxa on different biotic or abiotic conditions.

Acknowledgements

We thank Jeremy Bonvin for his help in culturing the different *R. irregularis* and cassava *in-vitro* plants. We are grateful with Lucile Muneret for her help in the fungal phenotypic measurements. We also thank Nicolas Ruch for his help in the greenhouse and CIAT (Cali, Colombia) for providing the cassava plantlets.



Counting *R. irregularis* spores

Ivan Mateus

Chapter 3. Using the intra-specific variability of host plant and the fungal partner to understand the molecular mechanisms of the mycorrhizal symbiosis.

Ivan D. Mateus-Gonzalez[1], **Frédéric G. Masclaux**[1,2] , **Edward Rojas**[1], **Consolée Aletti**[1], **Romain Savary**[1] , **Cindy Dupuis**[1] **& Ian R. Sanders**[1]

[1]*Department of Ecology and Evolution, University of Lausanne, Biophore building, 1015 Lausanne, Switzerland*
[2] *Vital-IT, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*

Abstract

The symbiosis between 74% of plants species and arbuscular mycorrhizal fungi is probably the commonest plant-microbe association on earth[1]. These fungi impact plant growth and are a major driver of plant diversity and productivity[2,3]. All globally important crops form this symbiosis. Because of their potential to improve food security[4] there is enormous interest in understanding which genes are involved in the symbiosis. Forward genetics approaches allowed the identification of a "mycorrhizal toolkit", common to most plants[5–7]. This approach is largely limited to discovering individual plant genes necessary for the establishment of the symbiosis. We obtained a wider picture of the molecular mechanisms of an established functioning symbiosis by looking how the intra-specific variability of both partners affected the gene-transcription of the other partner. We conducted large-scale RNA sequencing in different cassava (*Manihot esculenta*) cultivars inoculated with genetically different isolates of the mycorrhizal fungus *Rhizophagus irregularis*. We observed several cassava genes that were significantly differentially transcribed among the plants inoculated with different fungal isolates. We also observed a large number of fungal genes that were significantly differentially transcribed among the fungus coexisting with different cassava cultivars. While the expected regulation in the common mycorrhizal toolkit occurred, a surprisingly large number of genes from both partners, 1634 and 1807 from cassava and *R. irregularis*, respectively, were significantly influenced by the type of plant and type of fungus. Clustering genes into co-expression networks, based on their transcription patterns, revealed 317 *R. irregularis* and 120 cassava 'key genes' that displayed either a positive or negative interaction with genes of the same or the partner organism, allowing us to describe important gene-functions. We showed several cassava and *R. irregularis* genes involved in the cell organization inside the roots which transcription was highly correlated. Furthermore, we identified *R. irregularis* and cassava genes which transcription was highly correlated with fungal colonization. Our study reveals that both plant and fungal genetic variation plays a strong role in the shaping of the mycorrhizal transcriptome and shows that such approaches are necessary in uncovering the molecular mechanisms of this globally important plant-microbe symbiosis.

Identification of genes involved in the symbiosis has largely relied on forward genetics, using plant mutants in which colonization by the fungus is disrupted. These studies were mostly conducted at very early stages of the association, usually up to 21 days following inoculation, even though the

association exists during the complete lifetime of the plant[8]. This approach has uncovered plant genes that are crucial for the development of the fungus inside the plant; the so-called “mycorrhizal toolkit”. These genes are highly conserved throughout much of the plant kingdom[9]. Very few fungal genes involved in the symbiosis have been discovered[10–12]. The lack of a reliable transformation system for the fungus has made the discovery of fungal genes in this symbiosis extremely challenging. The forward genetics approach is limited in that because it relies on identifying the cause of a clear qualitative phenotype and has been restricted the investigation of a small number of genes in the symbiosis that essentially determine whether or not a symbiosis between the partners will be established; usually on the basis of arrested fungal development. Genes from the two partners are involved in an established functioning symbiosis beyond these early stages, and how they interact, have not been deeply investigated. This is necessary because the beneficial outcome of the association for the plant is not a foregone conclusion just because the fungus becomes established. Plant species vary enormously in their growth response to a given AMF taxa[3], but also the growth response of one plant species varies enormously according to the identity of the AMF taxa, with many interactions being negative[13,14]. Presumably, in such interactions the ‘toolkit’ of mycorrhizal genes is also switched on. Even different sibling fungi of the same AMF parent can lead to enormous variation in plant growth[15]. These studies indicate that underneath an overlying “mycorrhizal toolkit” that determines whether the symbiosis will be established, there must be also an important transcription variation of different co-regulated genes in the plant and the fungus that allows a variable plant growth response to the fungus and *vice versa*. Such interactions have never been investigated at the whole transcriptome level. Given that the overriding interest in the symbiosis is due to the capability of the fungus to greatly improve plant growth, the lack of knowledge about these genes and their interactions represents a lacuna in our much-needed understanding of the molecular genetics of this symbiosis.

We sequenced RNA transcripts from the roots of five genetically diverse cassava cultivars (CM6438-14, COL2215, BRA337, CM4574-7, CM523-7), each inoculated with one of two genetically different isolates of the mycorrhizal fungus *Rhizophagus irregularis* (isolate DAOM197198 and B1) or mock-inoculated (Extended Data Figure 1; Supplementary Table 1). We chose cassava because it is a globally important crop species that feeds almost 1 billion people daily[16]. Significant yield increases have been shown in numerous field trials in response to AMF inoculation, including *R. irregularis*[4,17]. The fungus significantly differentially colonized the cassava cultivars (Extended Data Figure 2a;

Supplementary Table 2a). Height, aboveground, tuberized roots, belowground and total dry weight varied significantly at least in one cultivar between a mycorrhizal and non-mycorrhizal plant. (Extended Data Figure 2b-g; Supplementary Table 2b-g). Because of the use of the natural variation of cassava and AMF in the experimental design of this study, we had a diverse dataset for identifying plant and fungal transcripts that were significantly up- and down-regulated in symbioses that differed in benefit. This allowed us, to separately test and quantify, the contribution of the plant and fungal genetic background to gene transcription in the symbiosis and to show gene co-expression networks between the plant and the fungal partner, while reducing the need to rely strongly on fold-change values.

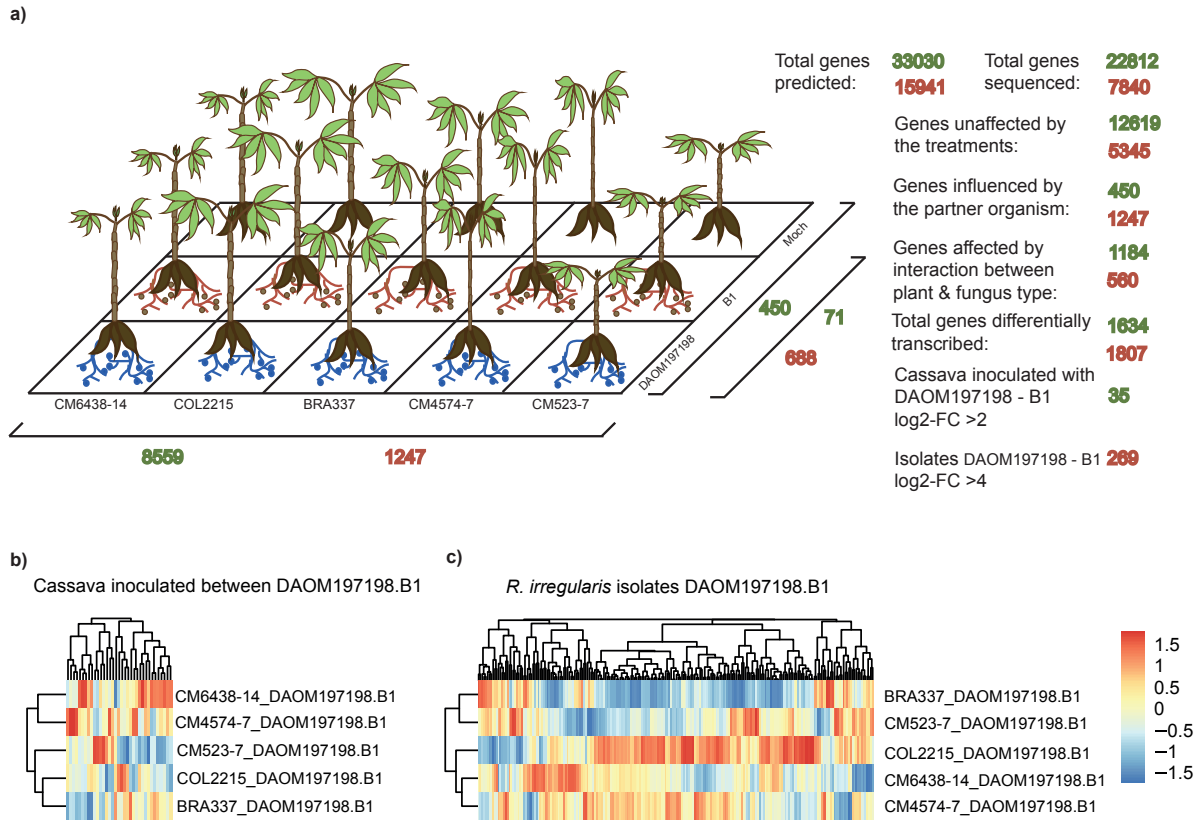


Figure 1. Variation in cassava and *R. irregularis* gene transcription accruing to plant cultivar and AMF isolate. **a)** Number of genes transcripts. Cassava genes are reported in green, while *R. irregularis* genes are reported in red. **b)** Log2Fold-change transcription between plants inoculated with DAOM197198 and B1 for each cassava cultivar. Genes where log2fold-change values higher to 2 are reported. Genes (columns) and fold-change by cassava cultivar (rows) are clustered by similarities in their fold-change. **c)** Log2fold-change transcription between *R. irregularis* isolates DAOM197198 and B1 for each cassava cultivar. Genes where log2Fold-change values higher to 4 are reported. Genes (columns) and fold-change by cassava cultivar (rows) are clustered by similarities in their fold-change.

A large number of gene transcripts were recovered from the five cassava cultivars, representing 69% of the total predicted gene number[18] (Figure 1a). The transcription of 8559 cassava genes differed significantly among the genetically different cassava cultivars but was not affected by the mycorrhizal

symbiosis (Figure 1a; Supplementary Table 3a). Another 450 cassava genes were differentially transcribed only among the three mycorrhizal treatments but not affected by different cassava cultivars (Figure 1a; Supplementary Table 3b). Only a very small number of cassava genes (71) were differentially transcribed between plants inoculated with the two different fungal isolates and that also showed consistent transcription patterns across all five cassava cultivars (Figure 1a; Supplementary Table 3b). In contrast, the transcription of a surprisingly large number of cassava genes (1184) was significantly affected by the genetic identity of both the plant and the mycorrhizal fungus (Figure 1a; Supplementary Table 3c). Some cassava genes displaying large fold changes in transcription (35 genes) between the two fungi, were affected by the identity of the host plant cultivar (Figure 1b; Supplementary Table 3d). We also tested whether expected homologs of several genes that are known to be up- or down-regulated in *Medicago trunculata*[19] in reponse to inoculation (part of the "mycorrhizal toolkit") were also affected by the three inoculation treatments. Out of 12 cassava homologs, of genes known to be up- or down-regulated in the mycorrhizal symbiosis at 21 days after inoculation, we found significant differential transcription in ten of the genes when comparing all cassava cultivars inoculated with AMF compared to the mock inoculated plants (Figure 2; Supplementary Table 4). The number of different cassava gene transcripts did not differ significantly among the mycorrhizal treatments, neither among the cassava cultivars (Extended Data Figures 3a-c), indicating that the dataset was reliable for detecting true transcriptional differences.

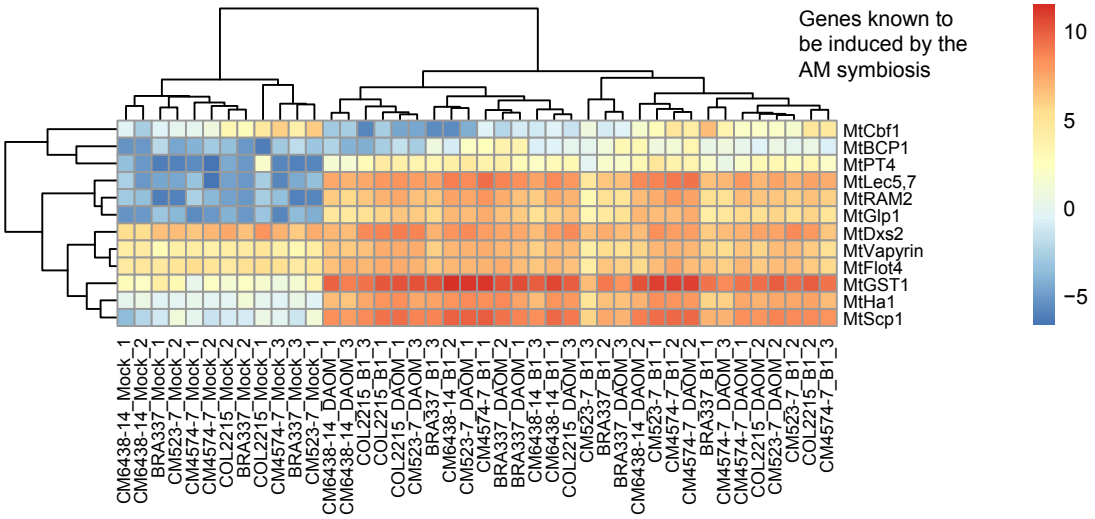


FIGURE 2. Log2-counts per million in sequenced samples (columns) on orthologs of different genes known to be involved in the arbuscular mycorrhizal symbiosis (rows), that were present in this study. Genes where chosen from Hogekamp & Küster, 2013 (Supplementary table 4). We clustered the samples by their similarities in log2-counts per million in each gene. We also clustered the genes by their similarity in transcription across samples.

The 7840 recovered fungal gene transcripts represented 49% of the total predicted gene number[20] (Figure 1a). Very few fungal transcripts were recovered from samples that were mock inoculated (Extended Figure 3a-c). Genetic variation among the cassava cultivars had a strong effect on fungal gene transcription, where 1247 fungal genes were differentially transcribed (Figure 1a; Supplementary Table 5a). Differential transcription of 688 fungal genes occurred between the two fungal isolates irrespective of the cassava cultivar (Figure 1a; Supplementary Table 5b). However, a substantial number of fungal genes (560 genes) display a strong interaction effect, were differentially transcribed due to both the fungal isolate and the cassava cultivar (Figure 1a; Supplementary Table 5c). As with the cassava genes, a substantial number of the fungal genes (269 genes) showed large fold changes in transcription between the two fungi but their expression was strongly influenced by the cassava cultivar (Figure 1c; Supplementary Table 5d).

Taking these results together, we show that genes are expressed in the symbiosis, and whether they are up- or down-regulated is highly dependent on both the plant and fungal genetic identity. We observed that cassava gene expression response to the two fungal isolates varied mainly in functional processes such as signalling, processes involving the cell wall and proteolysis (Extended Data Figure 4a). Furthermore, variation in fungal gene transcription between the fungal isolates DAOM17198 and B1 was mainly in functional processes related to auxin hormones, signalling, MAPK, proteolysis and the production of secondary metabolites (Extended Data Figure 4b). Remarkably, we found that the plant transcriptional response to the fungal isolates was not identical between the cultivars but dependant on the identity of the cassava cultivar (i.e. abiotic-stress processes displayed a high fold-change in cultivar COL2215 and CM523-7, but low differences in cultivar CM4574-7; Extended Data Figure 4a). We found the same pattern in fungal transcription, when looking at differences in gene transcription between isolates DAOM197198 and B1 on the different host plants.

No studies have recovered both plant and mycorrhizal fungal transcripts on the arbuscular mycorrhizal symbiosis and looked for a correlation in their transcription between the two organisms. Recovering cassava and fungal genes a substantial proportion of which varied greatly in transcription levels among both the different plant and fungal treatments, provided us with an ideal dataset to investigate if there were correlations in transcription between the plant and the fungal genes or if the transcription of fungal or plant genes was correlated to the plant phenotype. Clustering the genes into cassava gene

modules and fungal gene modules, using a gene co-expression network approach[21], allowed us to identify 9 and 20 modules respectively of co-transcribed genes, respectively (Extended Figure 5ab; Supplementary Table 6ab). Significant correlations occurred between a large number of the modules, where 8 out of the 9 cassava modules and 10 out of the 20 fungal modules were linked to at least one module of the other partner species (Extended Data Figure 5c; Figure 3a). We did not consider interactions between modules of the same species. To confirm that the modules were biologically meaningful, we randomly attribute the genes in the data set to the cassava and fungal modules. We applied the gene co-expression network approach to the modules containing random allocated genes. We observed that random cassava and fungal modules containing 100 random genes corresponded to no significant GO terms for both random modules (Extended Data Figure 6a). This random dataset yielded zero connections among the plant and fungal modules (Extended Data Figure 6a). This indicated that the observed gene co-expression network was unlikely to have occurred by chance. We then, applied the gene co-expression network approach to the 8 cassava and 10 fungal correlated modules. We found 'key genes' that were highly representative of the modules and highly correlated to the counterpart module (120 cassava genes and 317 *R. irregularis* genes; Supplementary Table 6cd). This information allowed us to disentangle the gene-gene interactions between fungal and plants genes.

We constructed correlograms that allow us to detect how genes involved in different functions were correlated to genes within the same organism, and to genes in the partner organism. As an example for the analysis of this data, in the *R. irregularis* 'key genes' dataset we found 10 fungal genes involved in the secretory pathway (*KEL1-like, Arrestin, Bph1p, Sfb3p, Srp102p*, ADP-ribosylation factor 5A and 6, *Tim17p, Ntf2p* and a Cysteine and glycine-rich protein) that reveal at least two major functions based in the correlogram to other fungal and plant genes (Figure 3c; Supplementary Table 6cd). In addition, we showed how these gene could have antagonistic functions (i.e. fungal genes involved in the secretory pathway: *Ntf2p*, ADP-ribosylation factor 5A, 6 and *Tim17p* were positively correlated to two fungal ABC transporters and negatively correlated to important plant genes such as phosphate transporters and two sucrose synthases. On the other hand fungal genes *KEL1-like, Arrestin, Bph1p* and *Sfb3p* display an opposite strong correlation to the previous fungal and plant genes; see Supplementary Table 7 for the correlogram analysis on fungal and plant functions).

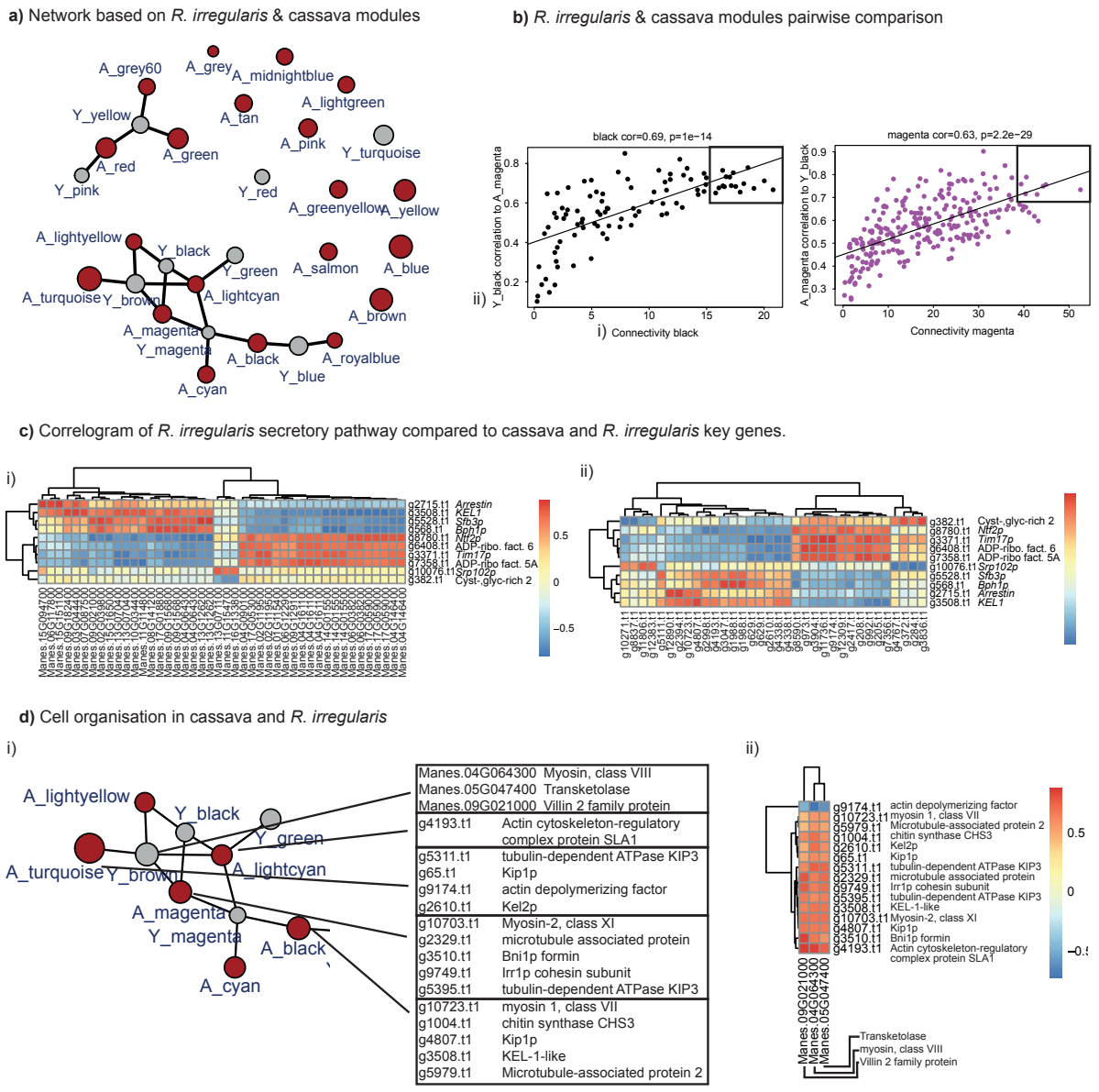


FIGURE 3. Gene co-expression network analysis. **a)** Network of *R. irregularis* and cassava modules. Correlated modules between the fungal and plant host are linked. Dark red circles represent *R. irregularis* modules. Grey circles represent cassava modules. The size of the circles is proportional of the number of genes contained in each module. Name of modules refer to modules colors identified in Extended Figure 5. **b)** Example of pairwise comparison between a fungal and a cassava gene. We characterized the genes inside each module by two measurements: i) connectivity of each gene (x-axis), which represents how a gene is connected to the other genes inside the module. The higher the value, the most representative this gene is inside the module. ii) correlation of each gene to the partner organism correlated module (y-axis). We selected the 10 % top genes (quantile 0.9) that display the higher connectivity value and a correlation score higher than the absolute value of 0.8, to analyse in downstream statistical analysis. All the module pairwise comparisons resulted in a dataset of 'key genes': 317 *R. irregularis* and 120 cassava genes that were highly connected within each module and highly correlated to the partner organism. **c)** i) Correlogram of *R. irregularis* genes involved in the secretory pathway correlated to a random sampling within the 120 cassava key genes. Each square represents the correlation coefficient between the gene in row and the gene in column. Genes in rows and columns were clustered by there similarity in their correlation coefficient. ii) Correlogram of *R. irregularis* genes involved in the secretory pathway correlated to a random sampling within the 317 *R. irregularis* key genes. Each square represents the correlation coefficient between the gene in row and the gene in column. Genes in rows and columns were clustered by there similarity in their correlation coefficient. **d)** i) *R. irregularis* and cassava genes involved in cell organisation. We show per organism and per module, the genes involved in cell organisation that we found that were highly correlated to the partner organism. ii) Correlogram of the *R. irregularis* and cassava genes involved in cell organisation. Genes in rows and columns were clustered by similarities in their correlation coefficient. The information of genes and function of the 'key genes' is found in Supplementary Table 6.

We found that fungal and plant 'key genes' issued from correlated modules between cassava and *R. irregularis*, involved in the cell organisation function, were highly positively correlated Figure3d; Supplementary Table 6d). As expected there is a coordinated re-organisation of the cytoskeleton of both partners showed by how actin and myosin related genes from both organism were highly positively correlated. Note that expected biological events were detected as we observed a strong positive correlation between actin and myosin involved genes, and a strong negative correlation between actin involved genes and an actin depolymerisation factor (Figure 3d).

We applied the same methodology to detect genes which transcription was correlated to the fungal colonization and the host plant growth. Despite that we found one cassava and one *R. irregularis* module that correlate to plant growth traits, we did not observed genes that highly correlates to the plant growth traits. However, we observed that two *R. irregularis* modules and one cassava module strongly correlates with fungal colonization. This analysis showed that fungal genes involved in cell organisation, a chitin synthase, a monoterpene hydrolase and a major facilitator superfamily transporter among others, and plant genes such as sucrose synthase and malate deshydrogenase could be involved in the fungal colonization of plant roots.

The methodology in this study allowed us to discriminate between different functions of fungal and plant genes. It also allowed us to found biological relevant processes of the mycorrhizal symbiosis. Finally, it allowed us to propose candidate plant and fungal genes that could be responsible of the fungal colonization of plant roots, demonstrating the power of the gene co-expression network analysis in the study of the mechanism of the arbuscular mycorrhizal symbiosis.

Plants become mycorrhizal a few days after their roots enter the soil and remain in the association their entire lives. However, almost all the genes we know of that are involved in the symbiosis are only important in the first few days of the interaction; mostly acting as part of a switch allowing the symbiosis to become established. The importance of the symbiosis, however, stems from the fact that the fungi have the capacity to potentially make plants grow better over their lifetime and we already know that this varies greatly according to the plant species and identity of the fungus[3]. We show that after these switches have occurred, the majority of genes and how both partners transcribe them is

highly dependent on the combination of plant and fungal genetics. Approximately 225000 different plant species form this symbiosis[1] and yet we show that an enormous amount of variation in gene transcription in this symbiosis is affected by genetic variation within one plant and one fungal species and how they interact together. The fact that such a large number of genes in the plant respond in a completely different way according to which fungus occupies their roots, and that the effect can be completely inversed due to the genetic identity of the plant, tells us that simple one plant – one fungus comparisons to a non-mycorrhizal control (a situation which essentially doesn't occur in nature), or the use of mutants with arrested development of the symbiosis are not going to enable researchers tease apart which genes are important and responsible for the improved plant growth capabilities afforded by this symbiosis. Had only one cassava and one AMF fungal then been chosen for this work then we would likely wrongly categorise given differentially transcribed genes into those that are up-regulated by the AM symbiosis and those that are down-regulated. In reality, even within the same species, the change in gene transcription of one gene could be in the opposite direction or the gene might simply not be affected at all when taking into account plant or fungal genetics. Given the high variation in the plant and fungus can lead to different plant growth responses to the symbiosis, such variation should be embraced by researchers as a tool. Combined with quantitative genetic techniques, such variation is an extremely powerful tool for resolving how variation in plant or fungal genomes results in a highly beneficial symbiosis; which surely should be one of the most important goals in research on this symbiosis.

Methods

Plant material and growth conditions

We used five different cassava (*Manihot esculenta* Crantz) cultivars (V1:CM6438-14, V4:COL2215, V5:BRA337, V6:CM4574-7, V8:CM523-7) obtained from CIAT. We micro-propagated the *in vitro* plants using cuttings from lateral and apical meristems of approximately 1 cm long. Explants were grown on MS medium with 14 h long daylight per day (light intensity 100 $\mu E.m^{-2}s^{-1}$) at 25°c in a culture chamber (Sanyo MLR-351 H). After 8 weeks of growth, plants were hardened off in greenhouse conditions (28°C, 16 h daylight and 70% RH) for four weeks. The hardening substrate was an autoclaved mixture of perlite and peat moss (1:1). Hardened plants were transplanted to final steam sterilized (100°C, 25 min) substrate; composed by perlite, moss peat, inert clay and sand (1:1:1:1) (v/v). After inoculation plants were grown in greenhouse conditions with enough irrigation during 18 weeks. We disposed the plants in a randomised block design in the greenhouse, where each block contained one replicate of the different treatments. Additionally, we randomised the position of the blocks every 4 weeks in order to avoid block effects.

AMF inoculation

We used *R. irregularis* isolates DAOM-197198 (originally collected in Pont Rouge, Canada) and isolate B1 (originally collected in Tänikon, Switzerland) to inoculate the plants. Individual spores from *in-vitro* culture split-plates were extracted by dissolving the agar in a solution of citric acid (6%), and passed through a sieve (30 microns). We then inoculated the plants with 300 spores of isolate DAOM-197198, 300 spores of isolate B1, and distilled water for the non-mycorrhizal plants. The inoculation was made diluting the spores in 10 ml of distillate water, and applying it directly to the root zone.

Phenotypic measures

After 18 weeks, plant height was measured. We then collected and dried the plants for eight days at 72°C. Dry weight was measured for the shoots, the tuberous roots and the fine roots separately.

Root colonization measurements

We randomly select some fine roots to measure the colonization by AMF. We cleared the roots with

10% KOH for 4 hrs, acidified them with HCl (1%) during 5 minutes and stained them with trypan blue (0.10% in a lactic acid-glycerol solution) overnight. Root colonization was determined using 10 replicates of each treatment by the grid line intersect method (Giovannetti & Mosse 1980).

RNA extraction, library preparation and sequencing

We sequenced 3 replicates of each treatment of each cultivar. Approximately 150 mg of fine roots were randomly collected. The RNA was extracted from the root tissue using the Maxwell plant RNA kit (Promega). We used the *TruSeq Stranded mRNA* Library Prep Kit, set B to make the libraries preparation. The libraries were sequenced using illumina Hi-seq 100nt paired-end technology. In all the steps of the RNA extraction, library preparation and sequencing, we randomized the samples in order to avoid batch effects at these steps.

Bioinformatic analysis

We processed the raw reads with the script Tagcleaner.pl to trim Illumina adapters (Schmieder *et al.* 2010). Reads were quality-filtered and trimmed using PrinSeq- lite.pl version 0.20.4 (Schmieder & Edwards 2011). Low quality 3'-ends were trimmed and reads containing uncalled bases (N) removed. Only reads longer than 50 bp were kept for further analyses.

To obtain the cassava and *R. irregularis* data sets separately, we aligned the total reads to the *Manihot esculenta* reference genome (*M. esculenta* V6.1, Phytozome V11) with a two-pass method with the STARstatic 2.4.0 aligner (Dobin *et al.* 2013). Then, for the cassava dataset, we extracted the aligned reads with samtools (Li *et al.* 2009) (bam2fq option). For the *R. irregularis* dataset, we ensured that the sequence reads were not from cassava by using only the unmapped reads to the cassava reference genome. By doing this, we obtained two separate datasets for cassava and for *R. irregularis*. We then, used the pseudo aligner kallisto (Bray *et al.* 2016) in order to produce count tables of the number of counts per transcript per sample. For *cassava* we used the online gene prediction produced for Mesculenta_V6.1. For *R. irregularis*, we produced a gene prediction of the reference genome N6 (Lin *et al.* 2014) with Augustus (see Gene prediction on *R. irregularis*).

Gene co-expression networks between cassava and AMF and the plant phenotype

We used a weighted gene co-expression network analysis (WGCNA) (Langfelder & Horvath 2008) in

order to create modules (cluster of similar expressed genes) of cassava and *R. irregularis*. For cassava we used the genes that were significantly differently expressed between the AMF treatments and the non-mycorrhizal treatment. For *R. irregularis* we selected all the genes expressed in the experiment. For Both organisms we used the default WGCNA 'step-by-step network construction' analysis. We first calculated the adjacencies between genes and constructed a topological overlap matrix. We then produced a hierarchical clustering tree with the dissimilarity of the topological overlap matrix and we selected the modules by using the dynamic three cut standard method. Finally, we merged similar modules by calculating the module eigengenes, clustering them and assigning a distance threshold. The parameters used were soft-power of 12 and minimum module size of 50 genes. We merged the modules with a Distance threshold cut of 0.1.

We associated the modules between the two organisms by correlating the cassava modules eigengenes against the *R. irregularis* modules eigengenes using the default WGCNA 'relating modules to external information' analysis. We calculated the gene significance for each gene to the correspondent correlated module (Ex: gene1-cassava vs. module1- *R. irregularis*) and we calculated the module membership of each gene to each module. We used this information in order to obtain the 'top genes' of a module (the genes that are the mostly correlated to the other organism modules (top 10% quantile) and that display a high module membership (>0.8)). For further investigations on the correlated genes between cassava and *R. irregularis*, we selected the modules that displayed a high correlation with a p-value < 0.001 between the two organisms. This cut-off threshold was chosen because when doing pairwise comparison between the cassava and *R. irregularis* modules (180 comparisons) with a probability of 0.05 we could expect to found by chance 9 correlated modules. By choosing a cut-off of 0.001 we could expect by chance only 0.18 correlated modules.

We also performed the same analysis as before ('relating modules to external information') to correlate the cassava and *R. irregularis* modules to the plant phenotype. We conserved the same parameters and threshold cut-off for these comparisons.

Gene Prediction on R. irregularis

Prediction of protein coding genes was performed with the *ab initio* gene prediction tool Augustus

based on hidden Markov model. Augustus was trained based on a dataset of Aspegillus proteins. The dataset of Aspegillus proteins was generated by retrieving proteins reviewed by Swiss-prot from www.uniprot.org. Scipio (Keller *et al.* 2008) and BLAT (Kent 2002) were used to align the Aspergillus proteins on the N6 genome and to define the gene structure.

The structure of the genes was used to train and optimize Augustus following Augustus instructions (http://www.molecularevolution.org/molevolfiles/exercises/augustus/training.html). We generated hints for Augustus predictions from DAOM197198 RNA-seq data using STAR aligner and cufflinks. Augustus predictions with hints were performed on the repeat-masked N6 genome using the trained parameters for the species.

Go terms, functional classification, pathway diagrams and Blastp

We obtained the GO terms correspondent to each predicted genes of both organisms using the Blast2GO tool with standard parameters (Götz *et al.* 2008). Blast was run against the NCBI-NR sequence database. We then used the R package 'GOSeq' to conduct the GO term enrichment analysis (Young *et al.* 2010). To detect which biological processes are the more representative in each module, we took as universe all the sequenced genes and perform the enrichment analysis on all the genes contained in the module.

We classified each gene into a functional category by using the Mercator sequence annotation tool (http://www.plabipd.de/portal/mercator-sequence-annotation) with the default parameters adding Interpro scan, ORYZA and CHLAMY parameters. The different pathways diagrams were made using MapMan 3.5.1R2 (Thimm *et al.* 2004) providing the functional classification obtained by the Mercator sequence annotation. We provided an experimental dataset to MapMan obtained by selecting the logFC of the significantly differently expressed genes between the conditions. In order to confirm the protein resulting for each gene, we performed a Blastp query to the non-redundant protein sequence (nr) database (NCBI).

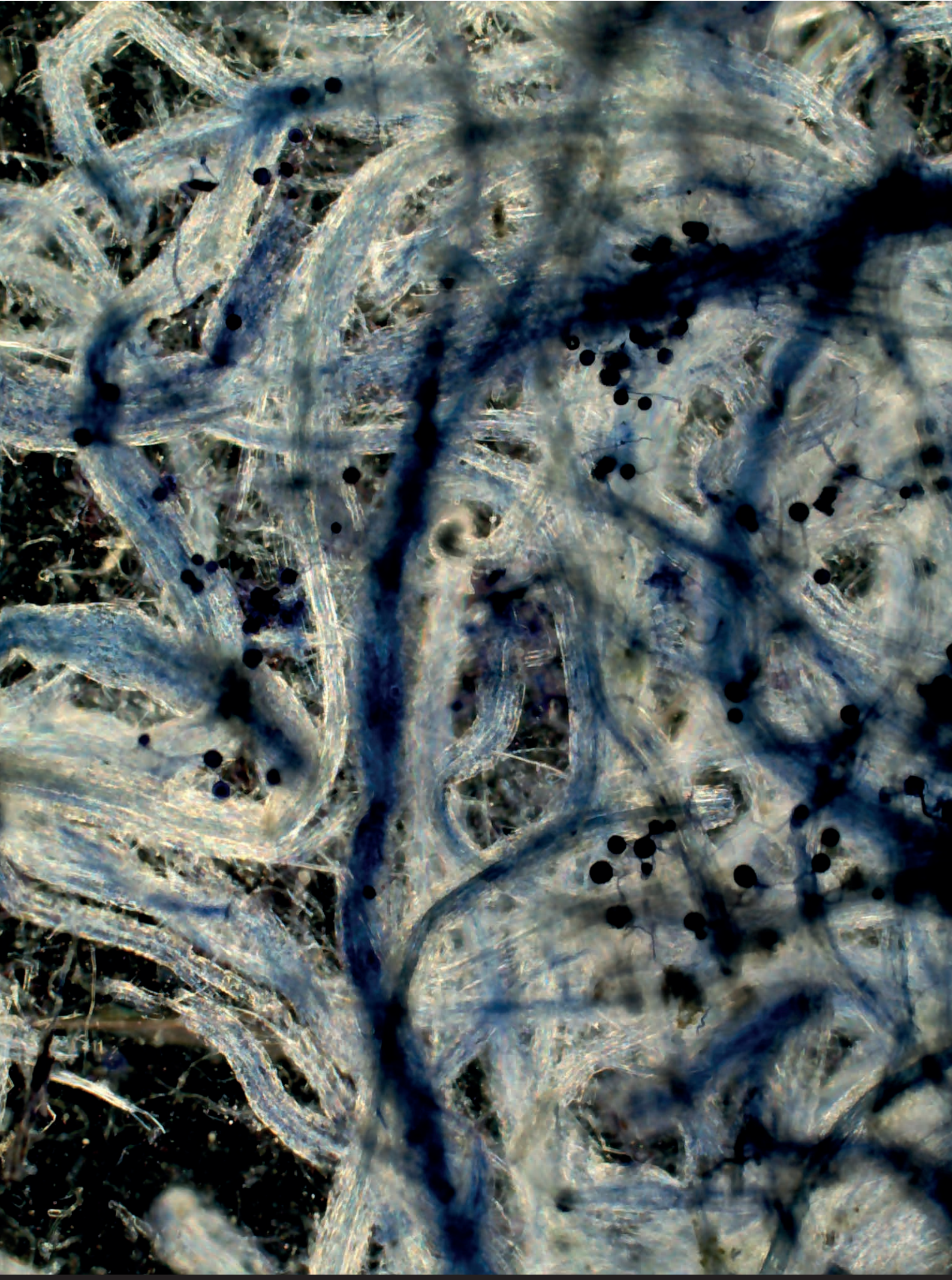
Phenotypic measurements

We used a binomial generalized linear model for the analysis of the fine roots colonization measurement. For all the other phenotypic variables, we used a mixed model including the block as a

random effect and tested the significance of the fixed effect of the cassava cultivar, the AMF treatment and their interaction.

Differential expression analysis

After the normalization by transcript length done by kallisto, we worked at the gene-level by using the tximport R function (Soneson *et al.* 2015). We then transform the raw data into a DGElist object (edge R package) (Robinson *et al.* 2009). We proceed to filter the genes that contained less than 100 counts for at least 3 samples and normalize the samples by the library size using the calcNormFactors function (edge R package). We proceeded to transform the count data to logCPM and estimate the mean-variance relationship using the voom function (limma R package) (Smith 2005). Finally, we used the lmFit, ebayes and topTable functions (limma package) to fit a linear model, compute the statistics for differential expression and extract the differentially expressed genes (DEG) between the treatments.



Staining of cassava roots inoculated with *R. irregularis*

Jeremy Bonvin

Chapter 4. Role of host intra-species variability on coexistence and effect on cassava (Manihot esculenta) of two isolates of Rhizophagus irregularis revealed by RNA-seq

Ivan D. Mateus-Gonzalez[1], Frédéric G. Masclaux[1,2] , Edward Rojas[1], Consolée Aletti[1], Romain Savary[1] , Cindy Dupuis[1] & Ian R. Sanders[1]

[1]*Department of Ecology and Evolution, University of Lausanne, Biophore building, 1015 Lausanne, Switzerland*
[2] *Vital-IT, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*

Abstract

Arbuscular mycorrhizal fungi (AMF) are soil microorganisms that can associate with ~74% of land plants and can enhance plants nutrients uptake. Many studies have attempted to understand the effect of AMF on plant growth, but their set-up mostly involved one plant – one fungus experimental designs. In natural conditions, the plants are colonized by more than one AMF. Therefore, it is important to understand how AMF coexist in the roots and what is their combined effect on the host plant.

In this experiment we performed RNA-seq to test if the host intra-species variability has an effect on the coexistence of two *R. irregularis* isolates. For this, we used two *R. irregularis* Isolates co-inoculated and single-inoculated in three different cassava cultivars.

We showed that the host-plant intra-specific variability influenced the coexistence of two *R. irregularis* isolates and that the coexistence of both isolates differently affected the different host plants. We also observed a conserved mycorrhizal gene-set on cassava. We observed that the two *R. irregularis* isolates possess a different repertoire of transcribed genes. Finally, we observed three *R. irregularis* genes that were probably related to the direct interaction between these two isolates.

We demonstrated a strong role of both the fungal and host plant intra-species variability on the coexistence of two *R. irregularis* isolates, and their effect on the host plant. This study suggests that both fungal and host genetic variability should be comprised in the study of the AMF-plant symbiosis. Finally, we showed that a *MATA-HMG* gene is only activated in the presence of a second isolate suggesting that there is probably recognition and mating between the two isolates.

Introduction

Arbuscular mycorrhizal fungi (AMF) are soil microorganisms that are present worldwide (Davison *et al.* 2015). These fungi are root endosymbionts that can associate with ~74% of land plants and can enhance plants nutrients uptake (Smith & Read 2008). It has been shown that AMF can impact plant growth (Koch *et al.* 2006) and plant community diversity (Heijden *et al.* 1998). Recently, AMF have been used in agriculture to enhance crop yield (Ceballos *et al.* 2013), demonstrating their potential to

be used as a non-chemical fertilizer. Many studies have attempted to understand the effect of AMF on plant growth, but their set-up mostly involved one plant – one fungus experimental designs. In natural conditions, the plants are colonized by more than one AMF. Therefore, it is important to understand how AMF coexist in the roots and what is their combined effect if compared to when they colonize independently the host plant.

Nevertheless, a few studies have addressed whether co-inoculation by two different AMF species has a bigger impact on plant growth compared to single-inoculations. Co-inoculation of two AMF species can: (1) enhance plant growth compared to the single-inoculations as in the case of *Andropogon gerardii* (Gustafson & Casper 2006), (2) decrease plant growth as seen in *Persea Americana* mill (Violi *et al.* 2007), or (3) not affect plant grown as observed in *Tripleurospermum inodorum* and *Calamagrotis epigejos* (Janoušková *et al.* 2009). These observations suggested, that the outcome of the co-inoculations compared to the single-inoculation is variable and dependent on the type of plant and fungus used. Furthermore, the host species plays a role in the effect of co-inoculations of different AMF species on the host growth. A mix of three different AMF inoculated on *Brachipodium pinnatum* and *Prunella vulgaris* resulted in a different effect on the biomass of these plants (van der Heijden *et al.* 2003). In addition, different host plants do the symbiosis with different combinations of AMF species suggesting a role of plant selectivity, which consequence is that some plant-AMF combinations are more likely to occur than others in a certain type of host (Helgason *et al.* 2002). While these studies analysed the role of host inter-species variation in the coexistence of two AMF species, the role of the host intra-specific variation in this mediation has never been tested.

To explain the effect of fungal coexistence on plant growth, it has been suggested that functional complementarity and competition between two AMF, could explain the effect of the coexistence of two AMF on the host plant (Roger *et al.* 2013). Extra-radical mycelium of different AMF species can exploit different parts of the soil (Jakobsen *et al.* 1992). Thus, two AMF species can jointly exploit a greater soil area compared to a single AMF species, and consequently increase the resources provided to the host plant and affecting plant growth. However, species that are in competition for resources can reduce resource allocation to the host in order to allocate more energy into competition and by consequence negatively affect the host plant growth (Roger *et al.* 2013). Furthermore, AMF can exchange genetic material between different isolates by anastomosis (Giovannetti *et al.* 1999).

Isolates that 'cross', can produce progeny with different phenotypes, and can produce different effect on plants (Angelard *et al.* 2010). Hence, anastomosis is another type of interaction that could occur between coexisting AMF isolates within roots and could affect plant growth.

At the intra-species level, coexistence between two different AMF could be influenced by the genetic relatedness of these organisms, where genetically related isolates tend to coexist in more even proportions in the plant roots compared to genetically distant isolates (Roger *et al.* 2013). Hence, the genetic relatedness between two co-inoculating isolates could be a driver of plant growth, as genetically related isolates have an increased effect on plant growth compared to genetically distant isolates (Roger *et al.* 2013). However, the host plant plays an active role in the coexistence of two AMF species because the plant host can mediate their competition by changing carbohydrate supply to the different fungi (Pearson *et al.* 1993; Kiers *et al.* 2011).

Understanding the mechanisms that explain the interactions and plant response to the coexistence of two fungi is still largely unknown, because of the complexity of measuring fungal metrics in soil and inside plant roots. An option to resolve this is to perform next-generation sequencing to produce genome-wide gene transcription data (RNA-seq) on the root tissues. RNA-seq is a technique that produces large amount of informative data on the organism, allowing the obtention of information about both, the fungal and the plant gene-transcription, as long as there is a reference genome for both organisms. Nowadays, we can access several *R. irregularis* genomes(Tisserant *et al.* 2013; Lin *et al.* 2014; Ropars *et al.* 2016), and non-model plant species genomes like cassava (Bredeson *et al.* 2016).

Gene-transcription data has been used to understand the impact of AMF on the host plant, and has revealed a 'mycorrhizal toolkit' of plant genes activated by AMF (Hogekamp & Küster 2013). However, to date, all the experiments in mycorrhizal research involving gene-transcription data used the simplistic approach of looking at the effect of one fungus on one plant. Hence, the use of RNA-seq, to understand which genes are involved in the coexistence of two fungal strains, and their effect on plant growth could reveal valuable information of which genes and biological processes are involved in this more complex scenario.

In this experiment, we aimed to understand the coexistence of two different *R. irregularis* isolates at

the phenotypic and gene level and test the role of the host intra-specific variation in the coexistence of two *R. irregularis* isolates. More precisely we asked:

1) What is the role of the host intra-species variability in the coexistence of two *R. irregularis* isolates?
2) Do the host plant react differently to the coexistence of two *R. irregularis* isolates compared to when the isolates are single-inoculated?
3) Do the host plant display a conserved mycorrhizal gene toolkit at the intra-species level?
4) Do the genetically related *R. irregularis* isolates display similar gene-repertoires?
5) Can we detect direct gene-interactions between two genetically related *R. irregularis* isolates?

We used three cassava (*Manihot esculenta*) cultivars, each inoculated with the *R. irregularis* isolates DAOM197198 or B1 or co-inoculated with both isolates. In addition, we used RNA-seq, to highlight which genes could be involved in the interaction between both isolates and between cassava and *R. irregularis*.

Methods

Plant material and growth conditions

We used three different cassava (*Manihot esculenta*) cultivars (COL2215, BRA337 and CM4574-7) obtained from CIAT. We chose these cultivars because they have already been used in field experiments by the lab. We micro-propagated the *in-vitro* plants using cuttings from lateral and apical meristems of approximately 1 cm long. Explants were grown on MS medium with 14 h daylight at 25°c in a plant growth chamber (Sanyo MLR-351 H). After 8 weeks of growth, plants were hardened off in greenhouse conditions (28°C, 16 h daylight and 70% RH) for four weeks. The hardening substrate was an autoclaved mixture of perlite and peat moss (1:1). Hardened plants were transplanted to final steam sterilized (100°C, 25 min) substrate; composed by perlite, moss peat, inert clay and sand (1:1:1:1) (v/v). After inoculation plants were grown in greenhouse conditions and watered regularly for 18 weeks. We harvested the plants at 18 weeks of growth because we observed a slowdown of growth after 4 months in similar conditions in the past. We positioned the plants in a randomised block

design in the greenhouse, where each block contained one replicate of the different treatments. Additionally, we randomised the position of the blocks every 4 weeks in order to avoid block effects.

AMF inoculation

We used *Rhizophagus irregularis* isolates DAOM197198 (originally collected in Pont Rouge, Canada) and isolate B1 (originally collected in Tänikon, Switzerland; (Jansa *et al.* 2002)) to inoculate the plants. Individual spores from *in-vitro* culture split-plates were extracted by dissolving the agar in a solution of citric acid (6%), and passed through a sieve (30 microns). We then inoculated the plants with 300 spores of isolate DAOM-197198 or with 300 spores of isolate B1 or with 150 spores of isolate DAOM197198 and 150 spores of isolate B1 as a co-inoculation treatment. Non-inoculated control plants were given distilled water. The inoculation was made by suspending the spores in 10 ml of distilled water, and applying it directly to the roots.

Plant phenotypic measurements

After 18 weeks, plant height was measured. We then collected and dried the plants for eight days at 72°C. Shoots, tuberous roots and fine roots were weighted separately. We randomly select some fine roots to measure the colonization by AMF. Roots were cleared with 10% KOH for 4 hrs, acidified them with HCl (1%) during 5 minutes and stained them with trypan blue (0.10% in a lactic acid-glycerol solution) overnight. Root colonization was determined on 10 replicates of each treatment using a grid line intersect method (Giovannetti & Mosse 1980).

RNA extraction, library preparation and sequencing

Approximately 150 mg of fine roots were randomly collected for the RNA extractions. We extracted RNA from 3 replicates of each treatment of each cultivar. The RNA was extracted from the root tissue using the Maxwell plant RNA kit, Promega. We used the *TruSeq Stranded mRNA* Library Prep Kit, set B to make the library preparation. The libraries were sequenced using illumina Hi-seq 100nt paired-end technology. In all the steps of the RNA extraction, library preparation and sequencing, we randomized the samples in order to avoid batch effects at these steps. Because of low library quality, we excluded 1 co-inoculated sample of cultivar CM4574-7, 1 non-inoculated control of cultivar

COL2215, 1 sample of cultivar CM4574-7 inoculated with B1 and 1 sample of cultivar COL2215 inoculated with DAOM197198.

Bioinformatic analysis

We processed the raw sequence reads with the script Tagcleaner.pl to trim Illumina adapters (Schmieder *et al.* 2010). Reads were quality-filtered and trimmed using PrinSeq- lite.pl version 0.20.4 (Schmieder & Edwards 2011). Low quality 3'-ends were trimmed and reads containing uncalled bases (N) removed. Only reads longer than 50 bp were kept for further analyses.

To obtain the cassava and AMF data sets separately, we aligned the total reads to the *Manihot esculenta* reference genome (*M. esculenta* V6.1,Phytozome V11) with a two pass method with the STARstatic 2.4.0 alligner (Dobin *et al.* 2013). Then, for the cassava dataset, we extracted the aligned reads with the samtools bam2fq option (Li *et al.* 2009). For the *R. irregularis* dataset, we ensured that the sequence reads were strictly from AMF and not from cassava by only using the reads that did not map to the cassava reference genome. By doing this, we obtained two separate datasets for cassava and for *R. irregularis*. We then, used kallisto (Bray *et al.* 2016) in order to produce count tables of the number of counts per transcript per sample. For cassava we used the gene prediction produced for Mesculenta_V.6.1. For *R. irregularis*, we produced a gene prediction of the reference genome N6 (Lin *et al.* 2014), with Agustus. We used the *R. irregularis* N6 reference alignment because was the best single-nuclei alignment compared to other single-nuclei alignments made in that study (Lin *et al.* 2014).

Gene Prediction on R. irregularis

Prediction of protein coding genes was performed with the *ab initio* gene prediction tool Augustus based on a hidden Markov model. Augustus was trained based on a dataset of *Aspergillus* proteins. The dataset of *Aspergillus* proteins was generated by retrieving proteins reviewed by Swiss-Prot from www.uniprot.org. Scipio (Keller *et al.* 2008) and BLAT (Kent 2002) were used to align the *Aspergillus* proteins on the N6 genome and to define the gene structure (http://www.molecularevolution.org/molevolfiles/exercises/augustus/scipio.html).

The structure of the genes was used to train and optimize Augustus following Augustus instructions

(http://www.molecularevolution.org/molevolfiles/exercises/augustus/training.html). We generated hints for Augustus predictions from DAOM197198 RNA-seq data using STAR aligner and cufflinks (Handa *et al.* 2015). Augustus predictions, with hints, were performed on the repeat-masked N6 genome using the trained parameters for the species.

Functional classification

Blast was run against the NCBI non-redundant protein sequence database to find gene identity. We classified each gene into a functional category by using the Mercator sequence annotation tool (http://www.plabipd.de/portal/mercator-sequence-annotation) with the default parameters adding Interpro scan, ORYZA and CHLAMY parameters.

Sequenced sites analysis

In order to confirm AMF colonization and isolates identity in the samples, we called variants from the AMF alignment files. We used Picard tools 1.130 to modify the header and mark the duplicated reads of the BAM files. We then, used GATK 3.5 (Auwera *et al.* 2014) to split and trim the reads that contained gaps resulting from splicing events. Lastly, we called variants with Freebayes 0.9.9.2 (Garrison & Marth 2012), and we used vcffilter to filter the SNP calling with a coverage depth higher than 10 and a calling quality higher than 30. Further analysis and figures were made using R (R 2016).

Statistical analysis on plant growth

We used a binomial generalized linear model for the analysis of the fine root colonization measurement. For all the other phenotypic variables, we used a mixed model including the block as a random effect and tested the fixed effect of the cassava cultivar, the AMF treatment and their interaction. Post-hoc multiple-comparison tests were made with the lsmeans R package.

Differential transcription analysis

After the normalization by transcript length performed by kallisto, we worked at the gene-level by using the tximport R function (Soneson *et al.* 2015). We then transformed the raw data into a DGElist object (edge R package; Robinson *et al.*, 2009). We proceed to filter the genes that contained less than 100 counts and normalize the samples by the library size using the calcNormFactors function (edge R package). We proceeded to transform the count data to logCPM and estimate the mean-variance

relationship using the voom function (limma R package; Smith, 2005). Finally, we used the lmFit, ebayes and topTable functions (limma package) to fit a linear model, compute the statistics for differential transcription and extract the differentially transcribed genes (DTG) between the treatments.

Results

Cassava phenotypic response to co-inoculation

We did not observed any colonization in the non-inoculated samples. We observed a different outcome of the effect of the co-inoculation treatment on the different cultivars. There was no significant difference in colonization between the single-inoculations and the co-inoculation treatment in the cultivar COL2215 (Figure 1a). However in cultivar BRA337, we observed that the co-inoculated treatment displayed significantly higher AMF colonization than the plants inoculated with isolate DAOM197198 (lsmean: -1.27(0.35), ratio: -3.56, p-val: 0.001; Figure 1a). In addition, we observed that CM4574-7 plants that were co-inoculated displayed higher levels of AMF colonization compared to single inoculations with B1 or DAOM197198 (B1 *vs.* co-inoculation: lsmean: -0.75(0.27), ratio: -2.73, p-val: 0.017; DAOM197198 *vs.* co-inoculation: lsmean: -0.75(0.27), ratio: -2.75, p-val: 0.016; Figure 1a).

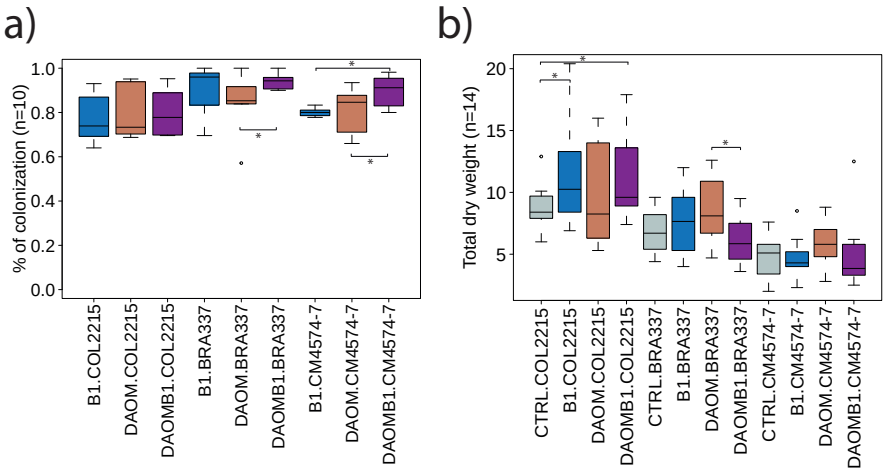


Figure 1. a) Percentage root length colonized by *R. irregularis* in all the inoculated treatments. b) Total dry weight of cassava cultivars. For all the measurements we used three different mycorrhizal treatment, a non-inoculated treatment and three different cultivars.

We observed a positive effect of the co-inoculation on the total dry weight of cultivar COL2215 compared to the non-inoculated control (lsmean: -2.90 (0.87), ratio: -3.309, p-val: 0.006; Figure 1b). The dry weight of the co-inoculated BRA337 plants was not different from the non-inoculated treatment, but significantly smaller compared to the plants inoculated with isolate DAOM197198

(lsmean: 2.44(0.87), ratio: 2.78, p-val: 0.031; Figure 1b). We observed no difference in total dry weight of the co-inoculated CM4574-7 plants compared to the other mycorrhizal and non-inoculated treatments on the cultivar CM4574-7 (Figure 1b). Other phenotypic variables measured at harvest such as height, aboveground dry weight and belowground dry weight support the previous findings of a different outcome of the co-inoculation treatment on the different cultivars (Supplementary Figure 1).

RNA-seq data analysis

The RNA-seq dataset on the cassava roots RNA allowed us to obtain for each sample, gene-transcription information on the cassava cultivars and the fungus *R. irregularis* (Supplementary Table 1). We achieved sequence saturation of the sequencing reaction for genes of both organisms meaning that the number of genes found per sample was not increasing significantly with the increase of the sequencing effort (Figure 2a-b). In addition, there was no observed mycorrhizal treatment bias on the cassava gene-transcription dataset (Figure 2c) or in the *R. irregularis* gene-transcription dataset (Figure 2c). Finally, we observed that 12 homologs of plant genes known to be involved in the symbiosis (Hogekamp & Küster 2013) were present and induced as well in our data set (Figure 2d).

Sequenced sites analysis

We extracted genomic variants for all the *R. irregularis* samples. The genomic variants found in the dataset confirmed that the colonization by AMF occurred in the expected samples. We observed that the non-inoculated treatments displayed very few *R. irregularis* variants (Figure 2e). We also observed that the number of sites did not differ between the *R. irregularis* treatments (Figure 2f). Finally, we observed that the co-inoculated samples displayed a similar proportion of alleles resulting from isolates DAOM197198 and B1, which demonstrate that both isolates colonized the roots in all the co-inoculated samples (Figure 2g).

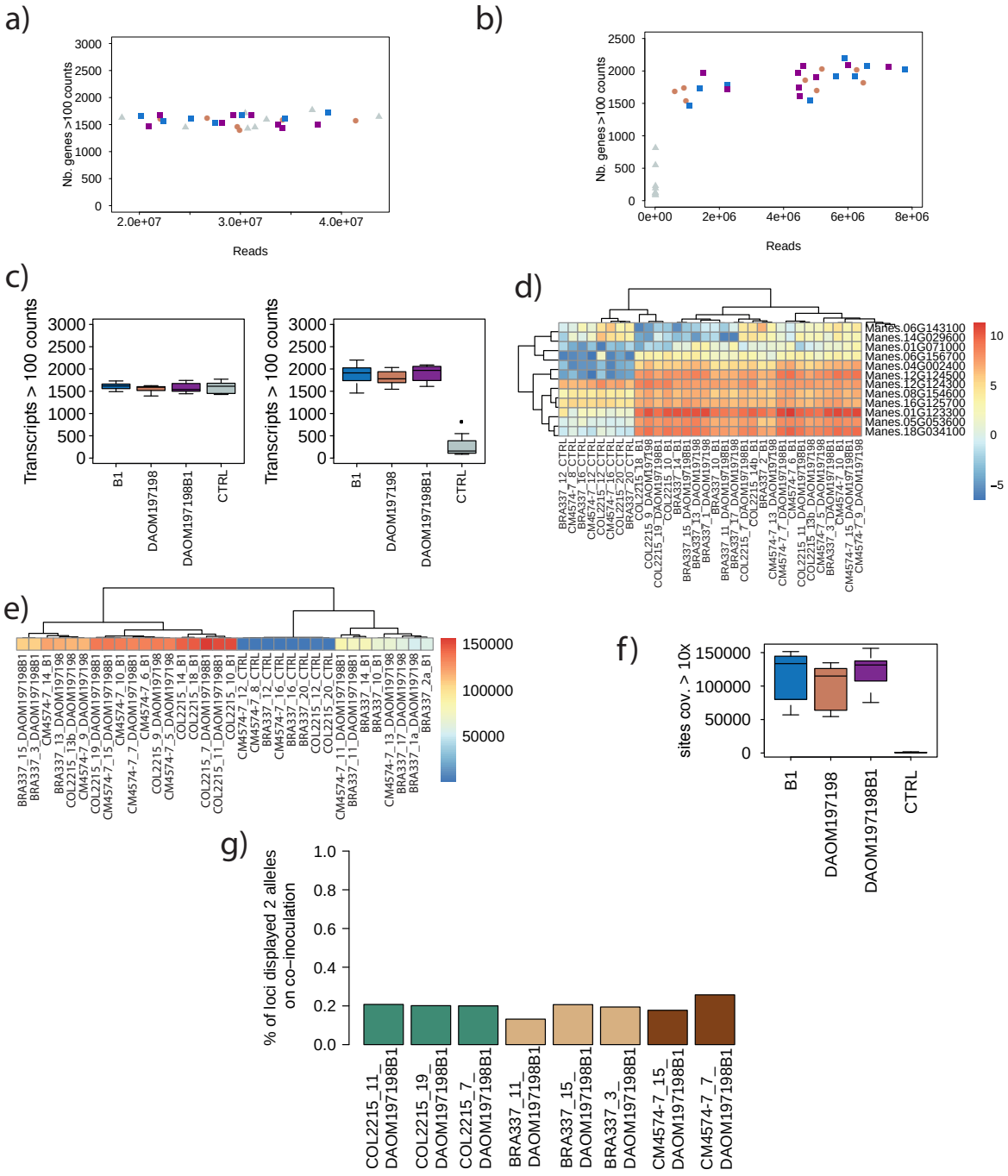


Figure 2. Summary results of the RNA-seq data-set. a,b) Cassava and *R. irregularis* rarefaction curves shows number of reads versus transcripts with more than 100 counts. c,d) Cassava and *R. irregularis* transcripts with more than 100 counts in the different mycorrhizal treatments. e) Heatmap showing the transcription of genes involved in the symbiosis (based on Hogekamp, C. & Küster, H. BMC Genomics 14, 306 (2013)). Sequenced sites analysis on *R. irregularis* samples : f) Number of sites with coverage > 10x. per sample. g) number of sites with coverage >10x by mycorrhizal treatment. h) Percentage of sites that displayed two alleles in the co-inoculations samples. We selected the sites that displayed a given reference allele for DAOM197198 and a SNP on isolate B1, we discarded sites that displayed more than one-allele. To test if the two isolates are present and to measure the intensity of the colonization, we look on the co-inoculation samples the percentage of sites that displayes both alleles (ref. for DAOM197198 and SNP for B1). V4 refers to cultivar COL2215, V5 refers to cultivar BRA337 and V6 refers to cultivar CM4574-7. DAOM refers to isolate DAOM197198, B1 refers to isolate B1, DAOMB1 refers to the co-inoculation treatments and CTRL refers to the non-inoculated treatment.

Cassava gene-transcription response

The comparison of non-inoculated cassava plants to the mycorrhizal treatments (Isolate DAOM197198, B1 and co-inoculation) resulted in 1158 genes that showed significantly different levels of transcription (Supplementary Table 2). We observed that the number of genes that significantly differ in expression between the non-inoculated plants and all the mycorrhizal treatments was different depending on the cultivar identity. From this comparison, we observed in cultivar COL2215 that 252 differentially transcribed genes were common to all the treatments, but only 36 genes specific to the co-inoculation treatment (Supplementary Figure 2a). While the cultivar BRA337 displayed 226 genes common to all the treatments and 284 genes specific to the co-inoculation (Supplementary Figure 2b). The cultivar CM4574-7 displayed 264 genes common to all the treatments but only 79 genes specific to the co-inoculation (Supplementary Figure 2c).

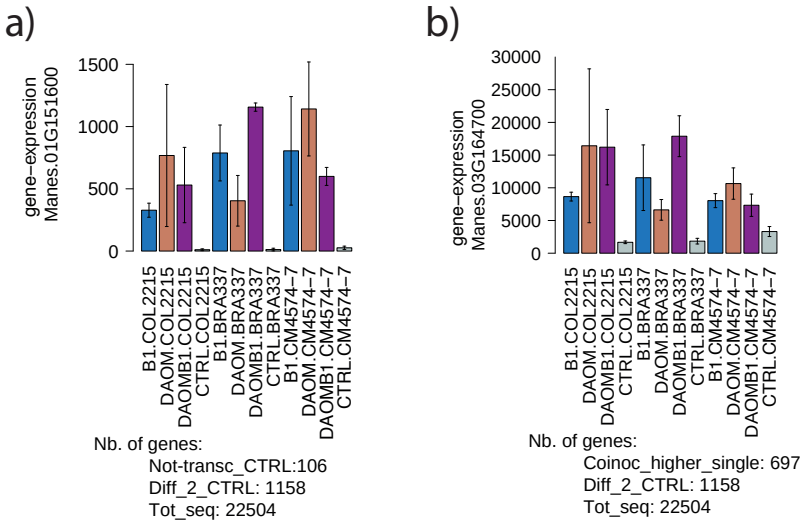


Figure 3. Gene-transcription levels of cassava genes in plants. a) case where non-inoculated plants do not display any transcription. b) Genes where the co-inoculation treatments display a higher gene-transcription compared to the single-inoculations in at least one cultivar. DAOM refers to isolate DAOM197198, B1 refers to isolate B1, DAOMB1 refers to the co-inoculation treatments and CTRL refers to the non-inoculated treatment. Below each graph there is a summary table of number of genes found. Not-transc_CTRL means genes non-transcribed in the control but in mycorrhizal treatments. Diff_2_CTRL means genes different between mycorrhizal treatments and the non-inoculated plants. Tot_seq means total number of genes sequenced in this experiment.

We found 106 cassava genes, involved in transport, signalling and protein degradation among others functions that were only activated in the mycorrhizal treatments and not in the non-inoculated samples (Figure 3a, Supplementary Table 3). We also observed that the co-inoculation treatment displayed a higher gene-transcription compared to the single-inoculations for 697 genes involved in important symbiosis functions as transport of phosphate, nitrate, protein degradation and post-translational modification and involved in signalling and in the plant biotic stress (*i.e.* pathogen-related response)

(Figure 3b, Supplementary Table 4). Remarkably, we found this effect only when the host plant was cultivar BRA337. On the other cultivars the genes displayed similar or intermediate values to the single-inoculations.

Differences in R. irregularis gene-transcription between single-inoculations and the co-inoculation

We observed that 985 *R. irregularis* genes were significantly differentially transcribed between the co-inoculated treatments compared to the single-inoculated treatments (Supplementary Table 5). From this set of genes, we observed that there were similar amount of genes differentially transcribed between the co-inoculated treatment and isolates DAOM197198 and B1 on cultivar COL2215 (Supplementary Figure 3a). However, on cultivar BRA337 we observed 3 times more genes differentially transcribed between the co-inoculated treatment and isolate DAOM197198 compared to the differentially transcribed genes between the co-inoculated treatment and isolate B1 (Supplementary Figure 3b). On cultivar CM4574-7, we saw 581 differentially transcribed genes between the co-inoculated treatment and isolate DAOM 197198 and only one differentially transcribed gene between the co-inoculated treatment and isolate B1 (Supplementary Figure 3c).

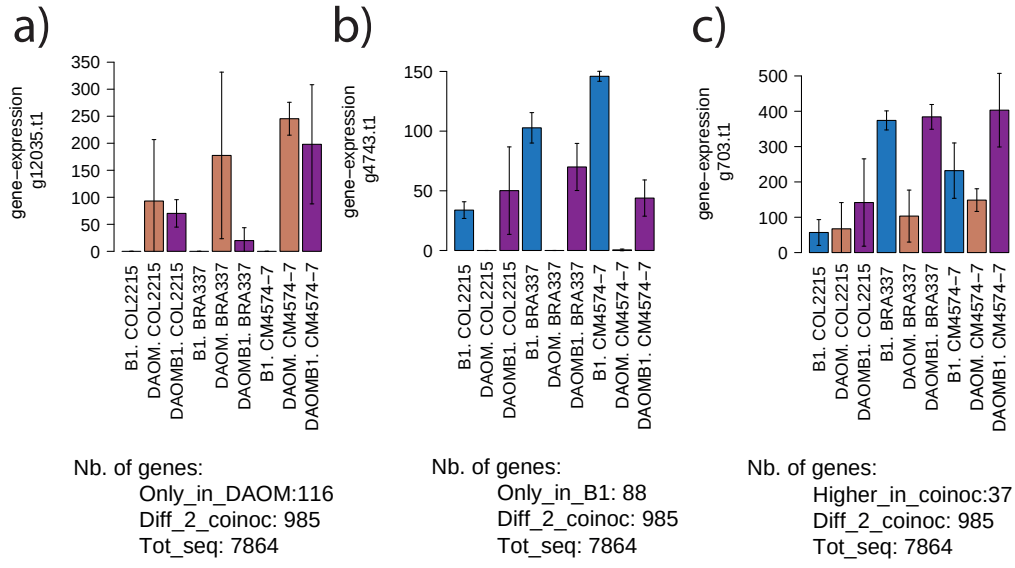


Figure 4. Gene-transcription of *R. irregularis* genes in plants inoculated with the mycorrhizal treatments. a) Genes where isolate B1 do not display any transcription but isolate DAOM197198 do. b) As opposite, genes where isolate DAOM197198 are not active but the same gene on isolate B1 is active. c) Genes where the co-inoculation treatment display a higher transcription compared to the single-inoculations in at least one cultivar. DAOM refers to isolate DAOM197198, B1 refers to isolate B1, DAOMB1 refers to the co-inoculation treatments. Below each graph there is a summary table of number of genes found. Only_in_DAOM means genes non-transcribed in B1 but in DAOM19198. Only_in_B1 means genes non-transcribed in DAOM197198 but in B1. Diff_2_coinoc means genes different between single-inoculation treatments and the co-inoculated plants. Tot_seq means total number of genes sequenced in this experiment.

We observed that each isolate had a different repertoire of transcribed genes, as several genes were only transcribed in one isolate and not in the other (For the complete list of genes: Supplementary Table 6). We found that isolate DAOM197198 has 116 transcribed genes that were not transcribed in isolate B1. As an example, isolate DAOM197198 had transcribed the gene *PMC1* involved in calcium transport, which is not transcribed in isolate B1 (Figure 4a). On the other hand, we found that isolate B1 had 88 transcribed genes that were not transcribed in isolate DAOM19798. As an example, isolate B1 has a fungal-cell wall regulatory gene (*Mkk1p*) transcribed that is not transcribed in isolate DAOM197198 (Figure 4b). For these genes the co-inoculation treatment displayed a lower gene-transcription compared to the gene transcription of the isolate that displayed the gene transcription (Figure 4c, Supplementary Table 7).

Finally, we found three genes (*MATA-HMG*, a *velvet* factor superfamily and *Skt5p;* Supplementary table 8) that displayed a high transcription level in the co-inoculation samples and almost no transcription in the single-inoculations (Figure 5). This suggests that these genes should be directly involved in the direct interaction and recognition between the two isolates.

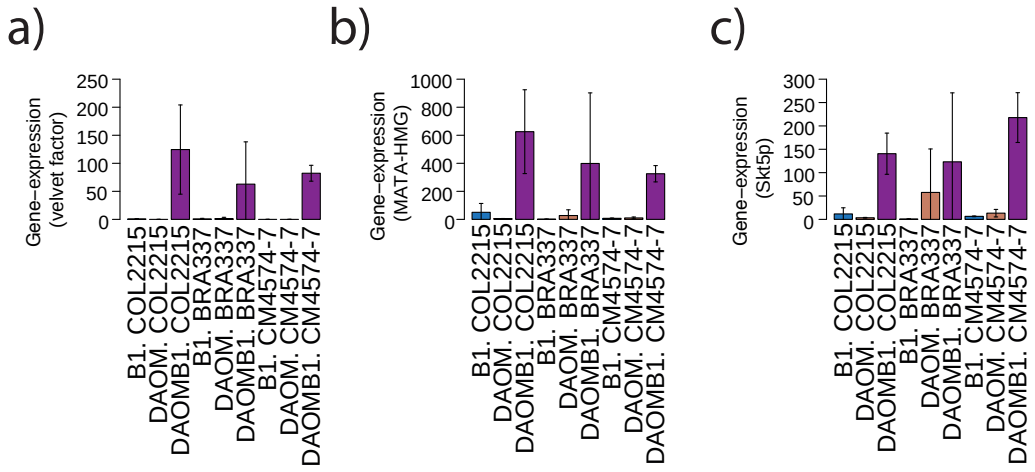


Figure 5. Genes where isolates DAOM197198 and B1 single-inoculation treatments do not display any gene transcription. For the same gene, the co-inoculation treatments display a high gene-transcription. DAOM refers to isolate DAOM197198, B1 refers to isolate B1 & DAOMB1 refers to the co-inoculation treatments.

Discussion

Our results showed that the host-plant intra-specific variability influenced the coexistence of two *R. irregularis* isolates and that the coexistence of both isolates differently affected the different host plants. We also observed a conserved cassava mycorrhizal gene-set, where in all three host cultivars,

there was no expression of these gene-set on the non-mycorrhizal plants. We observed that the two *R. irregularis* isolates possess a different repertoire of transcribed genes. Finally, we observed three *R. irregularis* genes that were probably related to the direct interaction between these two isolates.

Coexistence of two isolates differently affect different cassava cultivars.

We observed that the plant response to the coexistence of isolates DAOM197198 and B1 was variable among three cassava cultivars. We found that the co-inoculation treatment compared to the single-inoculation on the three different cassava cultivars: 1) displayed different colonization levels, 2) different effect on the dry weight and 3) different gene-transcription levels of several cassava genes.

It has been shown that different isolates of *R. irregularis* can generate a different response of the host plant (Koch *et al.* 2006), as well that different cultivars of the same species can have a different response to AMF (Hetrick *et al.* 1996). We observed that there was a higher cassava gene-transcription of the co-inoculation samples compared to the single-inoculations in BRA337, on genes involved in the plant biotic stress, signalling, protein degradation and protein modification. We observed that co-inoculation in cultivar BRA337 displayed a lower total plant dry weight compared to the single-inoculation of DAOM197198. The high gene-transcription on the previous genes suggests that in cultivar BRA337, the coexistence of two mycorrhizal fungi, could generate a bigger stress on the host plant, which could help to explain the lower dry weight of the co-inoculation treatment, compared to the single-inoculations.

Although it is expected that AMF increase plant growth, it has been reported that AMF could also negatively affect the growth of the host plant (Johnson *et al.* 1997; Klironomos 2003). Furthermore, genetic relatedness among the coexisting fungi has been reported to influence of the coexistence of two *R. irregularis* isolates on plant growth (Roger *et al.* 2013). In this study, we observed that the effect of the coexistence of two AMF isolates on the host plant growth was different among three genetically different cassava cultivars. In addition, to the importance of the fungal intra-isolate variability on the effect of the coexistence of two isolates on the host plant, we showed that the effect of two coexisting isolates was different depending on the host identity. This highlight also the importance of the host intra-species variability on the effect of coexisting isolates on plant growth.

Different gene repertoire and coexistence between isolates DAOM197198 and B1

We observed both, a shared and a different gene-repertoire, transcribed in isolates DAOM197198 and B1. The genes that were transcribed in one isolate and not in the other were involved in regulatory functions, signalling and transport. One confounding factor for this result is the effect of the reference alignment on the gene prediction. If there is a high divergence of these genes between the two isolates, the gene sequences can be divergent, generating no mapping on one of the isolates. However, by using other reference assemblies based on different *R. irregularis* isolates, we still observed this pattern (Supplementary Figure 4), meaning that the reference alignment is not generating a bias. The differential gene repertoire found on these isolates, in addition to the genes that were differently transcribed between isolates DAOM197198 and B1, should contribute to explain the phenotypic existing differences between these two isolates.

Furthermore, genes uniquely transcribed in one isolate, allowed us to test if the presence of another isolate, affects the transcription of these genes. If there is not interference between the two isolates, we could expect that the gene-transcription of active genes in only one isolate will not be different on the co-inoculation treatment. However, we observed that a big number of these genes (coding for a wide range of functions) were down regulated in the co-inoculation treatment, suggesting that the presence of another isolate is interfering the activity of these genes. We observed this pattern on genes only transcribed in DAOM197198 and also genes only transcribed in B1. We observed that the isolates when single-inoculated, displayed a very high level of roots colonization (>0.8 %). At the co-inoculation treatment, fewer roots were available for each isolate to colonize and the levels of colonization were very similar as for the single-inoculation treatments (> 0.8%). This demonstrates that the host roots were a limiting resource for the co-inoculating isolates, and that there was interference between both isolates.

Host plant intra-species variation affects the coexistence of two AMF isolates.

We observed that the different host plants differently affected the fungal gene-transcription of the single-inoculations and the co-inoculation treatments. We also observed that the fungal colonization of the co-inoculated samples compared to the single-inoculations, differed among the cassava cultivars.

Research in the AMF-plant symbiosis has been focused on how AMF affect the host plant phenotype and gene-transcription. But to date, there is very little results showing how the host plant influences

the *R. irregularis* phenotype, and no studies investigated the effect of the host plant on the fungal genes. In the nature, a mycorrhizal fungus interacts with other fungi inside the roots and associate to different host plants (Montesinos-Navarro *et al.* 2012). In this study we observed how the host plant influences differently single-inoculations and co-inoculations of two *R. irregularis* isolates. Hence, understanding how the host plant influences the fungal partner is highly important to understand the ecology and interactions of AMF.

The results on this study show that the fungal gene-transcription of co-inoculations is different from the single-inoculations of these isolates. Isolates DAOM197198 and B1 are genetically similar; they are comprised in the same genetic clade including several *R. irregularis* isolates issued from the same agricultural field (Wyss *et al.* 2016). Isolates DAOM197198 and B1 were not originated from the same place, but they have been grow in the same *in-vitro* culturing conditions for more than 20 years. Hence, maternal effects could not strongly influence the divergence between isolates DAOM197198 and B1. Remarkably, in this study, we observed that this two genetically similar isolates display different gene-transcription repertoires, and when co-inoculated display a very distinctive pattern from the single-inoculations. These results highlight also the importance of the study of the fungal intra-specific variation in the AMF-plant symbiosis in more complex conditions, than the popular one-plant, one fungus experimental setup.

Conserved mycorrhizal gene-set at the intra-species level in cassava

In this experiment we found 106 cassava genes that were only activated in the presence of AMF in all the three cassava cultivars. To date, there are very few genes, such as *PT4, STR, STR2, RAM1* and *RAM2* that could be considered as conserved mycorrhizal genes only present in plants that make the symbiosis with AMF (Harrison *et al.* 2002; Zhang *et al.* 2010; Gobbato *et al.* 2012; Wang *et al.* 2012). Furthermore, several genes could be considered as mycorrhizal because they are heavily activated on the presence of one AMF species (Hogekamp & Küster 2013). However, new methodologies as phylogenomics, allows the identification of bigger sets of mycorrhizal genes that are conserved among mycorrhizal plants (Bravo *et al.* 2016). The fact that we observed in three different cultivars the activation of 106 genes only in mycorrhizal samples suggests that these genes are conserved among the cassava species. These sets of genes represent a first pool of genes only activated in the

presence of AMF and conserved within the cassava species. Furthermore, phylogenomics analysis of these genes could be used in the future to detect if these genes are only present in mycorrhizal plants. In addition, meta-analysis of RNA-seq data of mycorrhizal and non-inoculated samples could be very valuable to understand at the scale of the plant kingdom if these genes are still only activated in mycorrhizal plants and inactivated in non-inoculated samples. In conclusion, this list of genes could be a starting point to identify a conserved across species gene-set of plant mycorrhizal specific genes.

Recognition between isolates DAOM197198 and B1

We found three genes (*MATA-HMG*, *velvet factor* and Skt5p) that were heavily transcribed in the co-inoculation samples of isolates DAOM197198 and B1 and displayed no or very low transcription in the samples where the isolates were inoculated separately. The *MATA-HMG* locus has been reported as determining the sexual compatibility between two individuals in fungal species. The test of several *MATA-HMG* loci has been done by crossing different *R. irregularis* isolates, but for all the *MATA-HMG* loci tested, the test has resulted in ambiguous patterns of transcription of the co-inoculations compared to the transcription on the isolates alone (Riley *et al.* 2014). In this study we found that the *MATA-HMG* gene; which was not tested in the previous study, displayed a high transcription on the co-inoculation of isolates DAOM197198 and B1 compared to the single-inoculations, on three different plant hosts. Furthermore, we found a *velvet factor* for which subunit A is required in spore development (Bayram & Braus 2012) and subunit B regulates the spore maturation by interacting with the subunit A on *Aspergillus nidulans* (Park *et al.* 2012). Finally, we found the *Skt5p* gene displaying the same transcription pattern. *Skt5p/Chs4p* is a regulatory gene that is involved in the septum formation in fungi (Matsuo *et al.* 2004), however this gene is also required for chitin synthesis during mating, as a knockout experiment of this gene on *Saccharomyces cerevisiae* showed a severe defect in mating (Trilla *et al.* 1997). Interestingly, it has been reported that septum formation can occur, if there is post-fusion incompatibility after hyphal fusion between two isolates (Croll *et al.* 2009). In conclusion, the over-transcription of the *velvet factor, Skt5p* and the *MATA-HMG* only in the co-inoculation samples, and the no transcription when each isolate was alone, suggests that the two isolates in the co-inoculation samples are involved in the recognition of each other, and potentially involved in mating.

Conclusions

In this study, we obtained significant evidence of a strong role of both the fungal and host plant intra-species variability on the coexistence of two *R. irregularis* isolates, and their effect on the host plant. If the majority of molecular research in AMF is made with the traditional one-plant, one fungus experimental design approach, how relevant are these results to natural conditions? This study suggests that both fungal and host genetic variability should be comprised in the study of coexistence of different AMF and more largely in the study of the AMF-plant symbiosis. Finally, we showed that two isolates of *R. irregularis* could directly recognize each other. Further co-inoculation experiments, coupled with sequencing of the region surrounding the *MATA-HMG* gene, which only displayed gene-transcription where two isolates coexisted, possibly could help to resolve the paradox of AMF as an ancient asexual scandal (Judson & Normark 1996).

Acknowledgements

We thank Keith Harshman from (Lausanne University Genomic Technologies Facility) for the sequencing. We thank Tania Wyss and Lucas Villard for advice. This research was funded by a Swiss National Science Foundation grant to IRS (Bonus of Excellence 310030B_144079). We also thank CIAT (Cali, Colombia) for providing the cassava plantlets.

General discussion and perspectives

The scope of this PhD thesis was to understand the role of the genetic variability in the fungus and in the host plant in the AMF-plant symbiosis. More precisely I aimed to understand the genomic organization of AMF; if *R. irregularis* isolates are homokaryotic or heterokaryotic. I tested if there is a genetic basis of the plant response to the AMF inoculation. I tested the influence of the fungal intra-species variability on the host plant gene-transcription, the influence of the host intra-species variability in the fungal gene-transcription and used the fungal and host intra-species variability to detect correlations in gene-transcription between the fungus and the host plant. Finally, I tested the role of the host intra-species variability on the coexistence of two *R. irregularis* isolates and the effect of the coexistence of two *R. irregularis* isolates on the host plant. These variety of experiments and conclusions, allowed us, not to focus on a single question but to understand different aspect of the AMF-plant symbiosis.

In chapter 1, we performed an experiment coupled to the multi-locus sequencing technique ddRad-seq to discriminate between the two current hypotheses of the genomic organization of AMF (presence of a single type of nucleus or multiple nuclei). We found evidence in two different *R. irregularis* isolates that single-spore siblings displayed genetic differences in terms of allele frequency differences among siblings. This confirms that there is more than one different nucleus in single-spore progenies produced for this experiment.

In chapter 2, we used the intra-species genetic variability of *R. irregularis* in order to test if there is a genetic basis of the fungal phenotype and plant response to *R. irregularis*. We found that genetically related isolates displayed similar phenotypes and similar effects on the host plant, demonstrating for the first time a link between the fungal genotype and the plant response to AMF.

In chapter 3, we used the intra-species variability of the host plant and the fungal partner in order to test the influence of each organism on the gene-transcription of the second partner using RNA-seq. We found that both the fungus intra-species variability and the host intra-species variability influenced the gene-transcription of the partner organism. In addition, this dataset allowed us to get an insight in

the symbiosis functioning at the gene-level by demonstrating an overview of gene-gene interactions within- and between- the host plant and the fungus. Furthermore, we found for the first time correlations between the transcription of fungal genes and the fungal colonization of host roots.

In chapter 4, I tested the role of the host intra-species variability on the coexistence of two R. irregularis isolates. I showed that the two *R. irregularis* isolates probably recognized each other. I showed that the two *R. irregularis* isolates possess different gene repertoires, that the coexistence of two isolates differently affect different cassava cultivars and that the different cassava cultivars affect differently the coexistence of the two isolates.

During all the different chapters of this PhD thesis, we observed a strong influence of either the fungus or the host genetic variability in different features of the AMF-plant symbiosis. We observed that the within-isolate genetic variability was also accompanied with phenotypic variability among single-spore siblings. Furthermore, we observed that the fungal intra-species variability was correlated to fungal phenotypic traits and the host plant response to the fungus. We observed that there was high gene-transcription variability between fungal isolates and host plant varieties in genes involved in the symbiosis. Finally, we observed that the coexistence of two AMF isolates has different effects on different cassava cultivars and that the host intra-species variability affected differently the coexistence of the two isolates.

All these results together showed that the interaction between AMF and the host plant is complex and highly variable even at the intra-species level. The majority of studies in the AMF-plant symbiosis use the model species (*Medicago trunculata* and *Rhizophagus irregularis*), however these studies neglect the effect of the intra- and inter- species variability by testing only one-plant one-fungus at the time. The intra- and inter- species variation has been tested in several studies to understand the phenotypic effect of AMF on the host plant (Klironomos 2003; van der Heijden *et al.* 2003; Koch *et al.* 2006), but to date, surprisingly there are very few RNA-seq or microarray studies that also addressed the intra- or inter- species variability in the AMF-plant symbiosis.

In chapter 3 and 4, we observed that several genes have a strong effect in one plant-fungus combination, but not in other plant-fungus combination. Furthermore, in chapter 4 we observed that several genes were expressed in some plant-fungus combination, but there was no expression at all in other plant-fungus combinations. The evidence shown in this PhD thesis suggests that the parsimonious approach adopted by the scientific community, of using one-plant, one-fungus combination, is underestimating the effects and variability observed in the AMF-plant symbiosis at the gene and phenotypic level.

The use of fungal and host intra-species genetic variability allowed us to draw a first draft of gene-gene interactions between the fungus and plant genes in the AMF symbiosis. This is a first major finding, because it shows first evidence of correlations in gene-transcription among fungal genes, among plant genes and between fungal and plant genes. Furthermore, this approach allowed us to find genes that could be involved in the fungal colonization of plant roots. This information allows understanding the functions and consequences of the transcription of relevant genes in the symbiosis. Despite that this approach is correlative, and no causality is demonstrated in these experiments, it is a first step to understand the 'interactome' of the AMF-plant symbiosis. This approach should be coupled with traditional knockout experiments; at least in plant genes, and coupled with 'interactome' analysis of the symbiosis, in order to reveal the genes that are responsible of the mycorrhizal effect on the host plant.

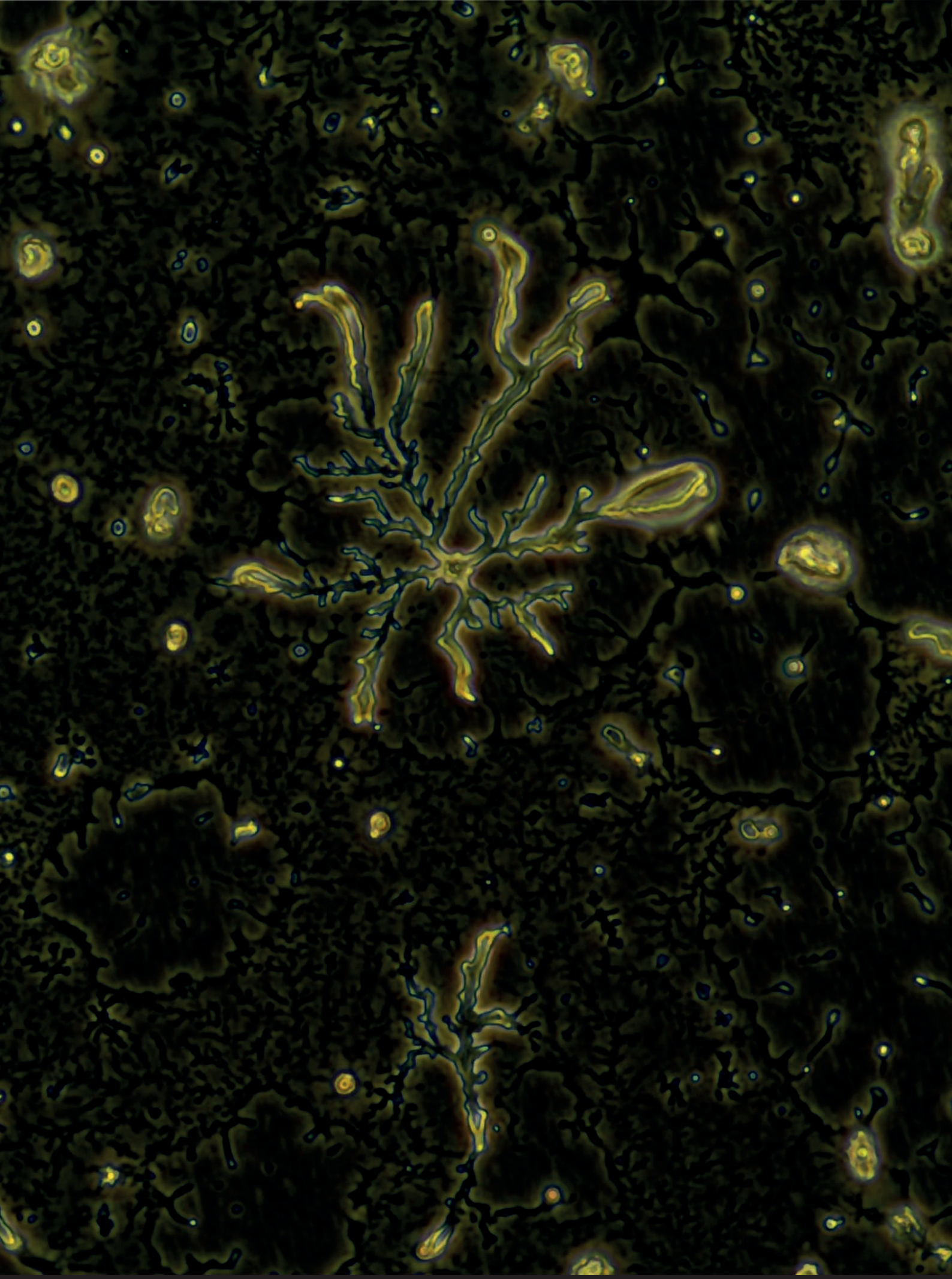
In recent years, there is increasing evidence that AMF should not be classified anymore as an 'ancient asexual case' (Halary *et al.* 2011; Riley & Corradi 2013; Tisserant *et al.* 2013; Riley *et al.* 2014; Ropars *et al.* 2016). In the chapter 4 of this PhD thesis we found evidence in three different host plants, of 3 fungal genes (*MATA-HMG*, *velvet factor superfamily* and *Skt5p*) that were only activated in presence of another fungal isolate. These genes were putatively involved in sporulation and sexual reproduction. One of these genes is a *MATA-HMG* gene, which could potentially be involved in the sex determination on this fungus (Riley *et al.* 2014). Previous studies on *R. irregularis*, have found several *MATA-HMG* in *R. irregularis* genome, but a selection of them did not show a pattern of gene-expression activation, only, in the presence of another isolate (Riley *et al.* 2014). In order to confirm that this gene is directly involved in sexual reproduction and activated only in presence of another

isolate, we started an experiment where: 1) we will sequence the region around this gene (20 kb), in order to observe if this region possess the common elements (HMG box transcription factor and Homeodomain class transcription factor) of known mating type loci in other sexual fungi (Lee *et al.* 2010). 2) We will test if the region around the *MATA-HMG* gene could be represented in two major forms as in other fungus that display sexual reproduction (Paoletti *et al.* 2005). 3) We will measure the frequency of the possible forms of this locus in the population and compare this information with other fungi that display sexual reproduction (Paoletti *et al.* 2005). 4) We performed crosses between different *R. irregularis* isolates, in order to detect by qPCR if the expression of this gene happens only when there is interaction between the two putative different forms of this locus (Riley *et al.* 2014). We expect after completing all these experiments to discard or confirm that this region is responsible of the mating type and sexual reproduction in *R. irregularis*. The consequences of these findings could allow us to understand why we observe a high genetic variability of this fungus and coexistence in same locations (Croll *et al.* 2008), and why these fungi exist for at least 352 million of years ago (Simon *et al.* 1993) and are not an extinct clade despite the accumulation of deleterious mutations across generations.

Several approaches ranging from community assembly (Davison *et al.* 2015) and population genetics (Sanders & Rodriguez 2016), attempt to understand the diversity of AMF on natural soils and host plants. The hypothesis of 'everything is everywhere' seems to be accurate when describing the distribution of AMF in different landscapes (Hazard *et al.* 2013). In addition, there are different biotic and abiotic variables that could explain the distribution of AMF communities (Chaudhary *et al.* 2008). However, phylogenetic signal approaches could be used to help to understand the distribution of AMF across environmental and biotic variables. In this thesis, we used the phylogenetic signal approach in chapter 2 to show that there is a genetic basis of plant response to AMF. This approach could be used at the inter-species level in order to test if genetically similar species displayed similar biotic or abiotic distributions. For example, the host-range of plant pathogens can be predicted by using phylogenetic signal methodologies (Gilbert & Webb 2007). Another example is the niche range of diatoms that could be predicted by using this approach (Keck *et al.* 2016b). In AMF, phylogenetic conservatism on the niche use could be tested in order to understand the distribution of different taxa across biotic and

abiotic gradients. This approach, could allow us to predict which taxa are more probable to be highly present in conditions of high PH, salinity or in some host plants.

In chapters 2 and 3, we found respectively that genetically similar isolates display similar effects on the host plant and we found that several fungal genes including major facilitator superfamily transporters and monoterpenes hydrolases are correlated to the fungal colonization of the host roots. The idea of using AMF to increase crops growth has been tested in the lab (Angelard *et al.* 2010) and in the field showing an increase in productivity of cassava (Ceballos *et al.* 2013). In this PhD thesis, we found that there is a genetic basis of the plant response to AMF. This means that despite gene-environment interactions, the response of the host plant to AMF could be predicted if AMF are inoculated to the same plant genotype in the same environment. This is a large step, on the way to develop a breeding program for using AMF to improve plant growth (Ceballos *et al.* 2013). Furthermore, because of the recent advances in gene-phenotype correlations, we could develop genetic markers that correlate to plant productivity. Approaches as the one used in chapter 3 Weighted-gene co-expression analysis (WGCNA) and Genome-wide association studies (GWAS) could be performed on the plant-AMF symbiosis to detect the variants that correlate to higher productivity of the host plant. Finally, the advances in the discovery of sexual reproduction in AMF (Halary *et al.* 2011; Ropars *et al.* 2016) suggest that in the future performing crosses between isolates and developing a 'classic' breeding program on AMF could be possible.



Contamination of *Petri* dish

Jeremy Bonvin

References

Abouheif E (1999) A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research*, **1**, 895–909.

Abràmoff MD, Hospitals I, Magalhães PJ, Abràmoff M (2004) Image Processing with ImageJ. *Biophotonics international*.

Akiyama K, Matsuzaki K, Hayashi H (2005) Plant sequiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature*, **435**, 824–827.

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

Auwera GA Van Der, Carneiro MO, Hartl C *et al.* (2014) *From FastQ data to high confidence varant calls: the Genonme Analysis Toolkit best practices pipeline*.

Baddeley A, Turner R (2005) spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal Of Statistical Software*, **12**, 1–42.

Bates D, Mächler M, Bolker B (2012) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software …*, 51.

Bayram Ö, Braus GH (2012) Coordination of secondary metabolism and development in fungi: The velvet family of regulatory proteins. *FEMS Microbiology Reviews*, **36**, 1–24.

Bécard G, Fortin JA (1988) Early events of vesicular–arbuscular mycorrhiza formation on Ri T-DNA transformed roots. *New Phytologist*, **108**, 211–218.

Bever JD, Morton J (1999) Heritable variation and mechanism of inheritance of spore shape within a population of Scutellosporea pellucida, An arbuscular mycorrhi. *American journal of botany*, **86**, 1209–1216.

Bever JD, Richardson SC, Lawrence BM, Holmes J, Watson M (2009) Preferential allocation to beneficial symbiont with spatial structure maintains mycorrhizal mutualism. *Ecology Letters*, **12**, 13–21.

Blilou I, Ocampo J a., García-Garrido JM, Garcia-Garrido JM (2000) Induction of Ltp (lipid transfer protein) and Pal (phenylalanine ammonia-lyase) gene expression in rice roots colonized by the arbuscular mycorrhizal fungus Glomus mosseae. *Journal of Experimental Botany*, **51**, 1969–1977.

Bravo A, York T, Pumplin N, Mueller LA, Harrison MJ (2016) Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nature Plants*, **2**, 1–6.

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, **34**, 525–527.

Bredeson J V, Lyons JB, Prochnik SE *et al.* (2016) Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nature biotechnology*, **34**, 562–570.

Breuillin F, Schramm J, Hajirezaei M *et al.* (2010) Phosphate systemically inhibits development of arbuscular mycorrhiza in Petunia hybrida and represses genes involved in mycorrhizal functioning. *Plant Journal*, **64**, 1002–1017.

Buee M, Rossignol M, Jauneau a, Ranjeva R, Bécard G (2000) The pre-symbiotic growth of arbuscular mycorrhizal fungi is induced by a branching factor partially purified from plant root exudates. *Molecular plant-microbe interactions : MPMI*, **13**, 693–698.

Burleigh SH, Kristensen BK, Bechmann IE (2003) A plasma membrane zinc transporter from Medicago truncatula is up-regulated in roots by Zn fertilization, yet down-regulated by arbuscular mycorrhizal colonization. *Plant Molecular Biology*, **52**, 1077–1088.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda, Md.)*, **1**, 171–82.

Ceballos I, Ruiz M, Fernández C *et al.* (2013) The In Vitro Mass-Produced Model Mycorrhizal Fungus,

Rhizophagus irregularis, Significantly Increases Yields of the Globally Important Food Security Crop Cassava (M Rillig, Ed,). *PLoS ONE*, **8**, e70633.

Chaudhary VB, Lau MK, Johnson NC (2008) Macroecology of microbes - Biogeography of the glomeromycota. *Mycorrhiza: State of the Art, Genetics and Molecular Biology, Eco-Function, Biotechnology, Eco-Physiology, Structure and Systematics (Third Edition)*, 529–563.

Corradi N, Croll D, Colard A *et al.* (2007) Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. *Applied and environmental microbiology*, **73**, 366–9.

Corratgé-Faillie C, Jabnoune M, Zimmermann S *et al.* (2010) Potassium and sodium transport in non-animal cells: The Trk/Ktr/HKT transporter family. *Cellular and Molecular Life Sciences*, **67**, 2511–2532.

Croll D (2016) Sex or no sex: Evolutionary adaptation occurs regardless. *BioEssays*, **38**.

Croll D, Giovannetti M, Koch AM *et al.* (2009) Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus Glomus intraradices. *The New phytologist*, **181**, 924–37.

Croll D, Wille L, Gamper H a *et al.* (2008) Genetic diversity and host plant preferences revealed by simple sequence repeat and mitochondrial markers in a population of the arbuscular mycorrhizal fungus Glomus intraradices. *The New phytologist*, **178**, 672–87.

Davison J, Moora M, Öpik M *et al.* (2015) Global assessment of arbuscular mycorrhizal fungus diversity reveals very lowendemism. *Science*, **127**, 970–973.

Delaux PM, Séjalon-Delmas N, Bécard G, Ané JM (2013) Evolution of the plant-microbe symbiotic "toolkit." *Trends in Plant Science*, **18**, 298–304.

Dickman MB, Ha YS, Yang Z, Adams B, Huang C (2003) A protein kinase from Colletotrichum trifolii is induced by plant cutin and is required for appressorium formation. *Molecular plant-microbe interactions : MPMI*, **16**, 411–421.

Dobin A, Davis CA, Schlesinger F *et al.* (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Ehinger MO, Croll D, Koch AM, Sanders IR (2012) Significant genetic and phenotypic changes arising from clonal growth of a single spore of an arbuscular mycorrhizal fungus over multiple generations. *New Phytologist*, **196**, 853–861.

Ehinger M, Koch AM, Sanders IR (2009) Changes in arbuscular mycorrhizal fungal phenotypes and genotypes in response to plant species identity and phosphorus concentration. *New Phytologist*, **184**, 412–423.

El-Sharkawy MA, Cock JH (1987) Response of cassava to water stress. *Plant and Soil*, **100**, 345–360.

Ellstrand NC., Roose ML (1987) Patterns of Genotypic Diversity in Clonal Plant Species. *American journal of botany*, **74**, 123–131.

Felsenstein J, Yokoyama S (1976) The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics*, **83**, 845–859.

Galili T (2015) dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, **31**, 3718–3720.

García-Garrido JM, Ocampo J a (2002) Regulation of the plant defence response in arbuscular mycorrhizal symbiosis. *Journal of experimental botany*, **53**, 1377–1386.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*, 9.

Genre A, Chabaud M, Balzergue C *et al.* (2013) Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca2+ spiking in Medicago truncatula roots and their production is enhanced by strigolactone. *New Phytologist*, **198**, 190–202.

Gilbert GS, Briggs HM, Magarey R (2015) The impact of plant enemies shows a phylogenetic signal. *PLoS ONE*, **10**, 1–11.

Gilbert GS, Parker IM (2016) *The Evolutionary Ecology of Plant Disease: A Phylogenetic Perspective*.

Gilbert GS, Webb CO (2007) Phylogenetic signal in plant pathogen-host range. *Proceedings of the National Academy of Sciences*, **104**, 4979–4983.

Giovannetti M, Azzolini D, Silvia A (1999) Anastomosis Formation and Nuclear and Protoplasmic Exchange in Arbuscular Mycorrhizal Fungi Anastomosis Formation and Nuclear and Protoplasmic Exchange in Arbuscular Mycorrhizal Fungi.

Giovannetti M, Mosse B (1980) An evaluation of techniques for measuring vesicular arbuscular mycorrhizal infection in roots. *New Phytologist*, **84**, 489–500.

Giovannetti M, Sbrana C, Avio L, Citernsei AS, Logi C (1993) Differential hyphal morphogenesis in arbuscular mycorrhizal fungi during pre-infection stages. *New Phytologist*, **125**, 587–593.

Gobbato E, Marsh JF, Vernié T *et al.* (2012) A GRAS-type transcription factor with a specific function in mycorrhizal signaling. *Current Biology*, **22**, 2236–2241.

González-Guerrero M, Escudero V, Saéz Á, Tejada-Jiménez M (2016) Transition Metal Transport in Plants and Associated Endosymbionts: Arbuscular Mycorrhizal Fungi and Rhizobia. *Frontiers in Plant Science*, **7**, 1–21.

Götz S, Garcia-Gomez JM, Terol J *et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, **36**, 3420–3435.

Govindarajulu M, Pfeffer PE, Jin H *et al.* (2005) Nitrogen transfer in the arbuscular mycorrhizal symbiosis. *Nature*, **435**, 819–823.

Grace EJ, Cotsaftis O, Tester M, Smith FA, Smith SE (2009) Arbuscular mycorrhizal inhibition of growth in barley cannot be attributed to extent of colonization, fungal phosphorus uptake or effects on expression of plant phosphate transporter genes. *New Phytologist*, **181**, 938–949.

Guether M, Neuhäuser B, Balestrini R *et al.* (2009) A mycorrhizal-specific ammonium transporter from Lotus japonicus acquires nitrogen released by arbuscular mycorrhizal fungi. *Plant physiology*, **150**, 73–83.

Gustafson DJ, Casper BB (2006) Differential host plant performance as a function of soil arbuscular mycorrhizal fungal communities: Experimentally manipulating co-occurring Glomus species. *Plant Ecology*, **183**, 257–263.

Gutjahr C, Banba M, Croset V *et al.* (2008) Arbuscular Mycorrhiza-Specific Signaling in Rice Transcends the Common Symbiosis Signaling Pathway. *the Plant Cell Online*, **20**, 2989–3005.

Gutjahr C, Parniske M (2013) Cell and developmental biology of arbuscular mycorrhiza symbiosis. *Annual Review of Cell and Developmental Biology*, **29**, 593–617.

Halary S, Malik SB, Lildhar L *et al.* (2011a) Conserved meiotic machinery in Glomus spp., a putatively ancient asexual fungal lineage. *Genome Biology and Evolution*, **3**, 950–958.

Handa Y, Nishide H, Takeda N *et al.* (2015) RNA-seq Transcriptional Profiling of an Arbuscular Mycorrhiza Provides Insights into Regulated and Coordinated Gene Expression in Lotus japonicus and Rhizophagus irregularis. *Plant and Cell Physiology*, **56**, 1490–1511.

Harrison MJ (1997) The arbuscular mycorrhizal symbiosis: An underground association. *Trends in Plant Science*, **2**, 54–60.

Harrison MJ, Dewbre GR, Liu JY (2002) A phosphate transporter from Medicago truncatula involved in the acquisiton of phosphate released by arbuscular mycorrhizal fungi. *Plant Cell*, **14**, 2413–2429.

Hazard C, Gosling P, van der Gast CJ *et al.* (2013) The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. *The ISME journal*, **7**, 498–508.

Hebert PD. N., Beaton MJ., Schwartz SS, Stanton DJ. (1989) Polyphyletic Origins of Asexuality in Daphnia pulex . I . Breeding-System Variation and Levels of Clonal Diversity. *Evolution*, **43**, 1004–1015.

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.

van der Heijden MGA, Wiemken A, Sanders IR (2003) Different arbuscular mycorrhizal fungi alter coexistence and resource distribution between co-occurring plant. *New Phytologist*, **157**, 569–578.

van der Heijden MGA, Klironomos J, Ursic M *et al.* (1998) Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature*, **74**, 69–72.

Helber N, Wippel K, Sauer N *et al.* (2011) A Versatile Monosaccharide Transporter That Operates in the Arbuscular Mycorrhizal Fungus Glomus sp Is Crucial for the Symbiotic Relationship with Plants. *The Plant Cell*, **23**, 3812–3823.

Helgason T, Merryweather JW, Denison J *et al.* (2002) Selectivity and functional diversity in arbuscular mycorrhizas of co-occuring fungi and plants from a temperate deciduous woodland. *Journal of Ecology*, **90**, 371–384.

Hetrick BAD, Wilson GWT, Todd TC (1996) Mycorrhizal response in wheat cultivars: relationship to phosphorus. *Canadian Journal of Botany*, **74**, 19–25.

Heupel S, Roser B, Kuhn H *et al.* (2010) Erl1, a novel era-like GTPase from Magnaporthe oryzae, is required for full root virulence and is conserved in the mutualistic symbiont Glomus intraradices. *Molecular plant-microbe interactions : MPMI*, **23**, 67–81.

Hijri M, Sanders I (2005) Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. *Nature*, **433**, 160–163.

Hildebrandt U, Schmelzer E, Bothe H (2002) Expression of nitrate transporter genes in tomato colonized by an arbuscular mycorrhizal fungus. *Physiologia plantarum*, **115**, 125–136.

Hogekamp C, Küster H (2013) A roadmap of cell-type specific gene expression during sequential stages of the arbuscular mycorrhiza symbiosis. *BMC genomics*, **14**, 306.

Howeler RH (2013) *Save and Grow: Cassava. A Guide to Sustainable Production Intensification*.

Jakobsen I, Abbott LK, Robson a. D (1992) External hyphae of vesicular-arbuscular mycorrhizal fungi associated with Trifolium subterraneum L. 1. Spread of hyphae and phosphorus inflow into roots. *New Phytologist*, **120**, 371–380.

Janoušková M, Seddas P, Mrnka L *et al.* (2009) Development and activity of Glomus intraradices as affected by co-existence with Glomus claroideum in one root system. *Mycorrhiza*, **19**, 393–402.

Jansa J, Mozafar A, Banke S, Donald BAMC, Frossard E (2002) Intra- and intersporal diversity of ITS rDNA sequences in Glomus intraradices assessed by cloning and sequencing , and by SSCP analysis. *Mycological Research*, **106**, 670–681.

Javot H, Penmetsa RV, Terzaghi N, Cook DR, Harrison MJ (2007) A Medicago truncatula phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 1720–1725.

Johnson NC, Graham JH, Smith FA (1997) Functioning of mycorrhizal associations along the mutualism-parasitism continuum. *New Phytologist*, **135**, 575–586.

Judson OP, Normark BB (1996) Ancient asexual scandals. *Trends in ecology & evolution*, **11**, 41–46.

Jung SC, Martinez-Medina A, Lopez-Raez JA, Pozo MJ (2012) Mycorrhiza-Induced Resistance and Priming of Plant Defenses. *Journal of Chemical Ecology*, **38**, 651–664.

Kaldorf M, Kuhn AJ, Schröder WH, Hildebrandt U, Bothe H (1999) Selective Element Deposits in Maize Colonized by a Heavy Metal Tolerance Conferring Arbuscular Mycorrhizal Fungus. *Journal of Plant Physiology*, **154**, 718–728.

Kamoun S (2006) A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual review of phytopathology*, **44**, 41–60.

Keck F, Rimet F, Bouchez A, Franc A (2016a) phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution*, **9**, 2774-2780.

Keck F, Rimet F, Franc A, Bouchez A (2016b) Phylogenetic signal in Diatom ecology: Perspectives for aquatic ecosystems biomonitoring. *Ecological Applications*, **26**, 861–872.

Keller O, Odronitz F, Stanke M, Kollmar M, Waack S (2008) Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC bioinformatics*, **9**, 278.

Kent WJ (2002) BLAT — The BLAST -Like Alignment Tool. *Genome research*, **12**, 656–664.

Kiers ET, Duhamel M, Beesetty Y *et al.* (2011) Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science*, **333**, 880–2.

Klironomos J (2003) Variation in plant response to native and exotic arbusculat mycorrhizal fungi. *Ecology*, **84**, 2292–2301.

Kloppholz S, Kuhn H, Requena N (2011) A secreted fungal effector of glomus intraradices promotes symbiotic biotrophy. *Current Biology*, **21**, 1204–1209.

Koch AM, Croll D, Sanders IR (2006) Genetic variability in a population of arbuscular mycorrhizal fungi causes variation in plant growth. *Ecology letters*, **9**, 103–10.

Koch AM, Sanders IR, Kuhn G, Fontanillas P, Fumagalli L (2004) High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. *PNAS*, **101**, 2369–2374.

Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, **27**, 3435–3436.

Koide R (1985) The Nature of Growth Depression in Sunflower Caused by Vesicular-Arbuscular Mycorrhizal Infection. *New Phytology*, **99**, 449–462.

Kuhn G, Hijri M, Sanders IR (2001) Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature*, **414**, 745–8.

Kuhn H, Küster H, Requena N (2010) Membrane steroid-binding protein 1 induced by a diffusible fungal signal is critical for mycorrhization in Medicago truncatula. *New Phytologist*, **185**, 716–733.

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, **9**, 559.

Lee SC, Ni M, Li W, Shertz C, Heitman J (2010) The evolution of sex: a perspective from the fungal kingdom. *Microbiology and molecular biology reviews : MMBR*, **74**, 298–340.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Lin K, Limpens E, Zhang Z *et al.* (2014) Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. *PLoS genetics*, **10**, e1004078.

Manck-Götzenberger J, Requena N (2016) Arbuscular mycorrhiza Symbiosis Induces a Major Transcriptional Reprogramming of the Potato SWEET Sugar Transporter Family. *Frontiers in Plant Science*, **7**, 1–14.

Marleau J, Dalpé Y, St-Arnaud M, Hijri M (2011) Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. *BMC evolutionary biology*, **11**, 51.

Matsuo Y, Matsuura Y, Tanaka K, Matsuda H, Kawamukai M (2004) Chr4, a Schizosaccharomyces pombe homologue of the Saccharomyces cerevisiae Chs4p/Skt5p protein, is related to septum formation and is required for the proper localization of Chs2. *Yeast*, **21**, 1005–1019.

Montesinos-Navarro a, Segarra-Moragues JG, Valiente-Banuet a, Verdú M (2012) The network structure of plant-arbuscular mycorrhizal fungi. *The New phytologist*, **194**, 536–47.

Moran PAP (1948) The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, **10**, 243–251.

Mullen RB, Schmidt SK (1993) Mycorrhizal infection, phosphorus uptake, and phenology in Ranunculus adoneus: implications for the functioning of mycorrhizae in alpine systems. *Oecologia*, **94**, 229–234.

Münkemüller T, Lavergne S, Bzeznik B *et al.* (2012) How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, **3**, 743–756.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.

Op den Camp R, Streng A, Mita S De *et al.* (2011) LysM-Type Mycorrhizal Receptor Recruited for Rhizobium Symbiosis in Nonlegume Parasponia. *Science*, **331**, 909–912.

Ouziad F, Hildebrandt U, Schmelzer E, Bothe H (2005) Differential gene expressions in arbuscular mycorrhizal-colonized tomato grown under heavy metal stress. *Journal of Plant Physiology*, **162**, 634–649.

Oyetunji OJ, Osonubi O (2007) Assessment of influence of alley cropping system and arbuscular mycorrhizal (AM) fungi on cassava productivity in derived savanna zone of Nigeria. *World Journal of Agricultural Sciences*, **3**, 489–495.

Paoletti M, Rydholm C, Schwier EU *et al.* (2005) Evidence for sexuality in the opportunistic fungal pathogen Aspergillus fumigatus. *Current Biology*, **15**, 1242–1248.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Parchman TL, Gompert Z, Mudge J *et al.* (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

Park HS, Ni M, Jeong KC, Kim YH, Yu JH (2012) The Role, Interaction and Regulation of the Velvet Regulator VelB in Aspergillus nidulans. *PLoS ONE*, **7**.

Parniske M (2008) Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature reviews. Microbiology*, **6**, 763–75.

Pawlowska TE, Taylor JW (2004) Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature*, **427**, 733–737.

Pearson JN, Abbott LK, Jasper DA (1993) Mediation of competition between two colonizing VA mycorrhizal fungi by the host plant. *New Phytologist*, **123**, 93–98.

Powell JR, Parrent JL, Hart MM *et al.* (2009) Phylogenetic trait conservatism and the evolution of functional trade-offs in arbuscular mycorrhizal fungi. *Proceedings. Biological sciences / The Royal Society*, **276**, 4237–45.

R CT (2016) R: A Language and Environment for Statistical Computing. , https://www.R-project.org/.

Redecker D (2000) Glomalean Fungi from the Ordovician. *Science*, **289**, 1920–1921.

Requena N, Mann P, Hampp R, Franken P (2002) Early developmentally regulated genes in the arbuscular mycorrhizal fungus Glomus mosseae: identification of GmGIN1, a novel gene with homology to the C-terminus ofmetazoan hedgehog proteins. *Plant and soil*, 129–139.

Riley R, Charron P, Idnurm A *et al.* (2014) Extreme diversification of the mating type-high-mobility group (MATA-HMG) gene family in a plant-associated arbuscular mycorrhizal fungus. *New Phytologist*, **201**, 254–268.

Riley R, Corradi N (2013) Searching for clues of sexual reproduction in the genomes of arbuscular mycorrhizal fungi. *Fungal Ecology*, **6**, 44–49.

Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Roger A, Colard A, Angelard C, Sanders IR (2013) Relatedness among arbuscular mycorrhizal fungi drives plant growth and intraspecific fungal coexistence. *The ISME journal*, **7**, 2137–46.

Ropars J, Toro KS, Noel J *et al.* (2016) Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nature Microbiology*, 16033.

Sanders IR, Croll D (2010) Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. *Annual review of genetics*, **44**, 271–92.

Sanders IR, Rodriguez A (2016) Aligning molecular studies of mycorrhizal fungal diversity with ecologically important levels of diversity in ecosystems. *The ISME journal*, 1–7.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England)*, 1–3.

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.

Sedzielewska K, Fuchs J, Temsch EM *et al.* (2011) Estimation of the Glomus intraradices nuclear DNA content. *New Phytologist*, **192**, 794–797.

Sędzielewska Toro K, Brachmann A (2016) The effector candidate repertoire of the arbuscular mycorrhizal fungus Rhizophagus clarus (supp. data). *BMC Genomics*, **53**, 1689–1699.

Sieverding E HR (1985) Influence of species of VA mycorrhizal fungi on cassava yield response to phosphorus fertilization. *Plant Soil*, **88**, 213–221.

Simon L, Bousquet J, Lévesque RC, Lalonde M (1993) Origin and diversifiaction of endomycorrhizal fungi and coincidence with vascular land plants. *Nature*, **363**, 67–69.

Smith GK (2005) limma: Linear Models for Microarray Data. *Bioinformatics and Computational Biology*

Solutions Using R and Bioconductor, 397–420.

Smith A, Hubley R, Green P (2010) RepeatMasker open-3.0.

Smith SE, Read DJ (2008) *Mycorrhizal Symbiosis*. Elsevier Ltd, London.

Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, **4**, 1521.

St-Arnaud M, Hamel C, Vimard B, Caron M, Fortin J a. (1996) Enhanced hyphal growth and spore production of the arbuscular mycorrhizal fungus Glomus intraradices in an in vitro system in the absence of host roots. *Mycological Research*, **100**, 328–332.

Tang Y-C, Amon A (2013) Gene copy number alterations: A cost-benefit analysis. *Cell*, **152**, 394–405.

Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE (2015) Clonal reproduction in fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 8901–8.

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.

Thimm O, Bläsing O, Gibon Y *et al.* (2004) MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal*, **37**, 914–939.

Tisserant E, Kohler A, Balestrini R *et al.* (2012) The transcriptome of the arbuscular mycorrhizal fungus Glomus intraradices (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. , 755–769.

Tisserant E, Malbreil M, Kuo A *et al.* (2013) Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences*, **111**, 562–563.

Trilla JA, Cos T, Duran A, Roncero C (1997) Characterization of CHS4 (CAL2), a gene of Saccharomyces cerevisiae involved in chitin biosynthesis and allelic to SKT5 and CSD4. *Yeast*, **13**, 795–807.

Veiga RSL, Jansa J, Frossard E, van der Heijden MGA (2011) Can arbuscular mycorrhizal fungi reduce the growth of agricultural weeds? *PLoS ONE*, **6**, 1–10.

Violi H a., Treseder KK, Menge J a., Wright SF, Lovatt CJ (2007) Density dependence and interspecific interactions between arbuscular mycorrhizal fungi mediated plant growth, glomalin production, and sporulation. *Canadian Journal of Botany*, **85**, 63–75.

Volpin H, Phillips D a., Okon Y, Kapulnik Y (1995) Suppression of an Isoflavonoid Phytoalexin Defense Response in Mycorrhizal Alfalfa Roots. *Plant physiology*, **108**, 1449–1454.

Vorburger C (2006) Temporal dynamics of genotypic diversity reveal strong clonal selection in the aphid Myzus persicae. *Journal of Evolutionary Biology*, **19**, 97–107.

Wang E, Schornack S, Marsh JF *et al.* (2012) A common signaling process that promotes mycorrhizal and oomycete colonization of plants. *Current Biology*, **22**, 2242–2246.

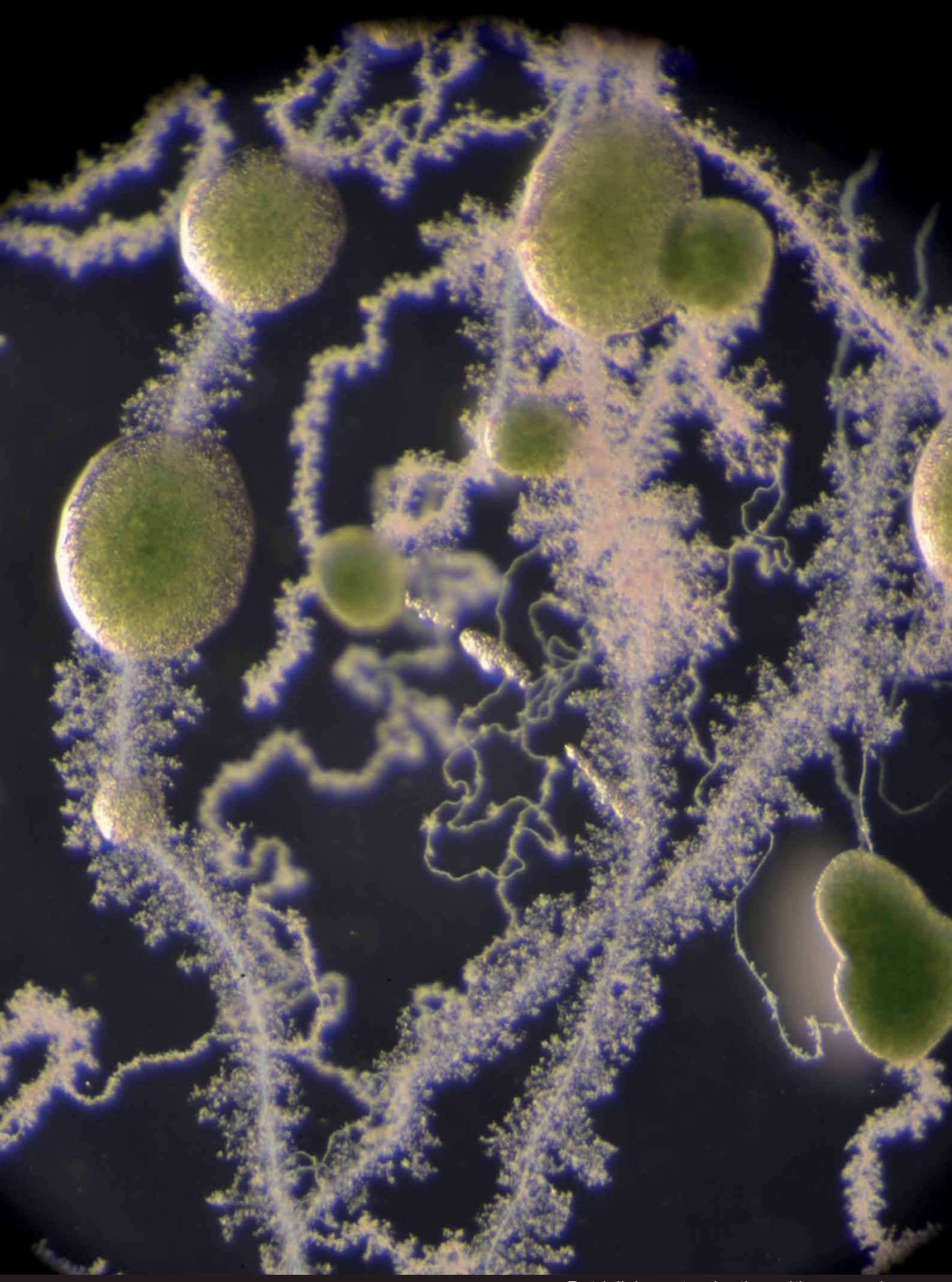
Wright S (1951) THE GENETICAL STRUCTURE OF POPULATIONS. *Ann. Eugenics,*, **15**, 323–354.

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 MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology*, **11**, R14.

Zaller JG, Frank T, Drapela T (2011) Soil sand content can alter effects of different taxa of mycorrhizal fungi on plant biomass production of grassland species. *European Journal of Soil Biology*, **47**, 175–181.

Zhang Q, Blaylock L a, Harrison MJ (2010) Two Medicago truncatula half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. *The Plant cell*, **22**, 1483–1497.



Petri dish contamination with green algae

Jeremy Bonvin

Annexe I. Quantitative genetic and phenotypic differences generated among clonal siblings of the symbiotic fungus Rhizophagus irregularis

SUPPLEMENTARY TABLE 1

Sequence information of each single-spore culture (ssc), and its replicates, of isolates B4 and DAOM 197198 that were mapped to the N6 and Rhiir2 reference assemblies. Number of mapped reads, mean coverage of sites with at least 10x, total sites sequenced, number of sites that displayed more than one allele (poly-allelic sites), and density of poly-allelic sites (shown as the number of poly-allelic sites per kilobase) are shown. Total number of base pairs sequenced, poly-allelic sites and number of poly-allelic sites / kb were estimated by averaging the pairwise 'sync' files (Popoolation2) output between each pair of cultures.

File	Isolate	Single-spore culture	Replicate	Reference assembly	Mapped reads	Mean coverage (>=10)	Estimation of total sites sequenced	Estimated number of poly-allelic sites	Number of poly-allelic sites/kb
B4_ssc1_a_Nu6.bam	B4	B4_ssc1	a	N6	2394106	53.65	1449101.8	5179.6	3.57
B4_ssc1_b_Nu6.bam	B4	B4_ssc1	b	N6	2675742	59.26	1449101.8	5105.6	3.52
B4_ssc1_c_Nu6.bam	B4	B4_ssc1	c	N6	1680447	44.8	1449101.8	5090.6	3.51
B4_ssc2_a_Nu6.bam	B4	B4_ssc2	a	N6	2481661	49.67	1428211.25	5262.25	3.68
B4_ssc2_b_Nu6.bam	B4	B4_ssc2	b	N6	1958151	48.6	1428211.25	5200.5	3.64
B4_ssc2_c_Nu6.bam	B4	B4_ssc2	c	N6	3027365	30.45	1428211.25	5238	3.67
B4_ssc3_a_Nu6.bam	B4	B4_ssc3	a	N6	1827459	55.62	1395961	4832.8	3.46
B4_ssc3_b_Nu6.bam	B4	B4_ssc3	b	N6	1565975	49.71	1395961	4880.6	3.5
B4_ssc3_c_Nu6.bam	B4	B4_ssc3	c	N6	1764742	39.71	1395961	4847.4	3.47
B4_ssc4_a_Nu6.bam	B4	B4_ssc4	a	N6	2303795	44.31	1412051.2	4957.6	3.51
B4_ssc4_b_Nu6.bam	B4	B4_ssc4	b	N6	1860074	36.09	1412051.2	5019	3.55
B4_ssc4_c_Nu6.bam	B4	B4_ssc4	c	N6	1649530	75.21	1412051.2	4966.8	3.52
B4_ssc5_a_Nu6.bam	B4	B4_ssc5	a	N6	2052012	40.98	1311523.2	4853.8	3.7
B4_ssc5_b_Nu6.bam	B4	B4_ssc5	b	N6	1286500	60.62	1311523.2	4810.8	3.67
B4_ssc5_c_Nu6.bam	B4	B4_ssc5	c	N6	1175479	43.72	1311523.2	4822.2	3.68
B4_ssc6_a_Nu6.bam	B4	B4_ssc6	a	N6	1213704	39.44	1338720.4	4958.6	3.7
B4_ssc6_b_Nu6.bam	B4	B4_ssc6	b	N6	3705652	34.39	1338720.4	5004.4	3.74
B4_ssc6_c_Nu6.bam	B4	B4_ssc6	c	N6	2059404	47.51	1338720.4	5032.4	3.76
B4_ssc1_a_Rhiir2.bam	B4	B4_ssc1	a	Rhiir2	2143455	49.2894	1797213	3276	1.82
B4_ssc1_b_Rhiir2.bam	B4	B4_ssc1	b	Rhiir2	2407434	51.5529	1797213	3276	1.82
B4_ssc1_c_Rhiir2.bam	B4	B4_ssc1	c	Rhiir2	1508656	37.6833	1797213	3276	1.82
B4_ssc2_a_Rhiir2.bam	B4	B4_ssc2	a	Rhiir2	2201499	53.9331	1798249.2	3336.8	1.86
B4_ssc2_b_Rhiir2.bam	B4	B4_ssc2	b	Rhiir2	1733443	45.262	1798249.2	3336.8	1.86
B4_ssc2_c_Rhiir2.bam	B4	B4_ssc2	c	Rhiir2	2690287	55.41	1798249.2	3336.8	1.86
B4_ssc3_a_Rhiir2.bam	B4	B4_ssc3	a	Rhiir2	1625888	40.9215	1736157.4	3124	1.8
B4_ssc3_b_Rhiir2.bam	B4	B4_ssc3	b	Rhiir2	1402430	36.5178	1736157.4	3124	1.8
B4_ssc3_c_Rhiir2.bam	B4	B4_ssc3	c	Rhiir2	1578950	40.2038	1736157.4	3124	1.8

B4_ssc4_a_Rhiir2.bam	B4	B4_ssc4	a	Rhiir2	1654837	40.3251	1770592.8	3177.8	1.79
B4_ssc4_b_Rhiir2.bam	B4	B4_ssc4	b	Rhiir2	2077101	45.86	1770592.8	3177.8	1.79
B4_ssc4_c_Rhiir2.bam	B4	B4_ssc4	c	Rhiir2	1467356	35.9259	1770592.8	3177.8	1.79
B4_ssc5_a_Rhiir2.bam	B4	B4_ssc5	a	Rhiir2	1039390	31.1254	1563714.8	2647.6	1.69
B4_ssc5_b_Rhiir2.bam	B4	B4_ssc5	b	Rhiir2	872971	28.0399	1563714.8	2647.6	1.69
B4_ssc5_c_Rhiir2.bam	B4	B4_ssc5	c	Rhiir2	981841	28.7454	1563714.8	2647.6	1.69
B4_ssc6_a_Rhiir2.bam	B4	B4_ssc6	a	Rhiir2	1074697	27.435	1759975.6	3144.6	1.79
B4_ssc6_b_Rhiir2.bam	B4	B4_ssc6	b	Rhiir2	3276397	68.4134	1759975.6	3144.6	1.79
B4_ssc6_c_Rhiir2.bam	B4	B4_ssc6	c	Rhiir2	1847772	43.6726	1759975.6	3144.6	1.79
CAN_ssc1_a_Nu6.bam	DAOM 197198	CAN_ssc1	a	N6	4553183	100.12	2636541	4398	1.67
CAN_ssc1_b_Nu6.bam	DAOM 197198	CAN_ssc1	b	N6	6217163	131.89	2636541	4398	1.67
CAN_ssc1_c_Nu6.bam	DAOM 197198	CAN_ssc1	c	N6	4760100	107.58	2636541	4398	1.67
CAN_ssc2_a_Nu6.bam	DAOM 197198	CAN_ssc2	a	N6	5280152	111.49	2625215	4397.5	1.68
CAN_ssc2_b_Nu6.bam	DAOM 197198	CAN_ssc2	b	N6	6062349	118.06	2625215	4397.5	1.68
CAN_ssc2_c_Nu6.bam	DAOM 197198	CAN_ssc2	c	N6	4694589	71.12	2625215	4397.5	1.68
CAN_ssc3_a_Nu6.bam	DAOM 197198	CAN_ssc3	a	N6	6310284	136.34	2657423	4348.5	1.64
CAN_ssc3_b_Nu6.bam	DAOM 197198	CAN_ssc3	b	N6	3888186	91.51	2657423	4348.5	1.64
CAN_ssc3_c_Nu6.bam	DAOM 197198	CAN_ssc3	c	N6	6371210	124.87	2657423	4348.5	1.64
CAN_ssc1_a_Rhiir2.bam	DAOM 197198	CAN_ssc1	a	Rhiir2	4321332	93.5283	2800064.5	1787.5	0.64
CAN_ssc1_b_Rhiir2.bam	DAOM 197198	CAN_ssc1	b	Rhiir2	5887060	123.064	2800064.5	1787.5	0.64
CAN_ssc1_c_Rhiir2.bam	DAOM 197198	CAN_ssc1	c	Rhiir2	4518289	100.126	2800064.5	1787.5	0.64
CAN_ssc2_a_Rhiir2.bam	DAOM 197198	CAN_ssc2	a	Rhiir2	5022784	105.327	2785369.5	1788	0.64
CAN_ssc2_b_Rhiir2.bam	DAOM 197198	CAN_ssc2	b	Rhiir2	5740897	110.494	2785369.5	1788	0.64
CAN_ssc2_c_Rhiir2.bam	DAOM 197198	CAN_ssc2	c	Rhiir2	4472800	66.1964	2785369.5	1788	0.64
CAN_ssc3_a_Rhiir2.bam	DAOM 197198	CAN_ssc3	a	Rhiir2	5983881	127.311	2795869	1774.5	0.63
CAN_ssc3_b_Rhiir2.bam	DAOM 197198	CAN_ssc3	b	Rhiir2	3695089	84.8371	2795869	1774.5	0.63
CAN_ssc3_c_Rhiir2.bam	DAOM 197198	CAN_ssc3	c	Rhiir2	6071808	117.166	2795869	1774.5	0.63

SUPPLEMENTARY TABLE 2

Reproducibility of replicates of each single-spore culture. A site is considered accurate if the p-value obtained after a Fisher exact test among the three replicates of the single-spore cultures was > 0.05 (*i.e*. allele frequencies at each site are very similar among the replicates).

Isolate	Reference assembly	Pair of cultures tested (A-B)	Number of sites	Proportion of accurate sites in replicates of culture A	Proportion of accurate sites in replicates of culture B
B4	N6	ssc1-ssc2	5070	0.941	0.904
B4	N6	ssc1-ssc3	4671	0.938	0.94
B4	N6	ssc1-ssc4	4785	0.937	0.899
B4	N6	ssc1-ssc5	4443	0.943	0.935
B4	N6	ssc1-ssc6	4760	0.94	0.884
B4	N6	ssc2-ssc3	4714	0.905	0.942
B4	N6	ssc2-ssc4	4832	0.906	0.901
B4	N6	ssc2-ssc5	4749	0.899	0.939
B4	N6	ssc2-ssc6	4952	0.908	0.882
B4	N6	ssc3-ssc4	4488	0.938	0.895
B4	N6	ssc3-ssc5	4333	0.94	0.935
B4	N6	ssc3-ssc6	4435	0.94	0.876
B4	N6	ssc4-ssc5	4407	0.898	0.935
B4	N6	ssc4-ssc6	4594	0.902	0.88
B4	N6	ssc5-ssc6	4383	0.935	0.871
B4	Rhiir2	ssc1-ssc2	3298	0.95	0.906
B4	Rhiir2	ssc1-ssc3	3020	0.947	0.95
B4	Rhiir2	ssc1-ssc4	3038	0.945	0.916
B4	Rhiir2	ssc1-ssc5	2372	0.945	0.954
B4	Rhiir2	ssc1-ssc6	3007	0.944	0.873
B4	Rhiir2	ssc2-ssc3	3087	0.904	0.95
B4	Rhiir2	ssc2-ssc4	3157	0.905	0.917
B4	Rhiir2	ssc2-ssc5	2438	0.89	0.951
B4	Rhiir2	ssc2-ssc6	3131	0.902	0.868
B4	Rhiir2	ssc3-ssc4	2887	0.949	0.909
B4	Rhiir2	ssc3-ssc5	2340	0.949	0.956
B4	Rhiir2	ssc3-ssc6	2771	0.95	0.867
B4	Rhiir2	ssc4-ssc5	2353	0.916	0.954
B4	Rhiir2	ssc4-ssc6	2894	0.92	0.859
B4	Rhiir2	ssc5-ssc6	2348	0.955	0.846
DAOM 197198	N6	ssc1-ssc2	3263	0.91	0.794
DAOM 197198	N6	ssc1-ssc3	3259	0.907	0.844
DAOM 197198	N6	ssc2-ssc3	3192	0.801	0.833
DAOM 197198	Rhiir2	ssc1-ssc2	1399	0.935	0.871
DAOM 197198	Rhiir2	ssc1-ssc3	1386	0.937	0.891
DAOM 197198	Rhiir2	ssc2-ssc3	1392	0.875	0.888

SUPPLEMENTARY TABLE 3

Genome-wide pairwise qualitative genetic differences among single-spore cultures (ssc) of isolate B4. Values represent: the total number of sites tested, number of sites and the proportion of sites that showed a mean F_{st} value higher than 0.2 between single-spore cultures, the mean genome-wide F_{st} between single-spore cultures, F_{st} values higher than 0.2 between replicates of the first single-spore culture (ssc A) and between replicates of the second single-spore culture (ssc B).

Isolate B4	**Cultures 1-2**	**Cultures 1-3**	**Cultures 1-4**	**Cultures 1-5**	**Cultures 1-6**
Total number of sites	4761	4374	4535	4176	4336
Number of sites with F_{ST} > 0.2 between cultures	1	0	0	1	8
% of sites F_{ST} > 0.2	**0.02%**	**0.00%**	**0.00%**	**0.02%**	**0.19%**
Mean F_{ST}	0.0145	0.0121	0.0128	0.0148	0.0151
F_{ST} > 0.2 in ssc A	0	1	0	0	0
F_{ST} > 0.2 in ssc B	1	0	3	0	4
	Cultures 2-3	**Cultures 2-4**	**Cultures 2-5**	**Cultures 2-6**	**Cultures 3-4**
Total number of sites	4376	4607	4508	4528	4317
Number of sites with F_{ST} > 0.2 between cultures	0	0	0	0	0
% of sites F_{ST} > 0.2	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
Mean F_{ST}	0.0127	0.0122	0.0119	0.0124	0.0119
F_{ST} > 0.2 ssc A	0	0	0	0	1
F_{ST} > 0.2 ssc B	0	3	0	3	2
	Cultures 3-5	**Cultures 3-6**	**Cultures 4-5**	**Cultures 4-6**	**Cultures 5-6**
Total number of sites	4091	3979	4230	4184	4003
Number of sites with F_{ST} > 0.2 between cultures	0	0	0	0	0
% of sites F_{ST} > 0.2	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**
Mean F_{ST}	0.0128	0.1448	0.0124	0.0136	0.0134
F_{ST} > 0.2 in ssc A	0	0	5	7	0
F_{ST} > 0.2 in ssc B	0	4	0	3	3

SUPPLEMENTARY TABLE 4

Genome-wide pairwise qualitative differences among single spore cultures (ssc) of isolate DAOM 197198. Values represent: Total number of sites tested, the number of sites and the proportion of sites that showed a mean F_{st} value higher than 0.2 between single-spore cultures, mean genome-wide F_{st} between single-spore cultures and F_{st} values higher than 0.2 between replicates of the first single-spore culture (ssc A) and between replicates of the second single-spore culture (ssc B).

Isolate DAOM 197198	Cultures 1-2	Cultures 1-3	Cultures 2-3
Total nb of sites	3160	3004	2991
Number of sites with F_{ST} > 0.2 between cultures	13	2	5
% of sites F_{ST} > 0.2	**0.41%**	**0.07%**	**0.17%**
mean F_{ST}	0.0123	0.01	0.0119
F_{ST} > 0.2 in ssc A	3	4	12
F_{ST} > 0.2 in ssc B	13	12	13

SUPPLEMENTARY TABLE 5

Summary of sites that were tested in all the pairwise comparisons among single-spore cultures of isolates B4 and DAOM 197198 that could not be classified into single-copy or multiple-copy sites. Values represent the number of sites that displayed quantitative genetic differences in allele frequencies among the single-spore cultures. For each comparison we used the data aligned to the N6 and Rhiir2 genome assemblies independently. We report the total number of sites tested and the number and percentage of sites that displayed significant differences in allele frequency.

Summary per locus	Isolate B4		Isolate DAOM 197198	
No info prediction	**N6**	**Rhiir2**	**N6**	**Rhiir2**
Total sites	1408	585	811	209
Significant sites (*p-value* >0.05)	476	177	99	25
% of significant sites	**33.81%**	**30.26%**	**12.21%**	**11.96%**

SUPPLEMENTARY TABLE 6

Summary of the number of scaffolds where significant quantitative genetic differences in allele frequencies among single-spore cultures were detected in isolates B4 and DAOM 197198. For each comparison, we used the data aligned to the N6 and Rhiir2 genome assemblies independently. We report the total number of scaffolds where sites were tested, and the number and percentage of scaffolds where significant differences in allele frequency occurred.

Summary per scaffold	Isolate B4		Isolate DAOM 197198	
	N6	Rhiir2	N6	Rhiir2
Total number of scaffolds	962	386	859	291
Number of scaffolds with significant sites	489	228	205	94
%	**50.83%**	**59.07%**	**23.86%**	**32.30%**

SUPPLEMENTARY TABLE 7

Pairwise comparisons of quantitative genetic differences in allele frequencies between single-spore cultures (ssc) of isolate B4. For each comparison we used the data aligned to the N6 and Rhiir2 genome assemblies independently. We reported the total number of sites tested, and the number and the percentage of sites that showed significant differences in allele frequency. All these values were calculated for: single-copy sites, multiple-copy sites, sites that could not be classified into either single-copy or multiple-copy sites, and the total number of sites tested.

	Reference alignment	Single-copy			Multiple-copies			No prediction			Total sites		
		Total	Significant	%	Total	Significant	%	Total	Significant	%	Total	Significant	%
ssc1-ssc2	**N6**	1343	109	**8.12%**	2661	283	**10.64%**	1066	147	**13.79%**	5070	539	**10.63%**
ssc1-ssc3	**N6**	1207	61	**5.05%**	2491	167	**6.70%**	973	107	**11.00%**	4671	335	**7.17%**
ssc1-ssc4	**N6**	1299	64	**4.93%**	2492	192	**7.70%**	994	70	**7.04%**	4785	326	**6.81%**
ssc1-ssc5	**N6**	1134	133	**11.73%**	2393	251	**10.49%**	916	110	**12.01%**	4443	494	**11.12%**
ssc1-ssc6	**N6**	1247	89	**7.14%**	2533	229	**9.04%**	980	113	**11.53%**	4760	431	**9.05%**
ssc2-ssc3	**N6**	1221	133	**10.89%**	2531	221	**8.73%**	962	83	**8.63%**	4714	437	**9.27%**
ssc2-ssc4	**N6**	1266	59	**4.66%**	2578	119	**4.62%**	988	60	**6.07%**	4832	238	**4.93%**
ssc2-ssc5	**N6**	1189	52	**4.37%**	2579	110	**4.27%**	981	38	**3.87%**	4749	200	**4.21%**
ssc2-ssc6	**N6**	1310	46	**3.51%**	2636	127	**4.82%**	1006	25	**2.49%**	4952	198	**4.00%**
ssc3-ssc4	**N6**	1182	32	**2.71%**	2393	142	**5.93%**	913	66	**7.23%**	4488	240	**5.35%**
ssc3-ssc5	**N6**	1076	107	**9.94%**	2377	155	**6.52%**	880	107	**12.16%**	4333	369	**8.52%**
ssc3-ssc6	**N6**	1130	88	**7.79%**	2427	146	**6.02%**	878	77	**8.77%**	4435	311	**7.01%**
ssc4-ssc5	**N6**	1113	58	**5.21%**	2388	128	**5.36%**	906	87	**9.60%**	4407	273	**6.19%**
ssc4-ssc6	**N6**	1209	29	**2.40%**	2448	147	**6.00%**	937	47	**5.02%**	4594	223	**4.85%**
ssc5-ssc6	**N6**	1091	26	**2.38%**	2414	73	**3.02%**	878	23	**2.62%**	4383	122	**2.78%**

Average	**N6**	1201.13	72.4	**6.06%**	2489.4	166	**6.66%**	950.53	77.33	**8.12%**	4641.07	315.73	**6.79%**
ssc1-ssc2	**Rhiir2**	1504	126	**8.38%**	1316	127	**9.65%**	478	59	**12.34%**	3298	312	**9.46%**
ssc1-ssc3	**Rhiir2**	1387	55	**3.97%**	1197	106	**8.86%**	436	39	**8.94%**	3020	200	**6.62%**
ssc1-ssc4	**Rhiir2**	1435	130	**9.06%**	1152	80	**6.94%**	451	43	**9.53%**	3038	253	**8.33%**
ssc1-ssc5	**Rhiir2**	1066	62	**5.82%**	962	82	**8.52%**	344	29	**8.43%**	2372	173	**7.29%**
ssc1-ssc6	**Rhiir2**	1399	90	**6.43%**	1180	126	**10.68%**	428	36	**8.41%**	3007	252	**8.38%**
ssc2-ssc3	**Rhiir2**	1429	74	**5.18%**	1234	87	**7.05%**	424	29	**6.84%**	3087	190	**6.15%**
ssc2-ssc4	**Rhiir2**	1459	56	**3.84%**	1254	55	**4.39%**	444	16	**3.60%**	3157	127	**4.02%**
ssc2-ssc5	**Rhiir2**	1065	50	**4.69%**	1018	54	**5.30%**	355	12	**3.38%**	2438	116	**4.76%**
ssc2-ssc6	**Rhiir2**	1408	49	**3.48%**	1293	55	**4.25%**	430	9	**2.09%**	3131	113	**3.61%**
ssc3-ssc4	**Rhiir2**	1353	37	**2.73%**	1122	54	**4.81%**	412	22	**5.34%**	2887	113	**3.91%**
ssc3-ssc5	**Rhiir2**	1048	43	**4.10%**	954	61	**6.39%**	338	17	**5.03%**	2340	121	**5.17%**
ssc3-ssc6	**Rhiir2**	1263	61	**4.83%**	1119	60	**5.36%**	389	16	**4.11%**	2771	137	**4.94%**
ssc4-ssc5	**Rhiir2**	1066	54	**5.07%**	941	53	**5.63%**	346	22	**6.36%**	2353	129	**5.48%**
ssc4-ssc6	**Rhiir2**	1334	70	**5.25%**	1148	45	**3.92%**	412	18	**4.37%**	2894	133	**4.60%**
ssc5-ssc6	**Rhiir2**	1044	40	**3.83%**	974	29	**2.98%**	330	23	**6.97%**	2348	92	**3.92%**
Average	**Rhiir2**	1284	66.47	**5.11%**	1124.27	71.6	**6.32%**	401.13	26	**6.38%**	2809.4	164.07	**5.78%**
		Total	**Significant**	**%**	**Total**	**Significant**	**%**	**Total**	**Significant**	**%**	**Total**	**Significant**	**%**

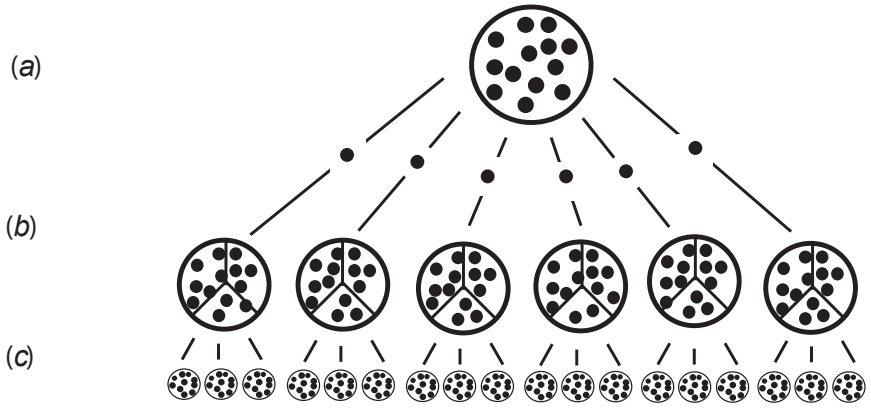
SUPPLEMENTARY TABLE 8

Pairwise comparisons of quantitative genetic differences in allele frequencies among single-spore cultures (ssc) of isolate DAOM 197198. For each comparison we used the data aligned to the N6 and Rhiir2 genome assemblies independently. We reported the total number of sites tested, and the number and the percentage of sites that showed significant differences. All these values were calculated for: single-copy sites, multiple-copy sites, sites that could not be classified into single-copy or multiple-copy sites, and the total number of sites tested.

		Single-copy			**Multiple-copies**			**No prediction**			**Total sites**		
		Total	**Significant**	**%**	**Total**	**Significant**	**%**	**Total**	**Significant**	**%**	**Total**	**Significant**	**%**
ssc1-ssc2	**N6**	853	90	**10.55%**	1771	115	**6.49%**	639	44	**6.89%**	3263	249	**7.63%**
ssc1-ssc3	**N6**	869	58	**6.67%**	1765	131	**7.42%**	625	59	**9.44%**	3259	248	**7.61%**
ssc2-ssc3	**N6**	828	55	**6.64%**	1740	67	**3.85%**	624	20	**3.21%**	3192	142	**4.45%**
Average	**N6**	850	67.67	**7.96%**	1758.67	104.33	**5.92%**	629.33	41	**6.51%**	3238	213	**6.56%**
ssc1-ssc2	**Rhiir2**	640	82	**12.81%**	588	36	**6.12%**	171	9	**5.26%**	1399	127	**9.08%**
ssc1-ssc3	**Rhiir2**	616	48	**7.79%**	609	33	**5.42%**	161	12	**7.45%**	1386	93	**6.71%**
ssc2-ssc3	**Rhiir2**	630	70	**11.11%**	591	20	**3.38%**	171	12	**7.02%**	1392	102	**7.33%**
Average	**Rhiir2**	628.67	66.67	**10.57%**	596	29.67	**4.98%**	167.67	11	**6.58%**	1392.33	107.33	**7.71%**

SUPPLEMENTARY FIGURES 1

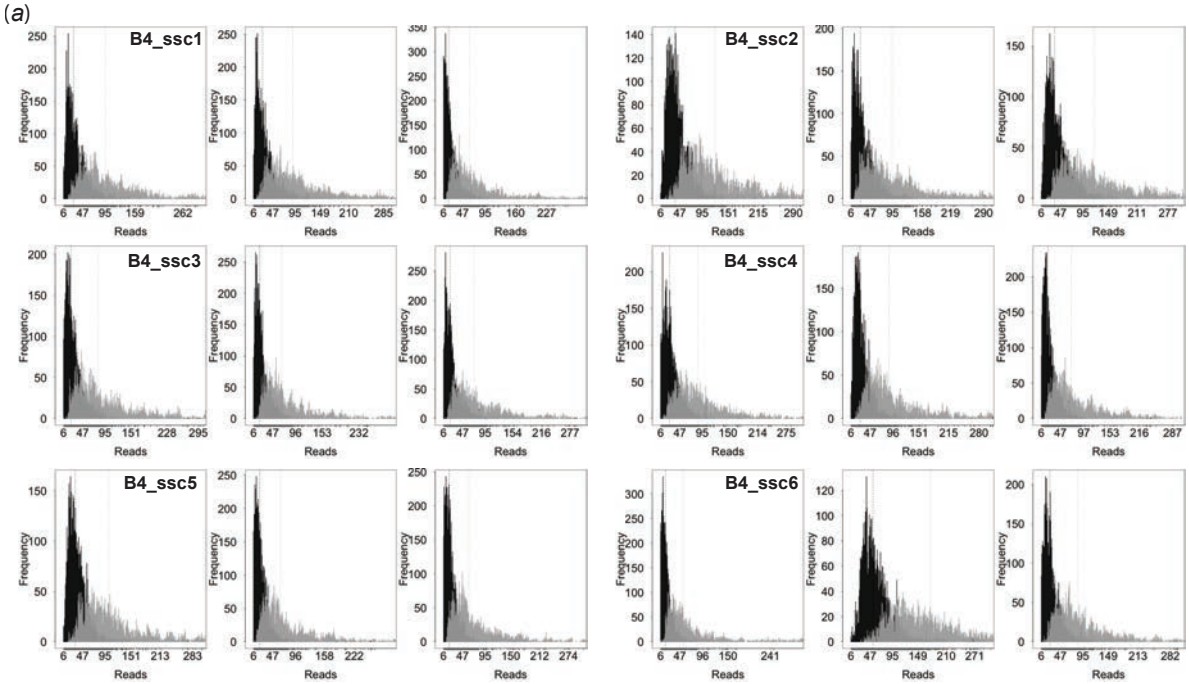
Experimental design. (*a*) Parental isolate (B4 or DAOM 197198). (*b*) Single-spores isolated from the parental isolates were used to initiate single-spore cultures. (*c*) In order to produce biological replicates, each single-spore culture was then divided in three equal parts and sub-cultured again. Large open circles represent Petri dishes and small black circles represent spores.



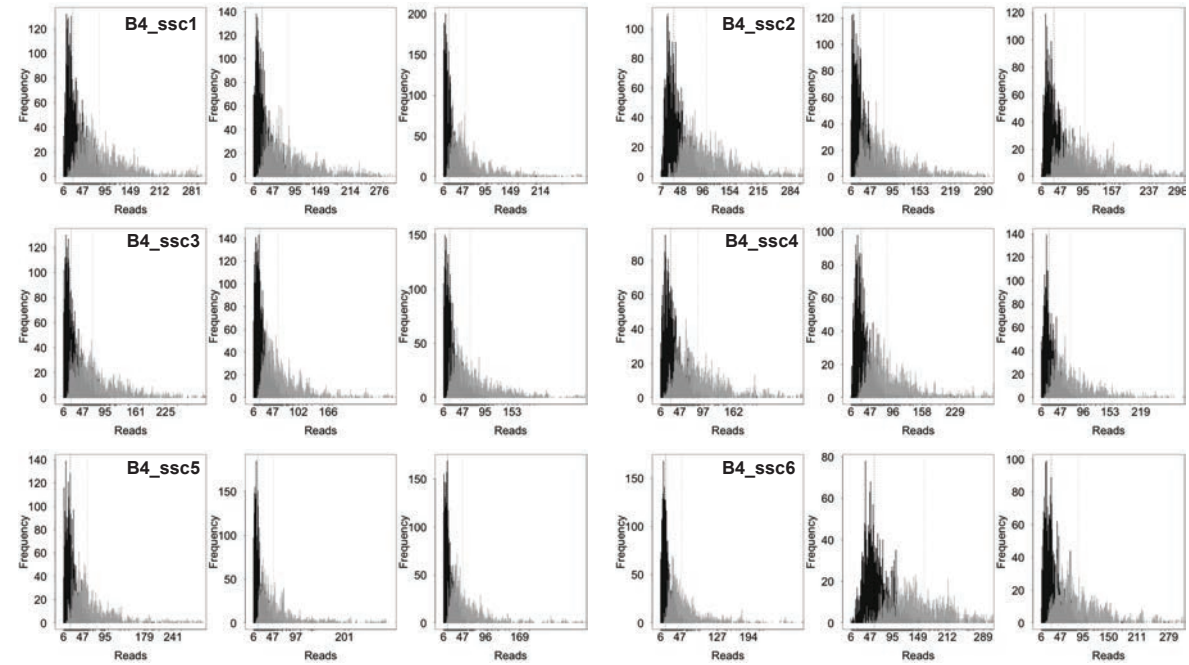
SUPPLEMENTARY FIGURE 2.

Coverage distributions of the poly-allelic site for each sample. The distribution of minimum allele coverage per site (black) and the total coverage per site (grey) is shown. The figure shows the three replicates per each single-spore sibling. (*a*) B4 aligned to N6 assembly, (*b*) B4 aligned to Rhirr2 assembly, (c) DAOM 197198 aligned to N6 assembly and (*d*) DAOM197198 aligned to Rhirr2 assembly.

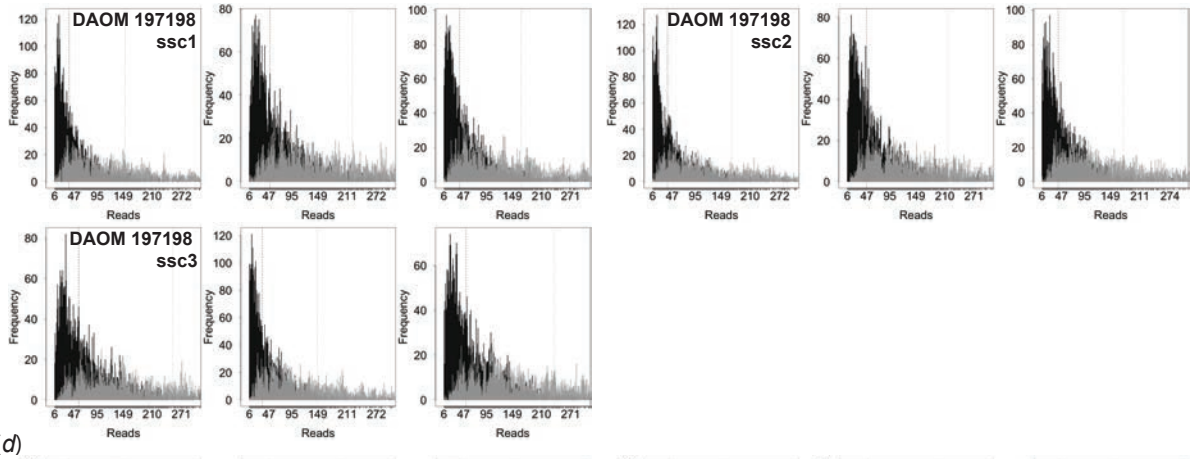
(*a*)



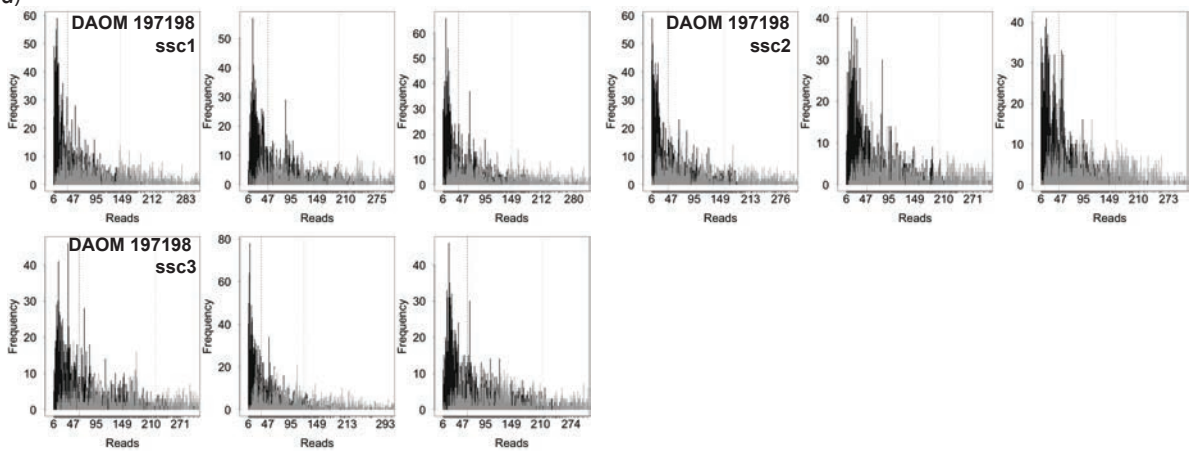
(*b*)



(*c*)

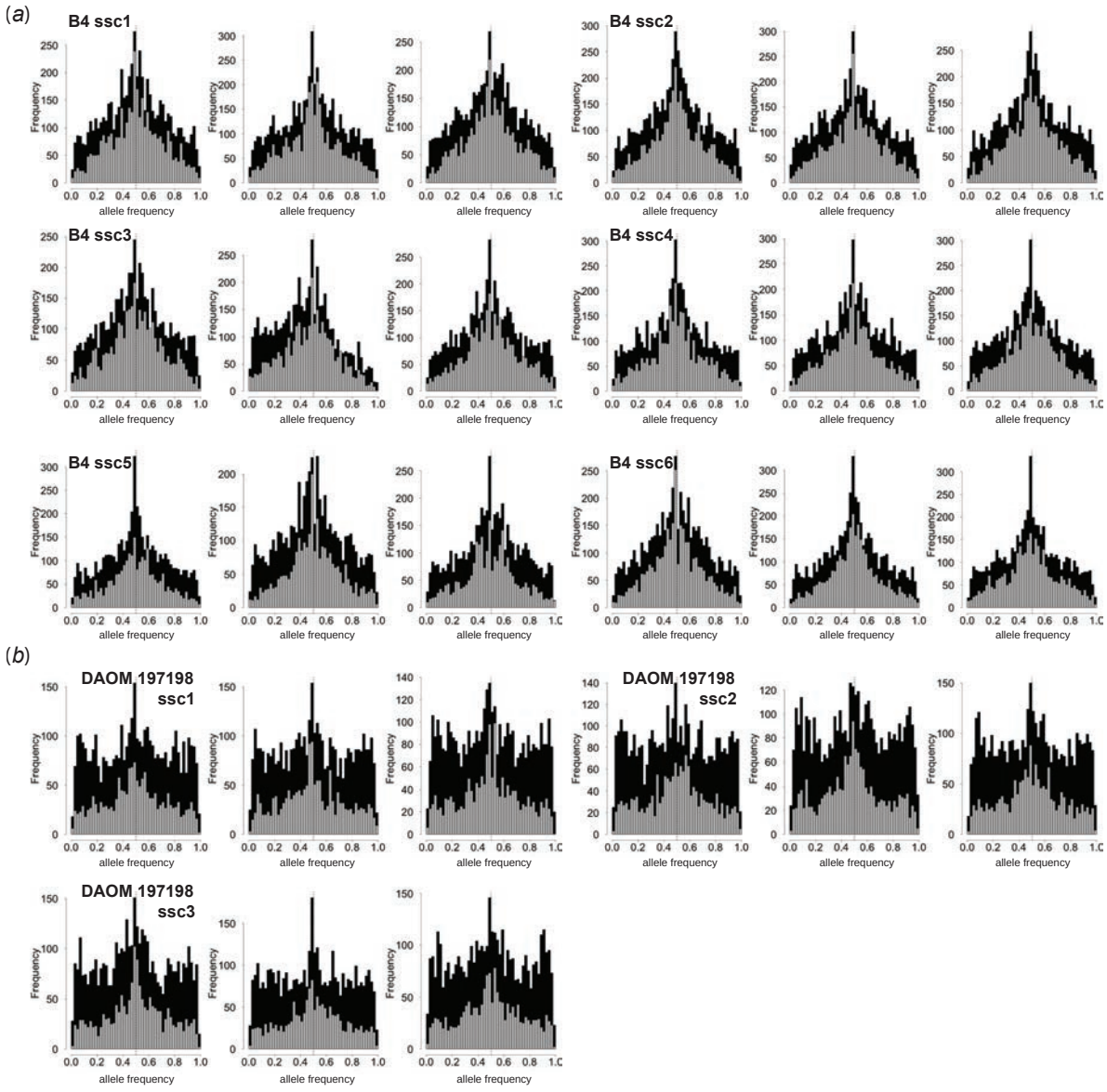

DAOM 197198
ssc1
DAOM 197198
ssc2
DAOM 197198
ssc3
Frequency
Reads

(*d*)


DAOM 197198
ssc1
DAOM 197198
ssc2
DAOM 197198
ssc3
Frequency
Reads

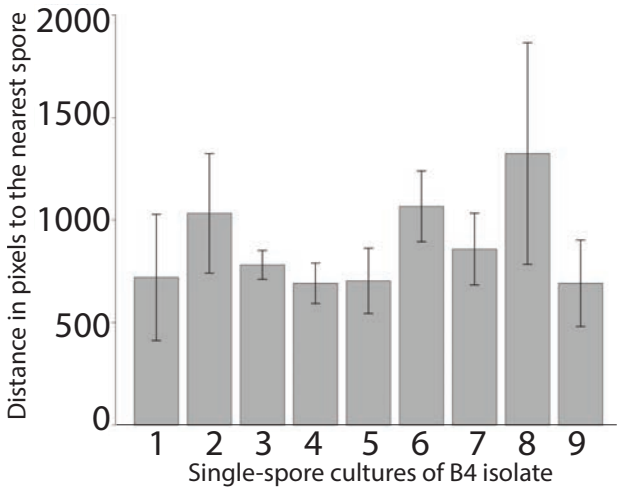
SUPPLEMENTARY FIGURE 3

Allele frequency distribution plots of poly-allelic sites of the dataset aligned to N6 (black) and to Rhiir2 (Grey). (*a*) Samples of single-spore cultures of isolate B4, each one with three replicates. (*b*) Samples of single-spore cultures of isolate DAOM 197198, each one with three replicates.



SUPPLEMENTARY FIGURE 4.

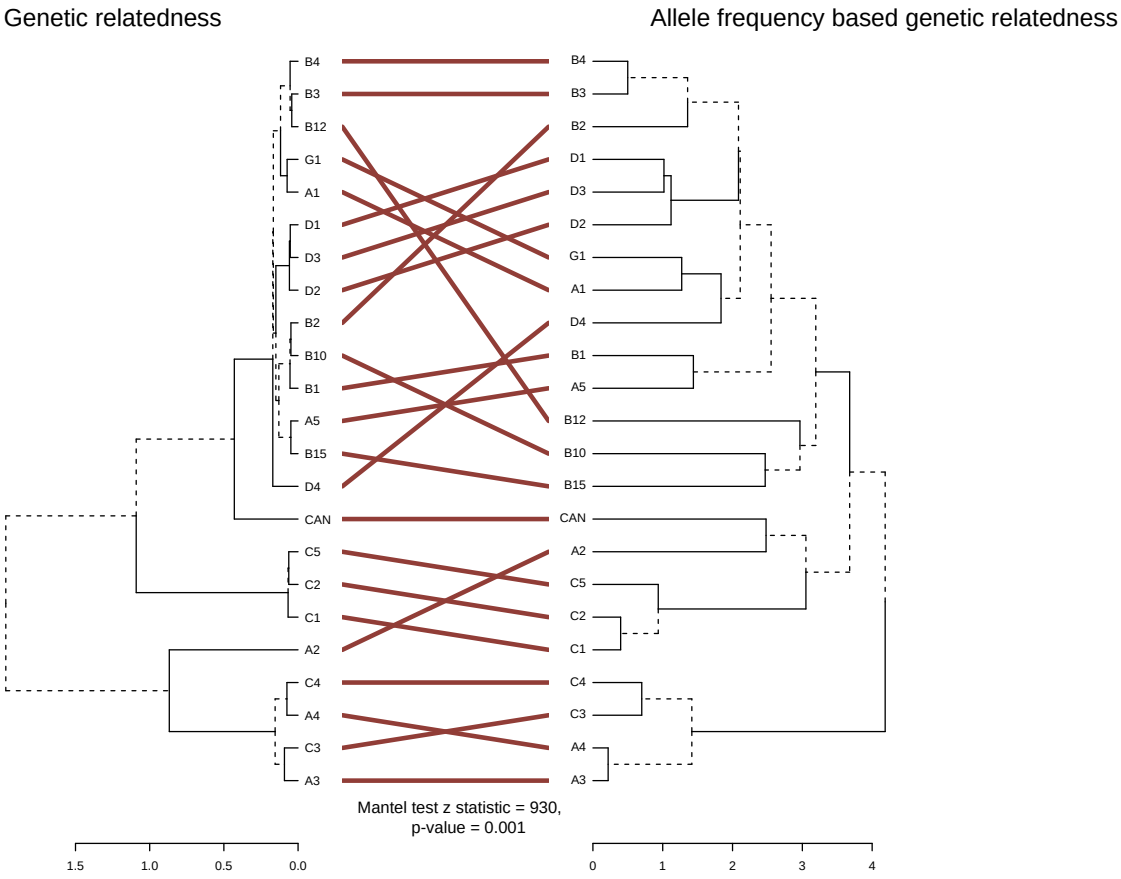
Spatial distribution of spores in the single-spore sibling cultures of *R. irregularis* isolate B4, as measured by the distance from random points to the nearest spore. For each isolate, 6 different areas of 2 cm^2 on three different Petri dishes were used for the counting. Mean value per single-spore culture and standard deviation are shown in the figure.



Annexe II. Phylogenetic signal shows that intra-specific genetic variation in Rhizophagus irregularis causes variation in fungal traits and growth of a globally important plant

SUPPLEMENTARY FIGURE 1

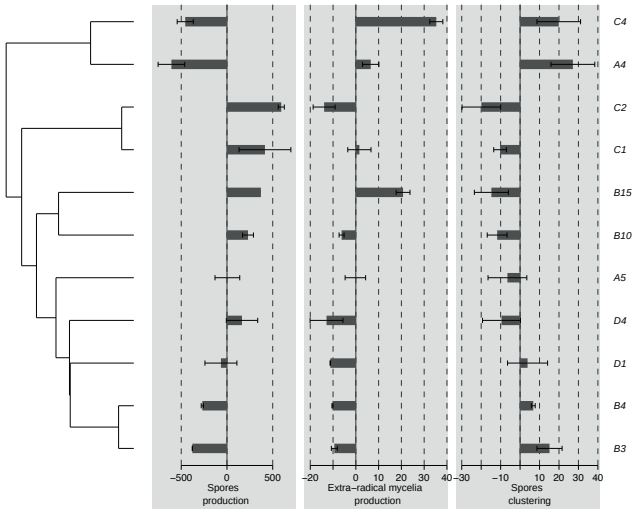
Genetic relatedness dendrogram (left), Genetic relatedness dendrogram based on allele frequency differences among isolates (right). Solid lines represent nodes which are conserved on both dendrograms. Dashed lines represent different nodes between the two dendrograms. Red lines links the same edges on both dendrograms.



SUPPLEMENTARY FIGURE 2

Phylogenetic signal in fungal and plant response traits using the among isolates differentiation based in the allele frequency changes among isolates. a,c) Representation of traits along with the dendrogram for different R. irregularis isolates. The trait values are centred (value – mean). b,d) Phylogenetic signal of fungal and plant traits. Summary statistic of Abouheif's Cmean and Moran's I. Metric statistics (above) and significance (p-value) are shown.

a)



b)

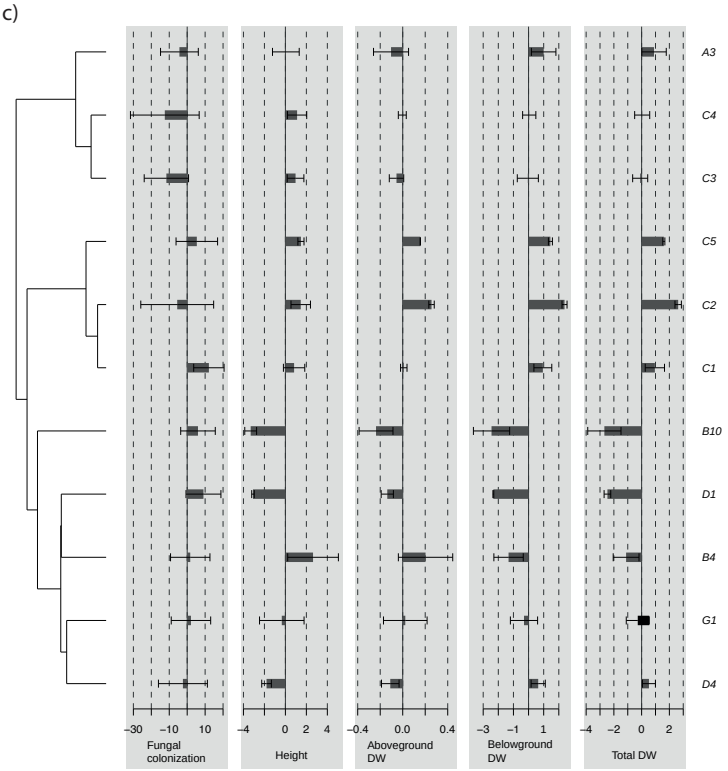
Statistic

	Cmean	I
Spore production	0.456	0.18327
Extra-radical mycelia	0.153	-0.00912
Spore clustering	0.446	0.10952

pvalue

	Cmean	I
Spore production	0.0056	0.0037
Extra-radical mycelia	0.0993	0.1173
Spore clustering	0.007	0.0349

c)



d)

Statistic

	Cmean	I
Fungal colonization	0.2859	0.0614
Height	-0.1009	-0.0229
Aboveground DW	-0.0962	-0.019
Belowground DW	0.3031	0.0895
Total DW	0.2826	0.0926

pvalue

	Cmean	I
Fungal colonization	0.0452	0.0818
Height	0.4404	0.2113
Aboveground DW	0.4391	0.203
Underground DW	0.0341	0.0418
Total DW	0.039	0.0417

SUPPLEMENTARY FIGURE 3

Genetic relationship of R. irreguaris isolates based on presence/absence of SNP's (Wyss et al., 2016).We report for each isolate the host plant used as the initial trap-culture of initial cultures of the R. irreguaris isolates (Croll et al., 2008). Highlighted isolates are the ones used in this study.

.

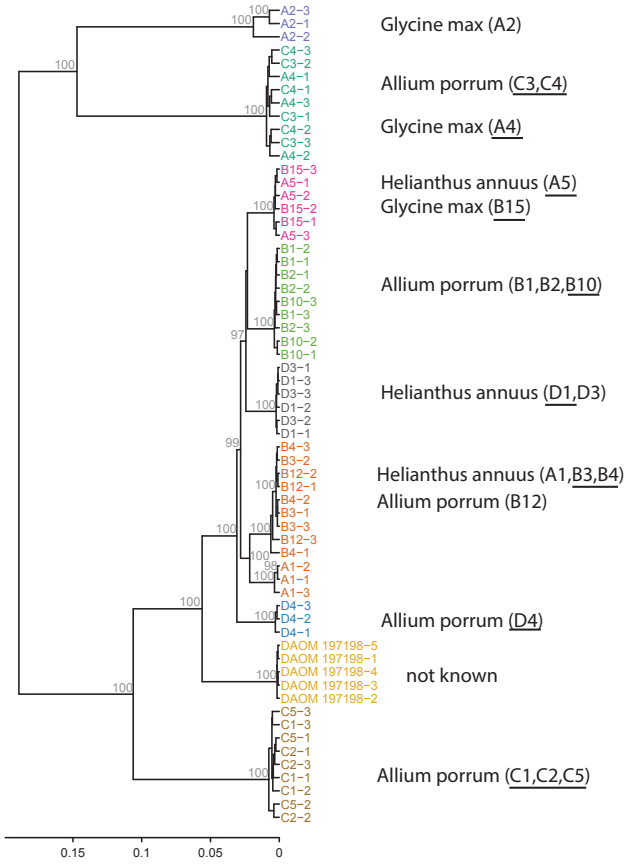
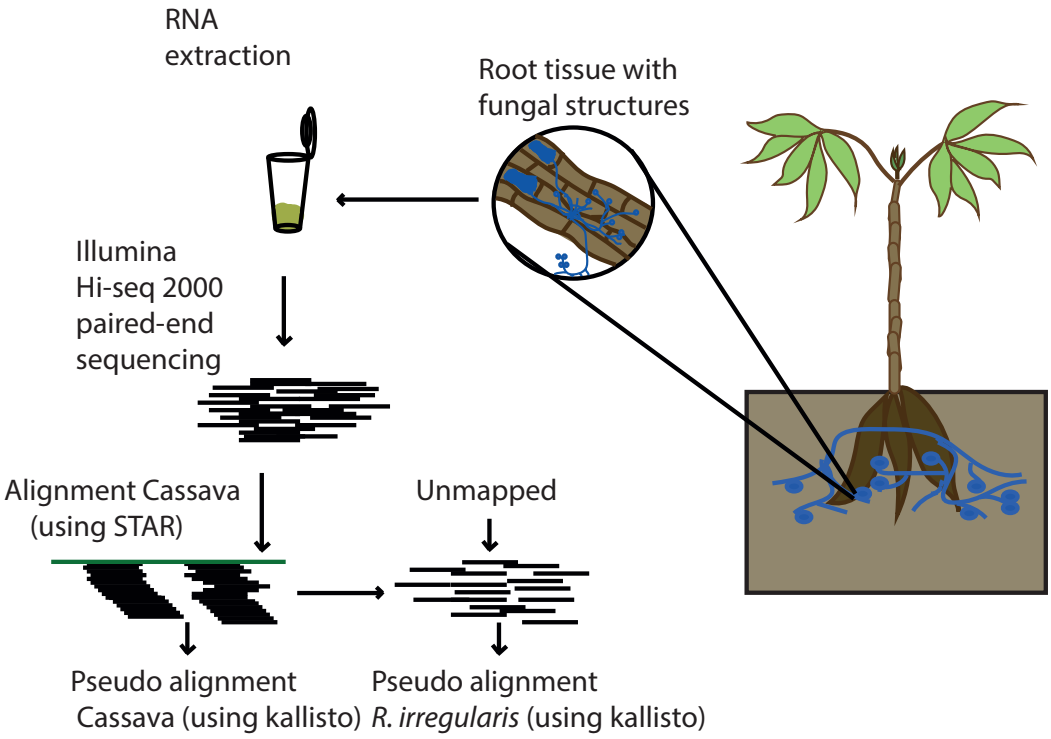


Figure from Wyss et al., 2016 Host used for initial trap-culture (Croll et al., 2008)

Annexe III. Gene-gene interactions between cassava and its mycorrhizal partner revealed by intra-specific variation of both partners

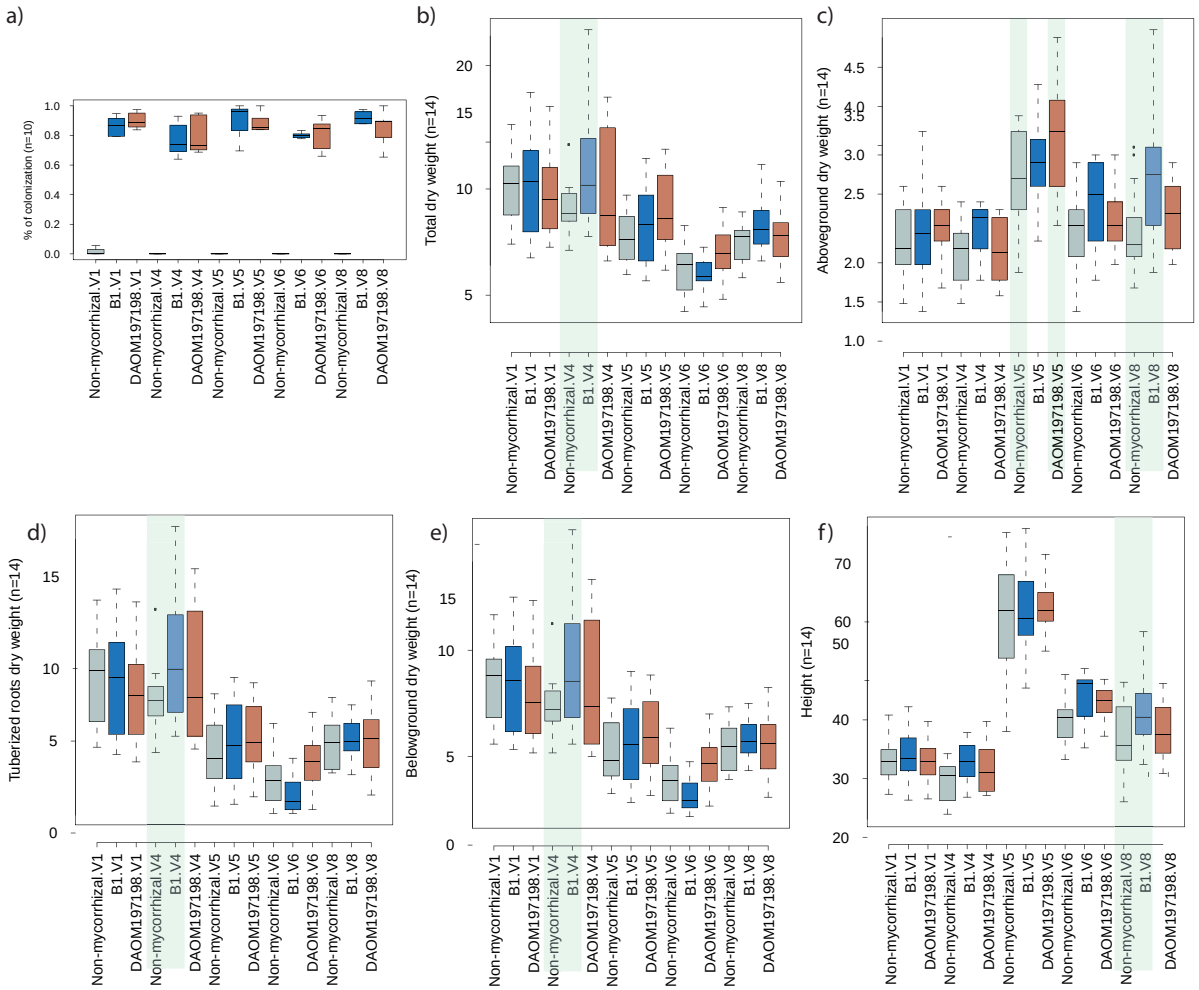
Extended data figure 1.

Experimental approach of the experiment. Randomly chosen fine roots (brown) containing fungal structures (blue) were isolated and used for RNA extraction and library preparation. We then aligned the obtained reads to the cassava reference genome. We obtained the cassava dataset by performing the pseudoalignment using the gene prediction on cassava. We then use the unmapped reads to cassava to obtain the R. irregularis dataset by performing the pseudoalignment using the gene prediction on R. irregularis.



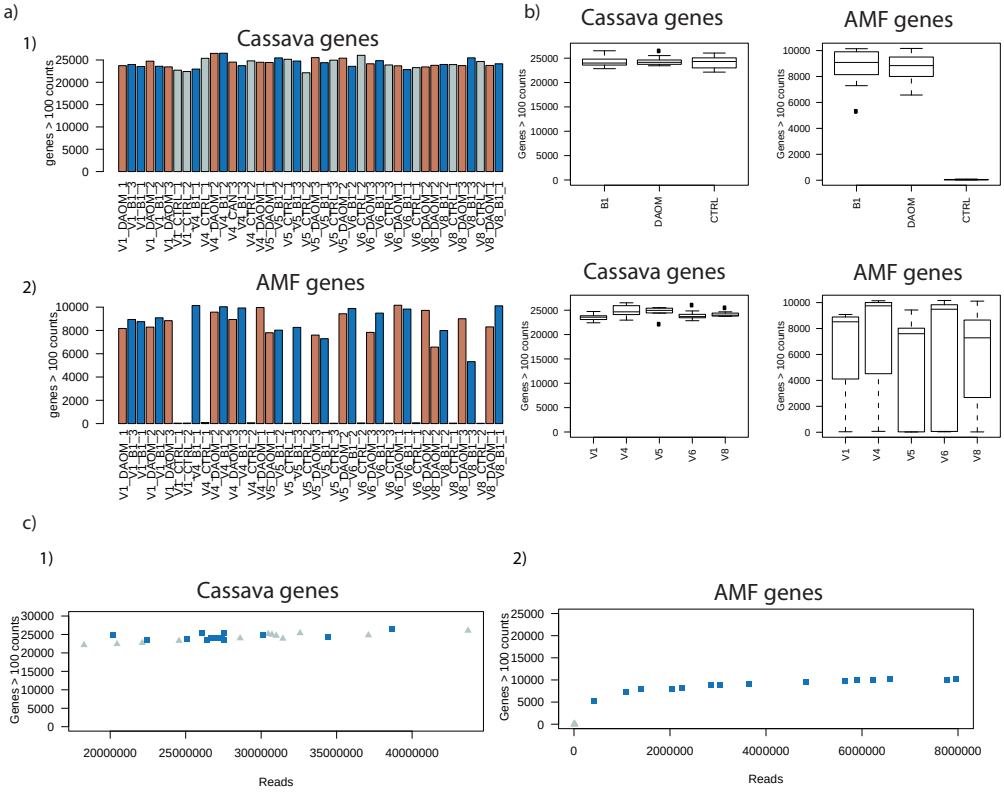
Extended data figure 2.

Fungal colonization and phenotypic response of cassava to R. irregularis. a) Percentage of root colonization by R. irregularis. b) Total dry weight, c) aboveground dry weight, d) tuberized roots dry weight, e) belowground dry weight and f) height plant measurements in function of the cultivar and mycorrhizal treatment. Red colour represent DAOM197198 samples, blue colour represent B1 samples and gray represent non-mycorrhizal plants. Green highlighted treatments represent fungal treatments that displayed significant statistical differences.



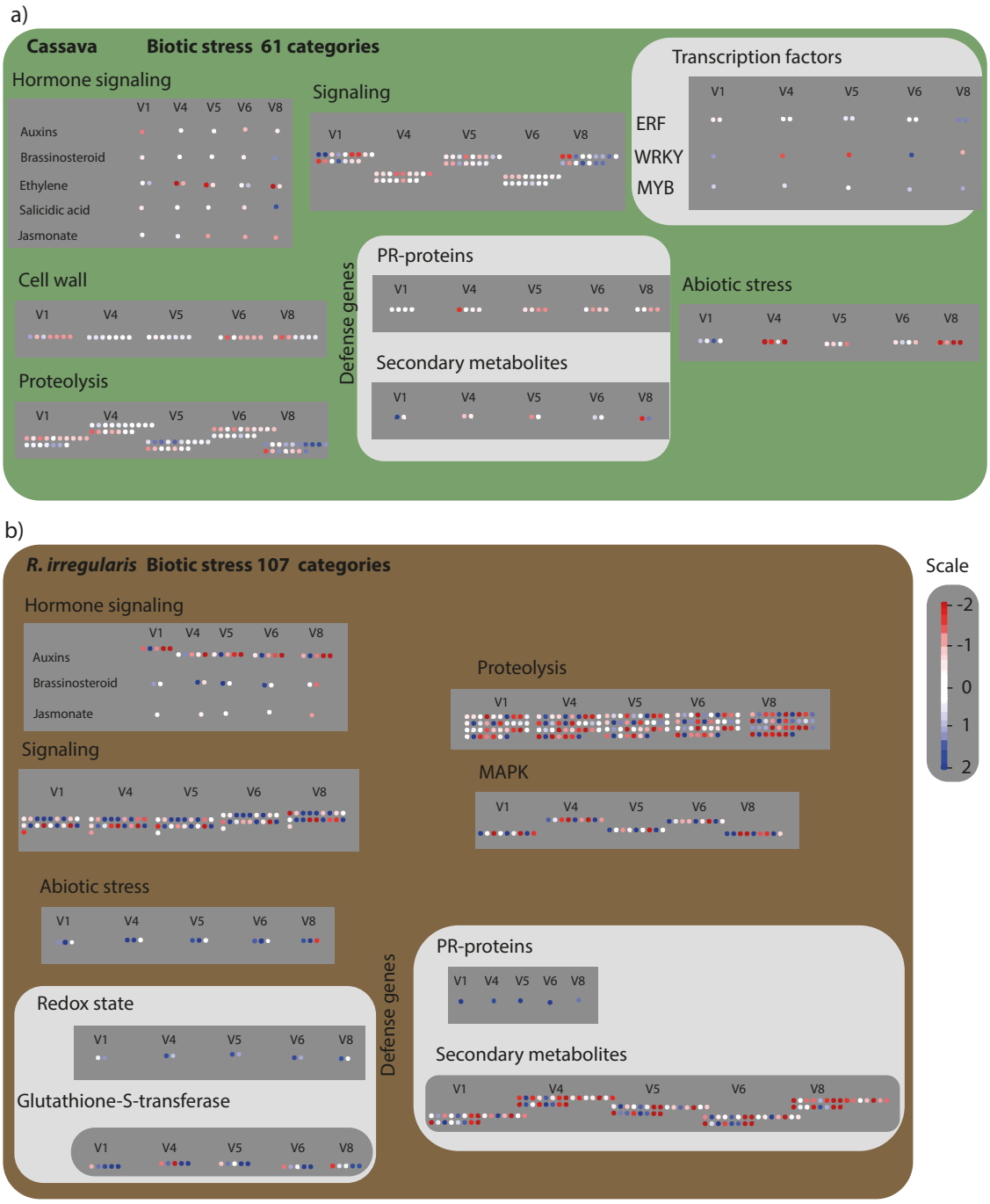
Extended data figure 3.

RNA-seq results on cassava and R. irregularis. a) Samples sequencing results. Barplot showing the number of transcripts that display more than 10 reads in 1) cassava and 2) R. irregularis. b) Boxplots displaying number of transcripts that display more than 10 reads in function of 1,2) mycorrhizal treatment or 3,4) host. 1,3) data shown for cassava and 2,4) data shown for R. irregularis. c) saturation curve of transcript number that display more than 10 counts by the numbers of reads for each sample. Red colour represent DAOM197198 samples, blue colour represent B1 samples and gray represent non-mycorrhizal plants. In cassava, the number of different cassava gene transcripts sequenced was not affected by the sequencing depth of the samples and did not differ significantly among the mycorrhizal treatments but differed among some cassava varieties (Extended Data Figures 3a-c). However, this did not influence the results as transcripts not common to all varieties were removed before analysis and the quantity of different transcripts sequenced per variety did not correlate with the pattern observed on the genes influenced by variety (fig 1b,c) (Extended Data Figure 3d). In R. irregularis, the read number did not influence the number of transcripts between inoculated treatments and there were no significant differences in the number of different transcripts among cassava varieties.



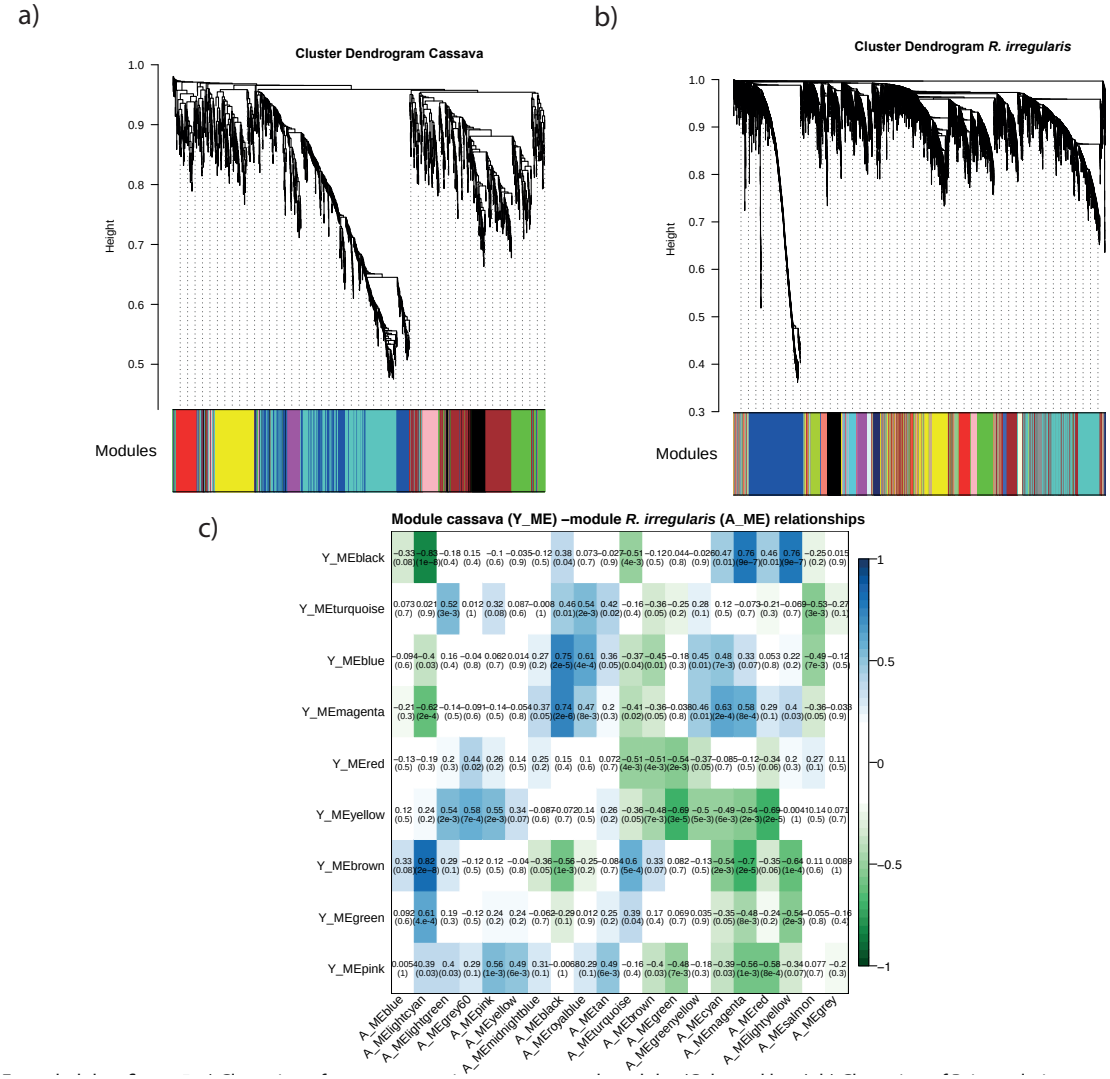
Extended Data Figure 4.

Mapman biotic stress categories log2fold-changes between R. irregularis isolates DAOM197198 and B1 for each cassava cultivar. a) Plant genes. b) fungal genes. Each small square means a gene were log2Fold-change is different between R. irregularis isolates DAOM197198 and B1. Red colours represent isolate DAOM197198 had lower foldchange compared to isolate B1. Blue colours represent that isolata DAOM197198 had higher foldchange comapared to isolate B1.



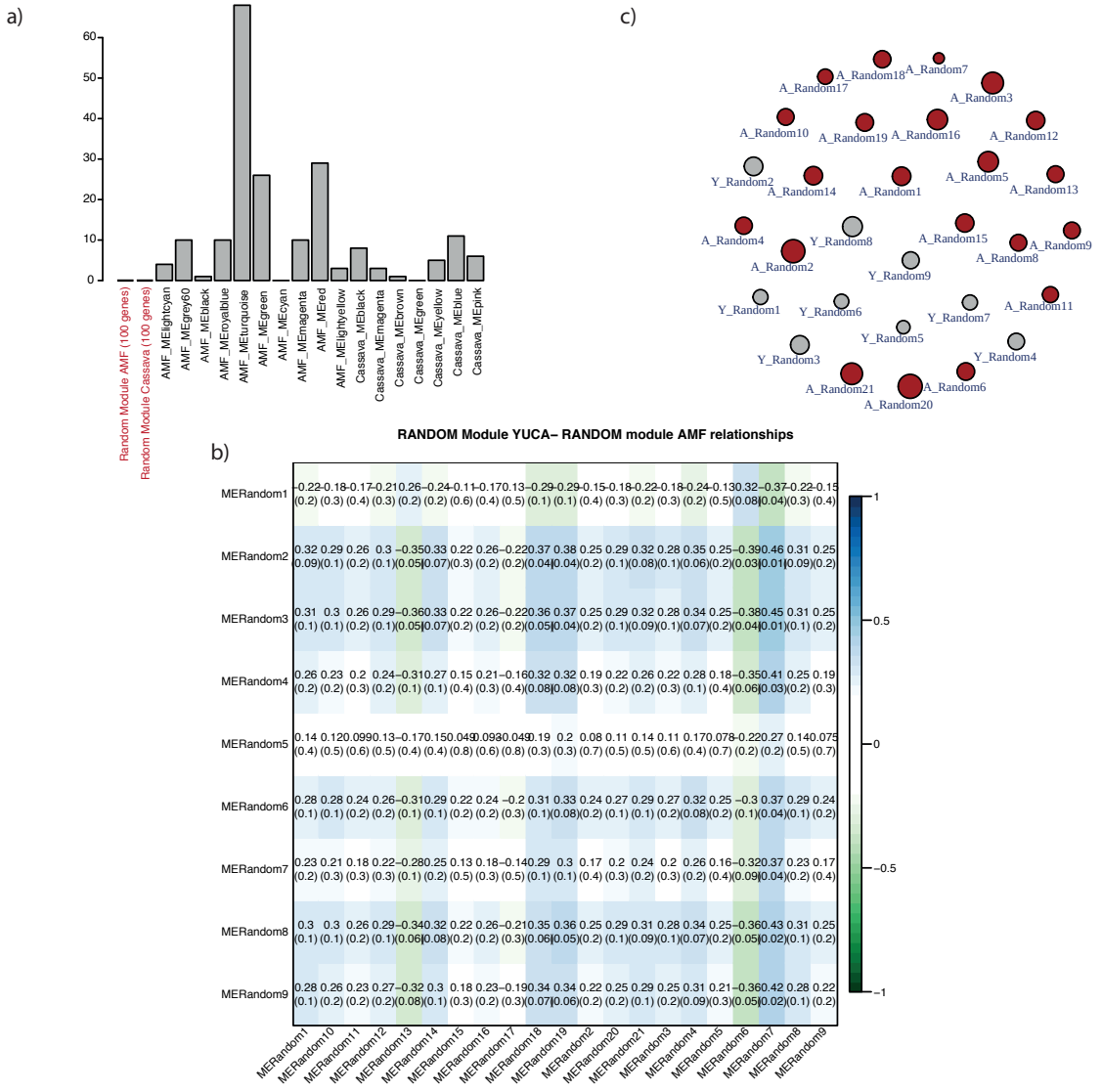
Extended data figure 5.

a) Clustering of cassava genes into co-expressed modules (Coloured bars). b) Clustering of R. irregularis genes into co-expressed modules (Coloured bars). c) Correlation between cassava modules (rows) and R. irregularis modules (columns). for each module-module correlation is reported the correlation coeficient and its p-value. Highly positive correlations are shown in dark blues, as opposite, highly negative correlations are shown in dark green. A threshold cutoff was set at a p-value = 0.001, representing the probability which can be found by chance 0.18 correlated modules in 180 comparisons.



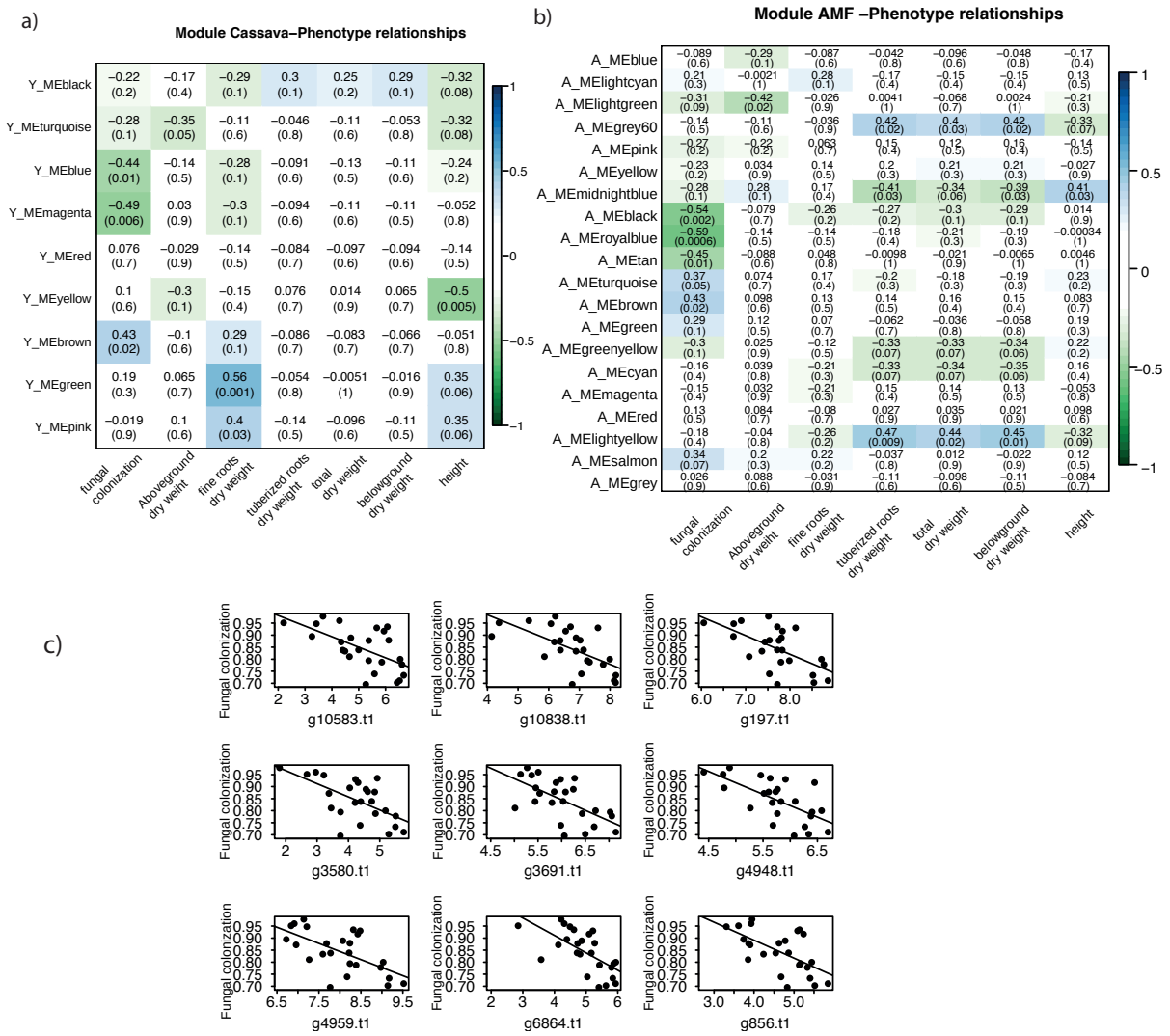
Extended data figure 6.

Random modules composition and network. a) Significant GO terms with more than 3 genes per GO term for random AMF and cassava modules, AMF correlated modules and cassava correlated modules. b) Correlation of random AMF and random cassava modules. Note that no correlations with p-value threshold of 0.001 are found between random AMF modules and random cassava modules. c) Network based on random AMF modules and random cassava modules.



Extended data figure 7.

Correlation of cassava and R. irregularis modules to phenotype a) Correlation of cassava modules and plant phenotypic measurements. b) Correlation of R. irregularis modules and plant phenotypic measurments. Note that a p-value threshold cutoff of 0.001 was used as estimated to be the threshold were 0.18 correlations by chance can occur between pairwise comparison between cassava and R. irregularis modules. c) Plots of most representative and correlated genes to the fungal colonization of the R. irregularis module royalblue.



Supplementary Table 1.

Sequencing information. 42 samples were sequenced. These samples were splited in 5 Cassava cultivars, 2 R. irregularis isolates and a non-mycorrhizal control. We count the number of reads for each sample, the numbers of reads used by kallisto that were used for the gene-count tables and the number of transcripts that possess more than 10 reads. Cassava data set: Total reads observed after quality filters and trimming. R. irregularis data set: Reads that were not mapped to Cassava reference genome (Unmapped)

					Cassava			*R. irregularis*		
Samples	File header	Variety	Cultivar	Treatment	Total reads	Pseudoaligned Cassava	Transcript > 10 counts Cassava (tpm10_C)	Unmapped reads	Pseudoaligned AMF	Transcript > 10 counts AMF(tpm10_A)
V1_DAOM_1	V1-1	V1	V1_CM6438-14	DAOM197198	40731000	32962696	10569	4333044	1794067	6558
V1_B1_3	V1-10	V1	V1_CM6438-14	B1	36745361	27105372	10170	6281081	3042201	6999
V1_B1_1	V1-2	V1	V1_CM6438-14	B1	33998994	26400896	11823	5269097	2858970	7052
V1_DAOM_2	V1-5	V1	V1_CM6438-14	DAOM197198	38772851	31068405	10157	4499642	1902265	6352
V1_B1_2	V1-6	V1	V1_CM6438-14	B1	36968177	27549708	10648	6905467	3655658	7014
V1_DAOM_3	V1-9	V1	V1_CM6438-14	DAOM197198	33622400	25364565	10310	4875538	2381836	6999
V4_B1_1	V4-10	V4	V4_COL2215	B1	28089092	15615196	13361	10267084	6591099	8188
V4_CTRL_1	V4-12	V4	V4_COL2215	CTRL	40861511	32580983	13001	5013584	12462	2925
V4_DAOM_2	V4-13b	V4	V4_COL2215	DAOM197198	57781123	41427389	13316	11676572	4674175	7702
V4_B1_2	V4-14b	V4	V4_COL2215	B1	57325234	38694763	13564	14423564	7764493	7852
V4_DAOM_3	V4-17	V4	V4_COL2215	DAOM197198	40181990	30261582	12832	6984404	3933950	6962
V4_B1_3	V4-18	V4	V4_COL2215	B1	39162831	25097740	13574	10536793	5890155	8195
V4_CTRL_2	V4-20	V4	V4_COL2215	CTRL	45681856	37104641	12713	4098123	11450	1939
V4_DAOM_1	V4-9	V4	V4_COL2215	DAOM197198	50745655	34187835	13192	12508099	6277627	7960
V5_DAOM_1	V5-1	V5	V5_BRA337	DAOM197198	44115391	36178369	13056	4171694	1293199	6564
V5_B1_2	V5-10	V5	V5_BRA337	B1	33952560	27516018	12297	4653392	1389347	7145
V5_CTRL_1	V5-12	V5	V5_BRA337	CTRL	35942843	30476187	13429	3516782	6623	279
V5_B1_3	V5-14	V5	V5_BRA337	B1	26710438	20195593	13000	5914378	2250463	6809
V5_CTRL_2	V5-16	V5	V5_BRA337	CTRL	22453636	18268306	12265	2289017	7956	620

V5_DAOM_3	V5-17	V5	V5_BRA337	DAO M197 198	35031403	29688754	9838	2761972	963263	6974
V5_B1_1	V5-2	V5	V5_BRA337	B1	41558031	34440766	11211	4364114	1076913	6313
V5_CTRL_3	V5-20	V5	V5_BRA337	CTRL	34800268	30716691	9935	1839737	7668	427
V6_B1_2	V6-10	V6	V6_CM4574-7	B1	38534288	22415261	12024	12470868	6222596	7792
V6_CTRL_2	V6-12	V6	V6_CM4574-7	CTRL	55386954	43708357	13042	5861506	11683	1201
V6_DAOM_3	V6-13	V6	V6_CM4574-7	DAO M197 198	38279561	29928531	10396	3343502	900292	7269
V6_B1_3	V6-14	V6	V6_CM4574-7	B1	42493661	30118763	12519	8470462	4828626	7228
V6_CTRL_3	V6-16	V6	V6_CM4574-7	CTRL	39367818	31438475	9285	2724939	11798	555
V6_DAOM_1	V6-5	V6	V6_CM4574-7	DAO M197 198	35815968	22081502	12703	9835933	5198774	8220
V6_B1_1	V6-6	V6	V6_CM4574-7	B1	30553935	14586040	10755	11537375	5641474	7771
V6_CTRL_1	V6-8	V6	V6_CM4574-7	CTRL	33392016	24559806	11817	3573643	10679	778
V6_DAOM_2	V6-9	V6	V6_CM4574-7	DAO M197 198	29904809	16661891	11149	9141363	5035687	7434
V8_DAOM_2	V8-13	V8	V8_CM523-7	DAO M197 198	36304871	30724871	10799	3623444	1245597	5233
V8_B1_2	V8-14	V8	V8_CM523-7	B1	34238143	27413040	12426	5036100	2033931	6637
V8_CTRL_1	V8-16	V8	V8_CM523-7	CTRL	33619848	28607364	10303	2422028	9155	1320
V8_DAOM_3	V8-17	V8	V8_CM523-7	DAO M197 198	43509886	34930496	10683	6039367	3334054	6592
V8_B1_3	V8-18	V8	V8_CM523-7	B1	30375350	26055792	13118	3042973	421247	6265
V8_CTRL_2	V8-20	V8	V8_CM523-7	CTRL	37441996	31000138	11525	2573894	7798	462
V8_DAOM_1	V8-5	V8	V8_CM523-7	DAO M197 198	35304373	28862016	10702	4673452	2251504	6166
V5_DAOM_2	m2V513	V5	V5_BRA337	DAO M197 198	61894858	45311336	12220	12672110	6480817	7038
V1_CTRL_1	V1-16	V1	V1_CM6438-14	CTRL	27490866	22127297	9254	1917370	7685	710
V1_CTRL_2	V1-20	V1	V1_CM6438-14	CTRL	25730307	20452924	7685	1854857	8974	923
V8_B1_1	V8-10	V8	V8_CM523-7	B1	43250047	26682043	12381	13132993	7957189	7787

Supplementary Table 2.

Fungal colonization and plant phenotypic response to isolates DAOM197198, B1 and the mock control. For each phenotypic measurement the analysis of variance is reported. a) fungal colonization on cassava roots. b) Total dry weight. c) Aboveground dry weight. d) tuberized roots dry weight. e) Belowground dry weight. f) Height of plants.

a) **Colonization 5 VAR**

Without non-mycorrhizal plants

*X ~AMF * PLANT CULTIVAR*

GLM(binomial-logit)

	Df	Deviance	Residual Df	Residual Deviance	Pr(>Chi)
NULL	51	213			
AMF	4	17.5	47	196	1.50E-03
CULTIVAR	1	0.16	46	196	0.6927
AMF:CULTIVAR	4	7.26	42	188	0.123

b) **TOTAL DRY WEIGHT 5 VAR**

*X ~AMF * PLANT CULTIVAR + (1 | Block)*

Mixed-linear model (binomial-logit)

	numDF	denDF	F-value	p-value	
(Intercept)	1	182	825	<.0001	***
AMF	2	182	3	0.073	.
CULTIVAR	4	182	33	<.0001	***
AMF:CULTIVAR	8	182	1	0.179	

c) **Aboveground dry weight 5 VAR**

*X ~AMF * PLANT CULTIVAR + (1 | Block)*

Mixed-linear model (binomial-logit)

	numDF	denDF	F-value	p-value	
(Intercept)	1	182	1970	<.0001	***
AMF	2	182	6	0.003	**
CULTIVAR	4	182	21	<.0001	***
AMF:CULTIVAR	8	182	1	0.165	

d) **Tuberized roots dry weight 5 VAR**

*X ~AMF * PLANT CULTIVAR + (1 | Block)*

Mixed-linear model (binomial-logit)

	numDF	denDF	F-value	p-value

(Intercept)	1	182	500	<.0001	***
AMF	2	182	1	0.39	
CULTIVAR	4	182	48	<.0001	***
AMF:CULTIVAR	8	182	1	0.25	

e) **Belowground dry weight 5 VAR**

*X ~AMF * PLANT CULTIVAR + (1 | Block)*

Mixed-linear model (binomial-logit)

	numDF	**denDF**	**F-value**	**p-value**	***
(Intercept)	1	182	588	<.0001	
AMF	2	182	1	0.29	***
CULTIVAR	4	182	45	<.0001	
AMF:CULTIVAR	8	182	1	0.26	

f) **Height 5 VAR**

*X ~AMF * PLANT CULTIVAR + (1 | Block)*

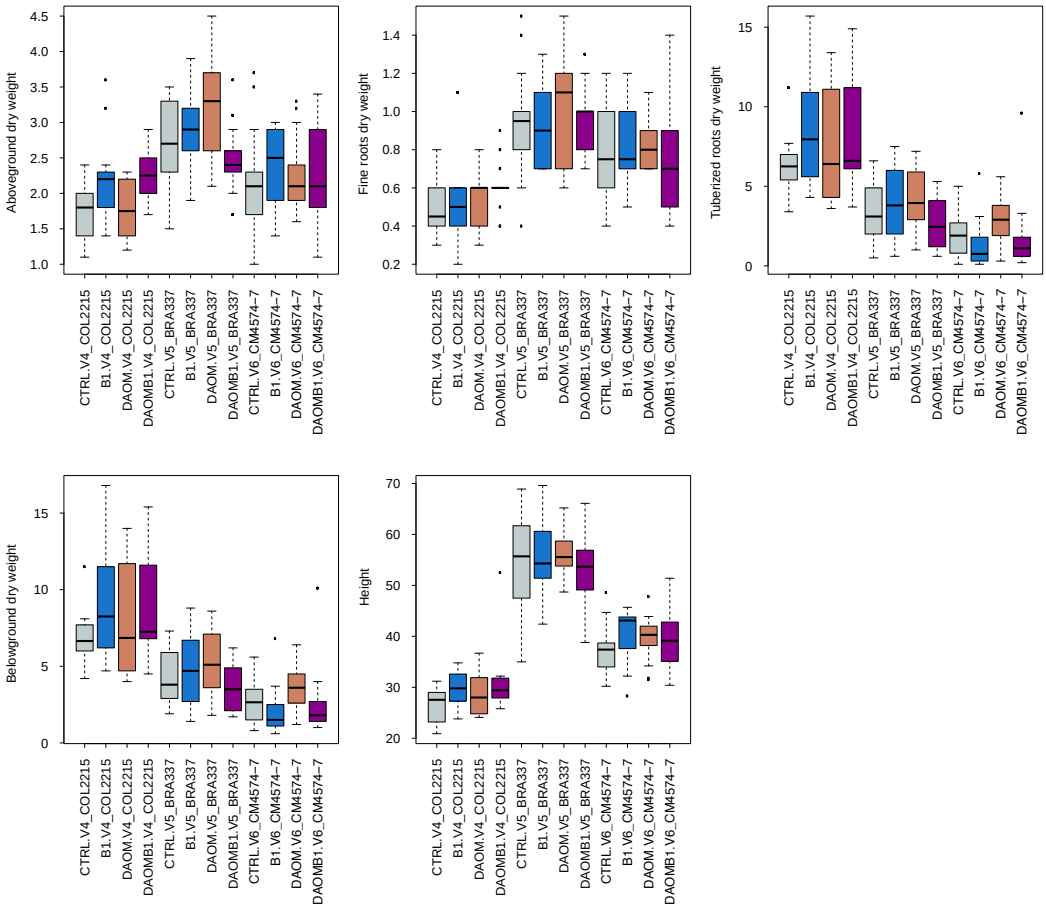
Mixed-linear model (binomial-logit)

	numDF	**denDF**	**F-value**	**p-value**	
(Intercept)	1	182	10037	<.0001	***
AMF	2	182	5	0.011	*
CULTIVAR	4	182	162	<.0001	***
AMF:CULTIVAR	8	182	1	0.821	

Annexe IV. Role of host intra-species variability on coexistence and effect on cassava (Manihot esculenta Crantz) of two isolates of Rhizophagus irregularis revealed by RNA-seq

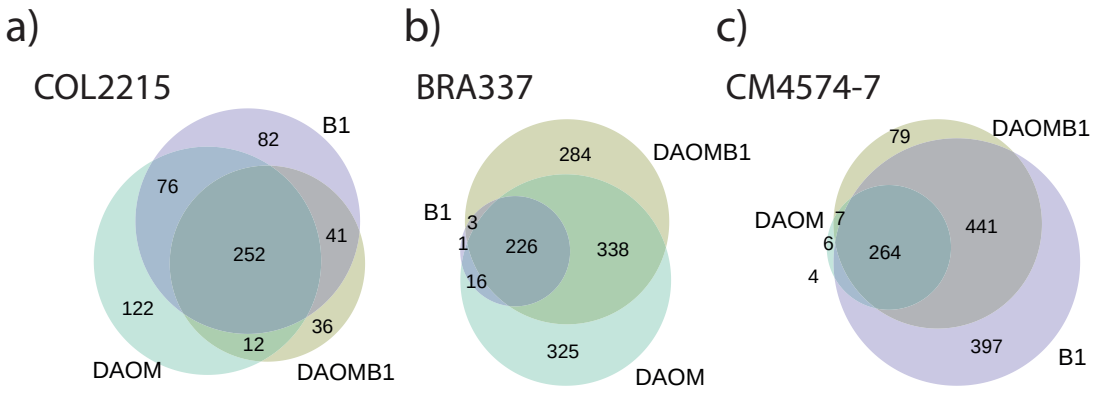
Supplementary Figure 1.

Phenotypic measurements of the different cassava cultivars. We used 14 replicates for each phenotypic measurement.a) Aboveground dry weight of plants per mycorrhizal treatment and cultivar. b) Fine roots dry weight of plants per mycorrhizal treatment and cultivar. c) Tuberized roots dry weight of plants per mycorrhizal treatment and cultivar. e) Belowground dry weight of plants per mycorrhizal treatment and cultivar. f) Plants height per mycorrhizal treatment and cultivar.



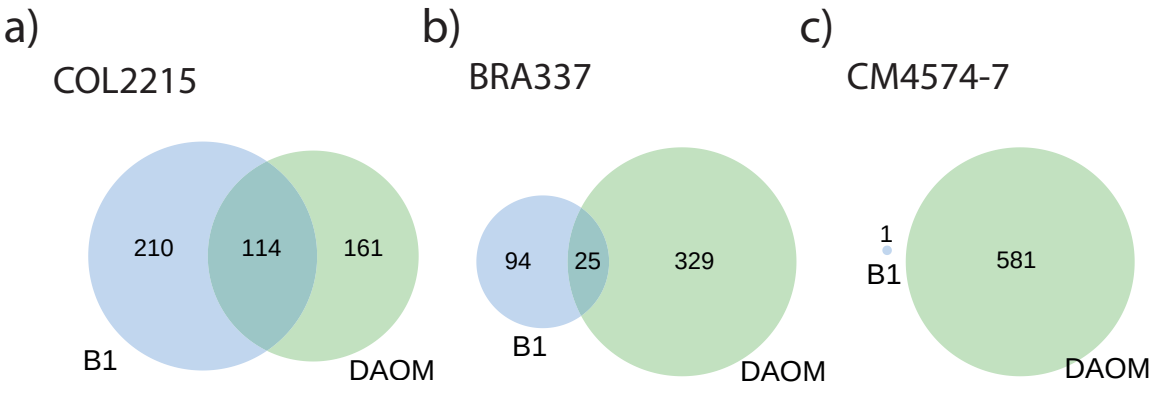
Supplementary Figure 2.

Venn diagram of the cassava genes that were significantly different between the mycorrhizal treatments and the non-mycorrhizal treatment. a) Cultivar COL2215, b) Cultivar BRA337 and c) cultivar CM4574-7. DAOM refers to isolate DAOM197198, B1 refers to isolate B1 and DAOMB1 refers to the co-inoculation treatment.



Supplementary Figure 3.

Venn diagram of the R. irregularis genes that were significantly different between the single-inoculations treatments and the co-inoculation treatment. a) Cultivar COL2215, b) Cultivar BRA337 and c) cultivar CM4574-7. DAOM refers to isolateDAOM197198 and B1 refers to isolate B1.



Supplementary Figure 4.

Influence on the reference alignment on the absence of expression on the different isolates. a,c,d) Plot of the gene-expression of the different mycorrhizal treatments on cultivar BRA337. b) alignement of the correspondant sequences of the hypothetical protein. DAOM refers to isolate DAOM197198, B1 refers to isolate B1 & DAOMB1 refers to the co-inoculation treatments.